Understanding the large-scale dynamics of the interstellar medium in barred galaxies

Mattia Carlo Sormani
Balliol College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity Term 2015

Understanding the large-scale dynamics of the interstellar medium in barred galaxies

Mattia Carlo Sormani

Balliol College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2015

Abstract

We study the large-scale dynamics of the interstellar medium in barred galaxies. The interest in doing this is two-fold. On the one hand, the hydrodynamic flow is the source of many interesting physical phenomena, such as shocks and spiral arms. On the other hand, it is a powerful tool to constrain the characteristics of individual galaxies.

Approximately half of this thesis is devoted to understanding the general characteristics of the gas dynamics. There are two key ingredients. The first is ballistic closed orbits. We investigate the connection between closed orbits and the full hydrodynamic flow. We show how they also form the basis to explain bar-driven spiral arms. The second key ingredient is shocks. These can arise at the transition between different families of closed orbits, or when the spiral arms become too strong. Under certain conditions, shocks become hydrodynamically unstable and produce density fluctuations that may have an observational counterpart in the form of asymmetries in the Milky Way molecular gas distribution. We also explore systematically how the gas flow is affected by a change in the parameters that characterise the bar, such as its length, strength and pattern speed.

Through the study of the gas dynamics, inferences can be made about the structure of galaxies. We have compared our models with observations of the Milky Way. First, we revisit and refine the Binney et al (1991) model, one of the most successful models for the gas flow in the Galaxy. Then, we present new models that while preserving the good properties of the previous model can also correct its main shortcomings. These models can qualitatively account for most of the observational signatures of the Galactic bar and allow us to constrain its characteristics, such as its pattern speed, length and strength. However, we show that is difficult to find a model that accounts for all the important observational features simultaneously due to the high dimensionality of parameter space involved. We argue that automatic fitting method are necessary. To this end, we develop a new quantitative method to fit Milky Way longitude-velocity diagrams that is based on feature matching.

This thesis is dedicated to my grandparents - Nonna Luigia, Nonna Nicla and Nonno Paolo
for the diversity of their teachings and their constant love and support

Publications Related to this Thesis

This thesis has used material from the following first-authored publications.

Chapter 3 uses material from Sormani M. C., Binney J., Magorrian J., 2015a, MNRAS, 449, 2421.
Chapter 4 uses material from Sormani M. C., Binney J., Magorrian J., 2015b, MNRAS, 451, 3437.
Chapter 5 uses material from Sormani M. C., Binney J., Magorrian J., 2015c, MNRAS, 454, 1818.
Chapter 6 uses material from Sormani M. C., Magorrian J., 2015, MNRAS, 446, 4186.

Acknowledgements

Firstly, I would like to thank Dr. John Magorrian for his supervision during my DPhil. He has been fundamental in the developing of this thesis, generously providing ideas and guidance. I also thank him for his invaluable help for many programming issues related to this thesis: not only he provided the bulk of the hydrodynamic code, but his almost inhuman ability to immediately find bugs in any piece of computer code sent his way has been exploited many times. But maybe even more impressive has been his patience against my difficult and grouchy character. He has often been not only a supervisor, but also a friend.

Thanks also go to the current and past members of the Oxford Galaxy Dynamics group. I would like to thank, in particular, James Binney, that in the second half of my DPhil has been like a second supervisor to me. I've been very lucky in having the chance to work with someone with his breadth of knowledge, creativity and wit. Special thanks also goes to Ralph Schönrich: his strenuous opposition to anything anybody says is invariably source of interesting discussions and original ideas. I also enjoyed coffee time and Monday meetings discussions with Paul McMillan, Michael Aumer, Eugene Vasiliev and Stuart Sale.

I'm grateful to the University of Oxford and the Oxford University Press, which granted me a Clarendon Fund Scholarship to fund my DPhil.

A lot of my time in Oxford has been spent in Room 2.9 of the Rudolf Peierls Centre and I would like to thank all the people that inhabited it. Jason Sanders has been a good friend and his advice has been precious on several occasions. Theo Le Bret and Payel Das have brought a breath of fresh air with their cheerfulness and enthusiasm towards any physics & non physics problem thrown at them.

A warm thanks goes to Steve Shore. He has been almost like a father, inspiring, advising and trusting me.

I also thank Stefano Bolzonella, Enrico Morgante, Laura De Stefanis, Emanuele Sobacchi, Federico Fabiano, Simone Surdi, Alessandro Moia, Maria Colombo, Nicolò Grilli, Stefano Recanatesi and Giacomo De Palma. Most of what I know of physics I owe to them, and this thesis would not have been possible without their legacy.

A special thank also goes to my best friends and housemates during the past two years, Arnold Mathijssen (or the Goodold) and Bruno Bertini. They have been my main companions in my Oxford life-adventure. Our internal mutual feedback system based only on "delirious glorification" and "harsh condemnation" has worked very well. I will never forget our evenings spent together and the fierce debates. I also thank Arnold in particular for making me discover how bad a meal can be and how much I hate opera, especially when sung in the shower every morning.

I thank Nicole Siri for persistently being my friend despite the distance, for the uncountable day and night conversations in which we shared euphoria and delusions, happiness and sadness, clever ideas and dull leitmotifs.

Finally, but most importantly, I thank Viviana. Two years ago she has changed my life and I've never been as happy as I am since I know her. With her, my days are full of love and joy. Every day is also a deep intellectual challenge, as I struggle and fail to imitate her fastness of mind. I hope in the next years I will be able to give her back at least a fraction of what she has given me.

Contents

CHAPTER 1

Introduction

The subject of this thesis is the study of the gas dynamics in barred galaxies, with a particular emphasis on the Milky Way. This chapter has the purpose of setting the scene that provides the context for the remaining chapters. A fundamental premise is that the Milky Way is a barred spiral galaxy. Thus we begin in the next two sections by briefly recounting first how we came to recognise the Milky Way as a spiral galaxy, and then how we became convinced that it is also barred. In the third section of this chapter we introduce the physical picture underlying the study of the gas dynamics. In the last part of this introductory chapter we make a small digression to discuss the motivation for writing this thesis and more in general for pursuing science. The chapter concludes with an overview of the thesis.

1.1 The Milky Way as a spiral galaxy

The realisation that we live in a spiral galaxy, a peer of the spiral nebulae found in the sky, did not come to astronomers overnight. It was instead the result of decades of speculation, debates and accumulation of evidence. The fact that our Galaxy might display some spiral structure was already explicitly suggested by Alexander (1852) only a few years after the discovery of spiral structure in the nebulae by Lord Rosse. But Alexander's suggestion took almost a century to be accepted at large by the community.[1]

By 1940, most astronomers agreed on a conceptual framework for what we would now call our nearby universe. A number of questions that provoked debates for decades or even centuries were finally settled, and most astronomers agreed that:

- We live in the Milky Way, a gravitationally-bound collection of stars arranged in the shape of a disk.
- Our position in the Milky Way is not near the centre, but in the periphery. Approximate values for the

[1] Of course, the present introduction is not meant in any way to be a proper historical account, but should be taken as a mere selection of facts, with the purpose of putting the reader in perspective. There is no point in giving here a mediocre account when excellent books on the subject are available (see for example Whitney, 1971; Berendzen et al., 1984).

size of our Galaxy and our distance from the centre were known, within a factor of two of the current value.

- The disk of the Milky Way rotates around a definite centre, and we participate to this rotation. A swarm of lesser systems, the globular clusters, surrounds the disk, but does not participate to its systematic rotation.
- Spiral nebulae are not systems inside the Milky Way, but "island universes", systems comparable in size to the Milky Way.
- The Milky Way contains dust that obscures our view of the plane of the disk and makes object look dimmer than they would appear in the absence of the dust.

Thus, by 1940, few astronomers doubted that we live in a spiral galaxy. But, as Whitney (1971) nicely phrases, nobody was able to point to an example and say "that's what our looks like". The two favorite candidates for a comparison were M31, that has a large nucleus and tightly wound arms, and M101, with a small nucleus and open arms. The astronomical community, that now finally had (and agreed on) a general conceptual framework (the list of items above), was eager to fill in the picture with the details.

However, progress was fragmentary until the advent of radio astronomy, and in particular before the 21-cm line of neutral hydrogen was measured and the first survey at this wavelength was carried out. A paper by Oort (1952) summarises the situation just before this revolution would take place (see Fig. 1.1). Until then, the vast majority of deductions on the large scale structure of the local universe were based on optical observations. Dust is not transparent to optical light, so sitting in the Galactic disk is like sitting in the middle of a foggy field. Knowledge was mostly limited to a cylindrical region with axis through the Sun and perpendicular to the Galactic plane with a radius of approximately 1 kpc, although outside the Galactic plane it was possible to see much further out (which is why Shapley had been able to study the globular clusters and infer from them the approximate location of the Galactic centre, see Chapter 18 in Whitney 1971 and references therein).

Suddenly, when the first survey in the 21-cm line was completed, the picture was filled with details and astronomers were able to produce a face-on map of the Milky Way, see Fig. 1.2.[2] However, puzzling phenomena that could not be explained on the basis of the simple assumptions (e.g., purely circular motion) used to produce the map were already noted, especially in the vicinity of the Galactic centre. For example, high velocity gas that did not appear to be participating in the circular rotation of the Galaxy perplexed astronomers. As we will see in more detail in the next section, the modern answer to this issue did not took long to be proposed (de Vaucouleurs, 1964). The key insight was that our Galaxy might host at its centre an elongated structure called the bar, and that the bar can influence the gas flow as to produce the puzzling observational features. But as for the recognition that we live in a spiral galaxy, also the recognition that we live in a *barred* spiral galaxy took decades to be fully accepted by the community. We review the evidence for the Milky Way bar in the next section.

[2]See item (3) in Sect. 2.2 for more details on how the data can be deprojected.

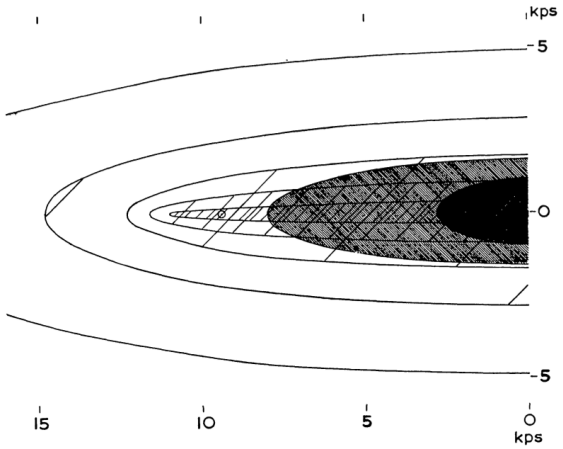


Figure 1.1: How astronomers thought our Galaxy just before the radio astronomy revolution. This schematic drawing is from Oort (1952). It shows a section perpendicular to the disk passing through the Sun (indicated by a small circle) and the Galactic centre. The conceptual picture of our Galaxy essentially coincided with the modern one, and astronomers were eager to fill this picture with details.

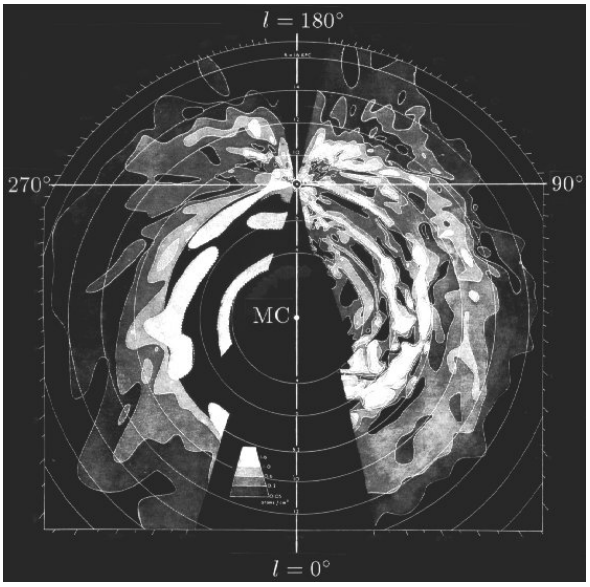


Figure 1.2: The famous map of the Milky Way produced after the first 21-cm survey was carried out, from Oort et al. (1958).

Today, we still struggle to produce a precise map for the spiral structure of our own Galaxy that improves the one shown in Fig. 1.2. One of the main difficulties lies in overcoming the assumption that gas flows on circular orbits. The Milky Way bar produces large non-axisymmetric motions, which invalidate the circular orbits assumption, and the effects are so severe that on several occasions they damped the enthusiasm of the early days. For example, Burton et al. (1992) did not feel confident in showing a map for the spiral arms in our Galaxy. However, progress is being made and we hope that the study of the gas dynamics will eventually teach us how to properly take into account the non-axisymmetric flows and to unravel how our mates living in external galaxies see us.

1.2 The Milky Way as a *barred* spiral galaxy

Today there is a consensus that the Milky Way is a *barred* spiral galaxy. This is not surprising after all when we consider that around two thirds of external spiral galaxies are barred (Sellwood & Wilkinson, 1993). The bar is a cigar-shaped structure composed of stars and lying in the Galactic plane. The non-axisymmetric gravitational field produced by the bar is believed to cause the gas flow to deviate significantly from purely circular motion, particularly in the region within 3 kpc of the Galactic centre. Astronomers believe that this can explain a number of features found in atomic and molecular gas observations. In this section, we briefly review how we came to this view. For other reviews on the Galactic bar see for example Gerhard (2002); Fux (2004); Merrifield (2004); Ferrière et al. (2007); Gerhard & Wegg (2014)

As we have seen in the previous section, 21-cm HI observatons made available to the astronomers a comprehensive view of the disk of our Galaxy. Being able to see through the dust enabled them to study for the first time the large scale structure of our Galaxy in a non fragmentary way. The new HI data were interpreted on the plausible assumption that the gas was participating to the mean rotation of the Galaxy, flowing on purely circular orbit. One of the results of this analysis was the already mentioned map shown in Fig. 1.2. But, as we mentioned in the previous section, already in the early days of radio astronomy it was noted that this assumption cannot be valid in the region within 3 kpc of the centre (e.g. Oort & Rougoor, 1959), as the gas has velocities that are forbidden to purely circular motion. The initial interpretation was that the hydrogen gas in the inner regions is expanding, perhaps as a consequence of some explosive phenomenon. It did not take long to de Vaucouleurs (1964) to propose an alternative explanation which did not involve a net outward mass flow: if the Milky Way were a barred galaxy, large observed non-circular motionshypotesis could be explained by gas streaming along the bar. His proposal came by analogy with external galaxies: as noted above, about two thirds of external disk galaxies are barred.

The expanding hypotesis and the streaming hypotesis stood as plausible alternatives for decades (Oort, 1977a,b). During the 70s and the 80s, advances in radio astronomy allowed other cold and dense components of the interstellar medium to be detected besides HI. In particular, spectral line coming from molecular species such as ^{12}CO, ^{13}CO and CS were detected and studied in large portions of the sky. During the same years, purely kinematical models (in the sense that they assumed that gas follows closed streamlines, but without a physical model for the origin of the assumed streamlines) were developed aimed at explaining the gas observations in terms of non-axisymmetric gas flow (Peters, 1975; Liszt & Burton, 1980). Astronomers gradually realised that the expanding hypotesis drawbacks were serious. For example, explosive phenomena would drive gas more out of the plane rather than in it, and N-body simulation started to show that bars occur naturally and are not dynamically peculiar objects, so the streaming hypotesis was more a priori probable. However, there was no convincing dynamical model based on the streaming hypotesis, and defender of both hypotheses remained.

The barred nature of the Milky Way was considered fully established only in the 90s, when two major advances took place. First, the presence of the bar was confirmed by near-infrared photometry (Blitz &

Spergel, 1991; Weiland et al., 1994; Dwek et al., 1995; Binney et al., 1997). This kind of data is still the most direct proof of the barred nature of the Milky Way. Fig. 1.3 offers a view of how our Galaxy appears at these wavelengths after the data have been corrected for extinction. The range 2 - 4 μm is optimal to observe stars near the centre of the Galaxy, because at these wavelengths, stars radiate a substantial fraction of their luminosity, while dust is much less opaque than at other wavelengths and its thermal emission is negligible. However, extinction due to dust is still non negligible and must be properly taken into account. The asymmetry between positive and negative longitudes caused by the bar is apparent in the figure. Indeed, a general observational signature of the Galactic bar is a departure from symmetry between positive and negative Galactic longitudes. Hence, the data are interpreted as an indication that we see a bar and that the Sun does not lie on one of the symmetry axes of the bar.

The second major advance was that purely kinematical models were refined into dynamical models. These can be divided in two categories. The first type is the ballistic models, in which the gas streamlines are approximated by closed orbits in an assumed underlying potential (Gerhard & Vietri, 1986; Binney et al., 1991). In particular, Binney et al. (1991) convinced the community that a dynamical explanation for the streaming hypotesis was possible. The second type of dynamical models is based on full hydrodynamical calculations. These involve finding quasi-steady solutions and/or running hydrodynamical simulations, usually under the assumption that the dynamics of the gas is governed by Euler's equation, complemented by the equation of state of an ideal isothermal gas. The first, pioneering simulation meant to reproduce the Milky Way was carried out by Mulder & Liem (1986), who found numerically a quasi-steady solution of gas-dynamical equations in a given barred gravitational potential. More recently, many simulations have been carried out with a variety of methods: sticky-particle codes (Jenkins & Binney, 1994; Rodriguez-Fernandez & Combes, 2008), Eulerian grid-based codes (Weiner & Sellwood, 1999) and smoothed particle hydrodynamic simulations (SPH) (Lee et al., 1999; Englmaier & Gerhard, 1999; Bissantz et al., 2003) in externally imposed potentials undergoing rigid rotation; SPH simulations coupled to self-consistent 3D N-body barred models of the Galaxy (Fux, 1999); SPH codes that include phenomenological terms to model heating and cooling processes such as radiative cooling, heating caused by UV radiation, star formation and SN feedback (Baba et al., 2010; Pettitt et al., 2014).

In addition to these two major proofs, other independent evidence have corroborated the view that our Galaxy hosts a bar. It is possible to see the positive-negative longitude asymmetry produced by the bar in discrete source counts (for example red clump stars) which show more or brighter objects on one side than on the other. This was already clear in the 90s (Stanek et al., 1994), but recently the 2MASS data (Skrutskie et al., 2006) allowed spectacular 3D reconstructions of the bar (Wegg & Gerhard, 2013; Wegg et al., 2015). Less spectacular evidence come from gravitational microlensing optical depth data (see for example Zhao & Mao, 1996; Popowski et al., 2005; Wyrzykowski et al., 2015). When a massive object (lens) passes between us and a bright background object, it can bend its light due to gravitational lensing effect. If the mass of the lens is low and we cannot resolve the image, we can still detect an apparent brightening of the background object. This is called a microlensing event. Counting such events provides a measure of the

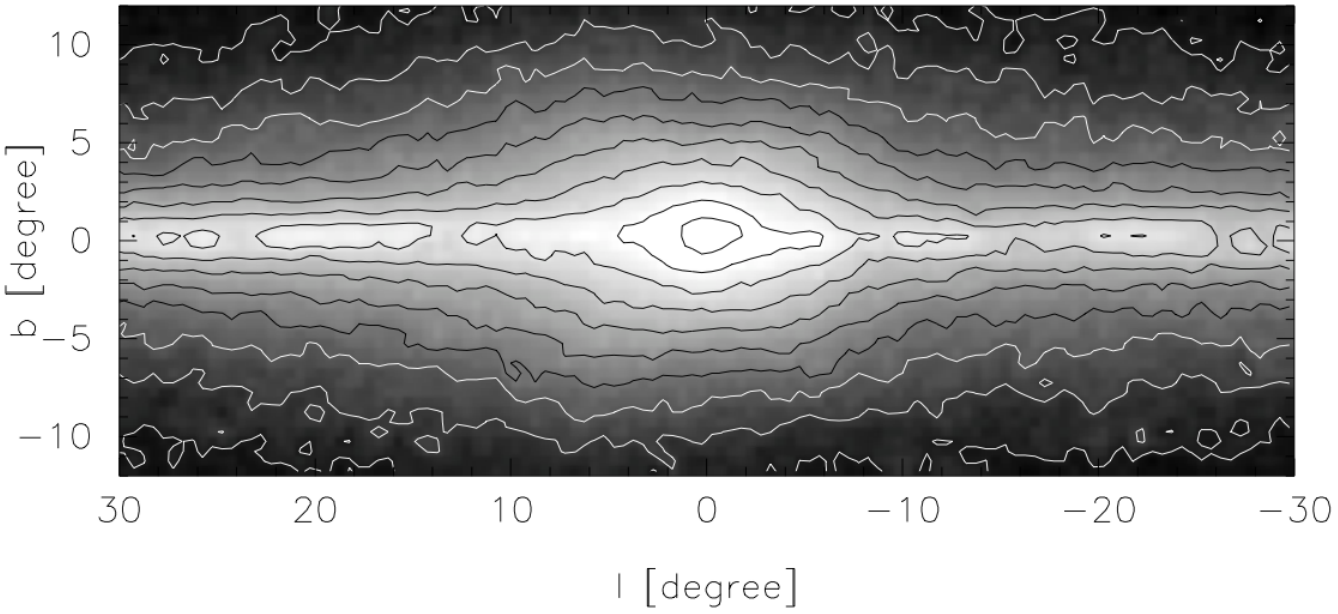


Figure 1.3: Extinction-corrected K-band (2.0 to 2.4 μm) map of the Galactic central regions from the DIRBE instrument on board the COBE satellite. Contours are spaced by 0.5 magnitude. The figure is reproduced from Fux (2004) and used data from Weiland et al. (1994).

integrated mass along the surveyed line of sight. Observations give values that are too high to be explained with an axisymmetric Galactic model which is symmetric between positive and negative l, thus providing evidence for an asymmetry which is interpreted as evidence for a bar.

Despite all the progress, as we will see in this thesis, a number of questions regarding the properties of the bar are still open. Bars are believed to rotate with a well defined pattern speed, otherwise we would observe winding rather than straight structures. The numerical value of the Milky Way bar pattern speed is subject to great uncertainty in the literature, with values ranging from $\Omega_{\rm p} \simeq 20$ km s^{-1} kpc^{-1} to $\Omega_{\rm p} \simeq 70$ km s^{-1} kpc^{-1}. The angle between the major axis of the bar and the Sun-Galactic centre line is also subject to great uncertainty, and determinations usually lie in the range $\phi \simeq$ 10-40°. Other parameters of interest are also not strongly constrained, such as the bar length, mass and axis ratio. One of the major difficulties in determining these parameters is that many are degenerate in observational space, i.e. the same observational result could be obtained with two (or more) different sets of parameters. For example, it is easy to see that the same sky-projected bar length could be obtained with different couples of values for the real physical length of the bar and for the angle ϕ. In Chapter 5 we will discuss some degeneracies and in Chapter 6 we will provide our own estimates of some of these quantities. In the next section we will focus on the dynamics of the gas flow.

Finally, we should mention that in the next few years, the recently launched *Gaia* satellite will provide us with an unprecedented detailed spectroscopic survey of the Milky Way, and will provide accurate parallax distance measurements for one billion stars at distances extending to the bulge region. This should provide independent constraints on the properties of the bar and a further boost in our understanding of the detailed structure of the Galactic bar.

1.3 Gas dynamics in galaxies

Understanding the large-scale gas dynamics in spiral galaxies is a challenging task that involves physical modelling on different scales. The first thing we need is a set of equations of motion that describes the coarse-grained dynamics at the scales we are interested in. To come up with these, we need to understand the microphysics of the interstellar medium on the fine-grained scales and prune as many inessential details as possible, often making dodgy simplifying assumptions (e.g., isothermal equation of state). Note that here the work *micro* does not necessarily mean an analysis at the molecular level, but also includes all the physical processes over which we want to coarse-grain and that we want to incorporate phenomenologically into the equations of motion, for example small-scale turbulence. On bigger scales, we need to understand large-scale flows and the dynamical processes that emerge from the equations of motion, for example shocks, hydrodynamical instabilities, generation of spiral arms, and so on. Of course, it is not always easy to draw a line between what we want to coarse-grain over and what not: for example, turbulence appears at many different scales. In the case we want to model a particular system - e.g., the Milky Way - we also need a procedure to compare and fit the models to observations. As we will see, this is also a non-trivial task.

In this section we briefly discuss these three aspects - the microphysics, the macrophysics and the fitting to individual models - that are essential to understanding the gas dynamics in galaxies. While each of the following subsections is dedicated to one aspect, it must be noted that these three aspects are linked by feedback loops. Hence, results regarding one of them can influence our understanding of the other, and we should bear in mind that they cannot be studied completely independently.

1.3.1 Understanding the microphysics

The interstellar medium (ISM) is a complex dynamical entity and it is not easy to justify a set of coarse-grained equations to describe its large-scale dynamics. It is made of several components coupled together, its three main constituents being gas, dust and magnetic fields. In addition to these three main constituents, minor constituents are also present, such as relativistically moving electrons and protons. We now briefly review how we can argue for a set of equations. For more comprehensive discussions on the structure and dynamics of the ISM see for example Burton et al. (1992), Shu (1992), Binney & Merrifield (1998), Spitzer (1998), Shore (2002), Klessen & Glover (2014), and references therein.

The ISM is highly inhomogeneous and shows structure on a vast range of scales, from sub-parsec to kiloparsecs. Most of the mass appears to be concentrated into clouds, but we also know that these are embedded in a warmer diffuse component and non-negligible exchanges of mass and momentum take place between the two. Clouds can be made of smaller clumps, holes and filaments, and these can contain even smaller subclumps, and so on. The physical state of clouds is also very inhomogeneous. The gas comes in a wide range of temperatures, densities and chemical compositions: from very cold, dense gas in molecular form to very hot, low density gas in ionised form. The dust, which is made of grains whose sizes range from

tens of molecules to small (0.1μm) solids, correlates with the gas but the proportion can vary. The magnetic fields have large scale as well as small scale variations.

In an ideal description of the ISM, we would have equations able to describe each component and its interaction with the others in detail. We would like to use this ideal description to reproduce the structure on all observed scales. For the moment, as one could guess from the short summary above, this is simply too complex. We must resort to successive approximations and phenomenological models that focus on reproducing structure on a particular scale of interest. In this thesis, we are interested in the large-scale structure, i.e., of orders of tens of parsecs or more.

One of the first dilemmas to face is whether we should take explicitly into account the granularity of the ISM in our description or whether it is better to think of it as a smooth fluid described by the Navier-Stokes equations. Both have been explored in the literature. The idea that ISM can be treated as a collection of cloudlets is behind one of the simplest ways of modelling the ISM, the ballistic approximation. This assumes that the gas streams along nested ballistic closed orbits in the underlying gravitational potential. The idea that the ISM is granular has also led to more sophisticated treatments, such as the sticky-particles schemes (see for example Jenkins & Binney, 1994; Rodriguez-Fernandez & Combes, 2008, and references therein). In these, the ISM is made of cloudlets that can interact with each other inelastically and can exchange mass and momentum. However, the application of such schemes has turned out to be cumbersome from the technical point of view, due for example to difficulties in imposing the detailed conservation of angular momentum during interaction between cloudlets.

We think that the smooth description is currently the best in practice, and we mostly stick with it in this thesis. Its application has led to successful explanations of many aspects of the large-scale structure of the ISM (see for example Sellwood & Wilkinson, 1993; Binney & Merrifield, 1998). It is computationally cheap and the underlying equations have been studied extensively in a countless number of other contexts and disciplines. Moreover, some of our simulations show (see Sect. 6.6) that on the scales we are interested in the results obtained with sticky particles schemes do not differ significantly from a smooth description based on Navier-Stokes, despite the underlying equations being different. But when is the smooth approximation justified? The generic answer is that we need to compare the collisional mean free path with the length scale of interest. If the mean free path is small compared to the latter, then a smooth coarse-grained description makes sense. However, in the case of the ISM we must proceed with caution. For it is true that ultimately the ISM is made (mostly) of molecules, and thus one can be tempted to follow the same arguments used for example for air in a room to justify the use of the Navier-Stokes equations. But due to the structure of the ISM it is not clear that molecules should be considered as the fundamental entities. It is possible for example that individuality should be recognised to clouds, and cloudlets regarded as the particles to be used in calculating the mean free path, and not the molecules. Ultimately, the fact that even from a coarse-grained picture of an external spiral galaxy we can discern clumpiness in the ISM should prompt us to keep in the back of our mind the shakiness of the smooth approximation.

Just saying that we describe the ISM by a smooth fluid does not complete the description. First of all, it

is well known that to close the system of fluid equations something relating the pressure to other quantities such as temperature and density of the gas is needed, e.g. an equation of state. Second, many ingredients can be added on top of the Navier-Stokes equations. For example, magnetic fields and are believed to contribute significantly to the total pressure inside the ISM, so we could ask whether we should explicitly include them. Other processes that appear to have major effects are heating and cooling processes: photoelectric heating, collisional cooling, supernova explosions, cosmic rays, stellar winds, and so on. Unfortunately, each of these ingredients is complex in itself and nobody fully understands the whole picture yet.

For most of this thesis we choose very simple models, which do not include many potentially important effects. We use the Euler fluid equation complemented by an isothermal equation of state. In a frame rotating with pattern speed $\mathbf{\Omega}_\mathrm{p}$ these read

$$\partial_t \mathbf{v} + (\mathbf{v}\cdot\nabla)\,\mathbf{v} = -\frac{\nabla P}{\rho} - \nabla\Phi - 2\mathbf{\Omega}_\mathrm{p}\times\mathbf{v} - \mathbf{\Omega}_\mathrm{p}\times(\mathbf{\Omega}_\mathrm{p}\times\mathbf{x})\ , \tag{1.1}$$

$$P = c_\mathrm{s}^2\rho\,, \tag{1.2}$$

where $\mathbf{x}$ is the position vector, $\mathbf{v}$ is the velocity in the rotating frame, ρ is the gas density, Φ is the gravitational potential and P is the pressure. Eq. (1.1) is the Euler equation, in which $-2\mathbf{\Omega}_\mathrm{p}\times\mathbf{v}$ is the Coriolis force and $-\mathbf{\Omega}_\mathrm{p}\times(\mathbf{\Omega}_\mathrm{p}\times\mathbf{x})$ is the centrifugal force. Eq. (1.2) is the isothermal equation of state, where c_s is the sound speed which is assumed to be a constant number. The details of the numerical schemes will be explained in later chapters.

One might ask why isothermal and not, for example, adiabatic. The traditional answer is that the pressure in Eq. (1.2) is a phenomenological quantity which also includes the effects of small-scale turbulence (e.g. Roberts, 1969; Cowie, 1980). The "temperature" of the isothermal assumption is thus an equivalent temperature related to the velocity dispersion of clouds rather than a microscopic temperature. The average energy content of the gas, and hence the temperature, is assumed to be held constant by some energy balancing that takes place in the ISM between heating and cooling processes, on the basis that the velocity dispersions of observed clouds seem to support this hypotesis. This means that any heating due to compression, for example at a shock, is instantaneously radiated away to restore the initial temperature. Thus, although due to the complexity of the physics involved it cannot be said that is very clear what one has to assume for the equation of state of the ISM, the crude isothermal approximation is believed to be more accurate than other equally simple assumptions such as the adiabatic approximation.

Is it always better to include more effects? For example, should we aim at including in our equations as many things as possible, such as magnetic fields, heating and cooling, and so on? We believe that the answer is no. Simultaneous inclusion of many physical effects that are not well understood separately can lead to results that are difficult to interpret and do not constitute real progress. The tendency of making things bigger and more complex but not clearer is a dangerous habit that is also affecting, for example, the study of structure and galaxy formation in a cosmological context. We believe that understanding simpler,

idealised systems is an unavoidable step in the path to comprehend the ISM in all its complexity. In fact, as we will see in the course of this thesis, despite the simplicity of the equations of motion, the resulting physics will display a richness that we still do not fully understand.

1.3.2 Understanding the macrophysics

There are several large-scale dynamical phenomena taking place in the ISM that we need to study if we want to move on from simple axisymmetric models of the gas flow.

One of the most ubiquitous of such phenomena issupernova shocks. In essence, shocks in a fluid flow are discontinuities in physical quantities. In astrophysics, they arise in a countless number of different contexts and situations, from atmospheric shocks in stars and planets to supernova explosions and up to intergalactic shock fronts that results from collisions of whole galaxy clusters. The subject is too broad to be covered here, and the interested reader should consult the literature (see for example Zel'dovich & Raizer, 1967; Landau & Lifshitz, 1987; Shu, 1992; Dyson & Williams, 1997; Shore, 2007). In this thesis, we are mostly concerned with large-scale galactic shock waves that occur in spiral galaxies (as opposed to small-scale shocks of at most few parsecs in size, which can be triggered for example by stellar activity). Examples of large-scale shocks in spiral galaxies can be found for example in Roberts (1969), Roberts (1971), Roberts (1979), Lin & Roberts (1981), Prendergast (1983), Athanassoula (1992) and Sellwood & Wilkinson (1993). In barred galaxies, as we will see in the course of the thesis, straight shocks that connect the end of the bar with an inner nuclear ring play a key role in understanding the gas flow, both dynamically and observationally.

Several other macroscopic phenomena play an important role in studying the large-scale gas flow. The mechanism that causes the generation of spiral arms driven by a bar will be the subject of Chapter 4. Turbulence appears in various phases of the ISM and on a wide range of spatial scales. In this thesis turbulence is not discussed explicitly, being buried in a single phenomenological parameter that is assumed to take care of its effects, the sound speed of the gas. This is done on the premise that we expect details of turbulence to be important only on scales smaller than those of interest. However, turbulence is always lurking below the surface and will demand some active attention when we meet an hydrodynamic instability which occurs along curved shocks called the wiggle instability.

Shocks, turbulence, instabilities and spiral arms are broad topics and each of them could be the subject of a thesis on its own. We cannot study all of them in detail in this thesis. The important piece of information contained in this subsection is that we cannot forget that they exist. Important macroscopic phenomena are often more general than the problem at hand and occur also in different contexts, which means that they could (and should) be understood singly, and only then we will have the proper tools to attack the problem of understanding how the ISM works as a whole dynamical entity.

1.3.3 Fitting models to individual systems

In astrophysics it is often convenient to model in detail a particular system which then serves as the basis to understand a whole class (e.g., the Sun seismology in relation to seismology of other stars). In our case, we might like to model a particular barred spiral galaxy in detail to understand barred galaxies in general. Thus a quantitative way of comparing models with the real system is necessary. Finding a quantitative fitting method for the result of smooth hydro simulations is not a trivial task, both for the case of the Milky Way and for external galaxies. The main reason is that traditional methods, such as χ^2 fitting, work poorly for this problem. As a result people have relied on by-eye comparisons, which is not a quantitative method, or employed quantitative methods that use only a small portions of the information available in the data. In the case of the Milky Way the problem is even harder because of our embedded position inside the disk. Unravelling a plausible bird-eye view of the Galaxy is a complicate and degenerate inverse-problem, and we would like to use all the available information. The problem of fitting our own Milky Way will be the subject of Chapter 6.

1.4 Why this thesis, and what for

Broadly speaking, the scope of this thesis is to advance our understanding of the gas dynamics in the Milky Way and in other barred galaxies. At this point, the reader may ask why one would be willing to do so.

There is no universal answer to this question, as different people have different motivations for pursuing science, ranging from pure intellectual entertainment to possible practical applications and passing from vanity. Some people just enjoy solving problems, and the quest of knowing more about our Galaxy can be seen as a giant puzzle, source of many mini-problems. Regarding this it is curious to note that according to Murray GellMann (2001), Everett didn't produce his many-worlds interpretation of quantum mechanics because he was particularly passionate about the subject, but just because he liked to solve problems, and the problem of the interpretation of quantum mechanics was just one problem. After completing his PhD on this subject, Everett spent most of the rest of his life working on military problems, which according to Gellmann he was happy with as a source of interesting problems to solve. This example is particularly interesting, because if true it means that even deep philosophical problems can be tackled for the simple enjoyment of solving problems rather than because of their philosophical interest. Therefore it must be certainly possible that for similar reasons someone is lead to solve problems in Milky Way science.

However, if the only enjoyment of science was the pleasure of solving problems, playing chess or a similar game would be enough and the same. Other people, consciously or not, may find among their motivations that they like to be part of an intellectual elite. The thrill of the discovery is, of course, another form of intellectual vanity (to which I'm not immune) as it presumes a social ecosystem to be at all possible. A probably apocryphal story told by Feynman (1988) is a good example of this:

[...] [Arthur Eddington] had just figured out that the stars get their power from burning hydrogen in a

nuclear reaction producing helium. He recounted how, on the night after his discovery, he was sitting on a bench with his girlfriend. She said, "Look how pretty the stars shine!" To which he replied, "Yes, and right now, I'm the only man in the world who knows how they shine." He was describing a kind of wonderful loneliness you have when you make a discovery.

Another simple motivation that can be identified is certainly the curiosity of knowing more about our universe and in particular about our home, the Milky Way. What would it really look like if we could fly over it? Studying the gas dynamics is a way for taking part in the bigger quest of knowing more about the Galaxy in which we live. As we will see in the following chapters, it allows us to constrain the gravitational potential and hence to infer the mass distribution of our Galaxy, as well as other physical properties of the bar at its centre. But to what extent do we really care about these questions if they are not just pieces of something more meaningful?

On the premise that all knowledge, intended as every possible information that can be communicated, is interesting and worth learning, some people argue that we do science to increase the amount of human knowledge. One should therefore pursue science because "there is so much to know" out there. But I believe one should be wary in making such statements. Is everything out there really equally worth learning just on the basis that it has not been recorded yet? After all, in the era of Big Data, we are so full of information that we don't know what to do with them. This is felt acutely in science, where many surveys await someone making sense out of them. I think that "there is so much to know" is exactly the reason why, given the limited amount of time we spend on Earth, we should be careful in choosing what to learn.

I believe that, ultimately, all the reasons listed above are nothing more than mere "bonuses" on top of the real motivation for pursuing science. The real reason has been nicely summarized by Bohm (2004):

[...] what he [the scientist] is really seeking is to learn something new that has a certain fundamental kind of significance: a hitherto unknown lawfulness in the order of nature, which exhibits unity in a broad range of phenomena. Thus, he wishes to find in the reality in which he lives a certain oneness and totality, or wholeness, constituting a kind of harmony that is felt to be beautiful.

In this thesis, we have not achieved this, but at least we tried.

1.5 Overview of thesis

This thesis is composed of several related studies regarding the dynamics of gas in barred galaxies and in the Milky Way. In this section we briefly describe the content of each chapter. In addition to these, a final chapter that reviews our findings and puts them in a broader context is included at the end of the thesis.

1.5.1 Chapter 2: Observations of the Milky Way

The longitude-velocity distributions of HI and CO spectral lines contain an incredibly rich and diverse amount of information about our Galaxy. They exhibit structure on a wide range of scales, from clumpiness on small scales to coherent, broad features on large scales. This chapter has the purpose of reviewing the

observational data and to serve as a reference for later chapters. Since the observational aspects of most interest for this thesis are best understood as a departure from a simple axisymmetric model, we start by discussing such a simple model. Then we highlight its limitations and the lines of evidence that require us to move to more complex models that take into account non-axisymmetric gas flows. We consider in more detail the observations of the central disk, the region inside Galactocentric radius $R \simeq 3\,\mathrm{kpc}$. We discuss how to identify the envelope of the observations (the terminal velocity curve) and review the main large-scale features that have been identified in the literature, some of which are believed to correspond to spiral arms.

1.5.2 Chapter 3: Gas flow in barred potentials

In this chapter we begin our study of the physics of the gas flow in barred potentials. By using a Cartesian grid, we simulate the flow of gas in a particular barred Galactic potential and investigate the effects of varying the sound speed in the gas and the resolution of the grid. We investigate the relation between ballistic closed orbits and the results of full hydrodynamical calculations in more detail than previous authors did by quantitatively comparing the velocity field along each orbit and the corresponding hydrodynamic velocity field. We find that for all sound speeds and resolutions, streamlines closely follow closed orbits at large and small radii. At intermediate radii shocks arise and the streamlines shift between two families of closed orbits. An important new result is that the point at which the shocks appear and the streamlines shift between orbit families, which plays an important role in the interpretation of the data, depends strongly on the resolution. For sufficiently large values of the sound speed and the resolution, the transfer happens at the cusped orbit as hypothesised by Binney et al. over two decades ago. This is the first time the prediction of the Binney at al. model to explain the high velocity peaks found in the (l, v) distributions of HI and CO near the Galactic centre is supported by a full hydrodynamical simulation. Previous simulations did not study the relation between closed orbit and hydro simulations in sufficient detail to draw conclusions. For sufficiently high resolutions the flow downstream of the shocks becomes unsteady. If this unsteadiness is physical, as appears to be the case, we suggest that it provides a promising explanation for the asymmetry in the observed distribution of CO.

1.5.3 Chapter 4: Bar driven spiral arms

Spiral arms that emerge from the ends of a galactic bar are important in interpreting observations of our and external galaxies. It is therefore important to understand the physical mechanism that causes them. In this chapter we investigate such mechanisms. We find that these spiral arms can be understood as kinematic density waves generated by librations around the underlying ballistic closed orbits. This is even true in the case of a strong bar, provided the librations are around the appropriate closed orbits and not around the circular orbits that form the basis of the epicycle approximation. An important consequence is that it is a potential's orbital structure that determines whether a bar should be classified as weak or strong, and not crude estimates of the potential's deviation from axisymmetry. The criterion is that a bar should be

classified as weak when stable almost circular orbits (e.g., x_2 orbits or round x_1 orbits) exist at most radii, while it should be classified as strong when such orbits are not present or their presence has extended gaps. We show that potentials that have been considered as weak bars do not in fact possess almost circular orbits over extended ranges in radius. The relation of these spiral arms with observations of the Milky Way will be explored in Chapter 5.

1.5.4 Chapter 5: Effects of the quadrupole

We run hydrodynamical simulations of a 2D isothermal non self-gravitating inviscid gas flowing in a rigidly rotating externally imposed potential formed by only two components: a monopole and a quadrupole. We explore systematically the effects of varying the quadrupole while keeping fixed the monopole and discuss the consequences for the interpretation of longitude-velocity diagrams in the Milky Way. We find that the gas flow can constrain the quadrupole of the potential and the characteristics of the bar that generates it. The exponential scale length of the bar must be at least 1.5 kpc. The strength of the bar is also constrained. Our global interpretation favours a pattern speed of $\Omega_\mathrm{p} = 40\ \mathrm{km\,s^{-1}\,kpc^{-1}}$. We find that for most observational features, there exist a value of the parameters that matches each individual feature well, but is difficult to reproduce all the important features at once. Due to the intractably high number of parameters involved in the general problem, quantitative fitting methods that can run automatic searches in parameter space are necessary.

1.5.5 Chapter 6: Fitting dynamical models to the Milky Way

We present a new method for fitting simple hydrodynamical models to the (l, v) distribution of atomic and molecular gas observed in the Milky Way. The method works by matching features found in models and observations. It is based on the assumption that the large-scale features seen in (l, v) plots, such as ridgelines and the terminal velocity curve, are influenced primarily by the underlying large-scale Galactic potential and are only weakly dependent on local ISM heating and cooling processes. In our scheme one first identifies by hand the features in the observations: this only has to be done once. We describe a procedure for automatically extracting similar features from simple hydrodynamical models and quantifying the "distance" between each model's features and the observations. Application to models of the Galactic bar region ($|l| < 30°$) shows that our feature-fitting method performs better than χ^2 or envelope distances to identify the correct underlying galaxy model.

References

Alexander S., 1852, AJ, 2, 97
Athanassoula E., 1992, MNRAS, 259, 345
Baba J., Saitoh T. R., Wada K., 2010, PASJ, 62, 1413
Berendzen R., Hart R., Seeley D., 1984, Man discovers the galaxies.
Binney J., Gerhard O., Spergel D., 1997, MNRAS, 288, 365
Binney J., Gerhard O. E., Stark A. A., Bally J., Uchida K. I., 1991, MNRAS, 252, 210

Binney J., Merrifield M., 1998, Galactic Astronomy. Princeton University Press
Bissantz N., Englmaier P., Gerhard O., 2003, MNRAS, 340, 949
Blitz L., Spergel D. N., 1991, ApJ, 379, 631
Bohm D., 2004, On Creativity, On creativity No. v. 13. Routledge
Burton W. B., Elmegreen B. G., Genzel R., eds., 1992, Saas-Fee Advanced Course 21: The Galactic Interstellar Medium
Cowie L. L., 1980, ApJ, 236, 868
de Vaucouleurs G., 1964, in IAU Symposium, Vol. 20, The Galaxy and the Magellanic Clouds, Kerr F. J., ed., p. 195
Dwek E. et al., 1995, ApJ, 445, 716
Dyson J. E., Williams D. A., 1997, The physics of the interstellar medium. Bristol: Institute of Physics Publishing
Englmaier P., Gerhard O., 1999, MNRAS, 304, 512
Ferrière K., Gillard W., Jean P., 2007, A & A, 467, 611
Feynman R. P., 1988, What do you care what other people think? Further adventures of a curious character as told to Ralph Leighton. Penguin
Fux R., 1999, A & A, 345, 787
Fux R., 2004, in Astrophysics and Space Science Library, Vol. 315, How Does the Galaxy Work?, Alfaro E. J., Pérez E., Franco J., eds., p. 213
GellMann M., 2001, Interview on the different interpretations of quantum mechanics (`http://www.webofstories.com/play/murray.gell-mann/1`)
Gerhard O., 2002, in Astronomical Society of the Pacific Conference Series, Vol. 273, The Dynamics, Structure & History of Galaxies: A Workshop in Honour of Professor Ken Freeman, Da Costa G. S., Sadler E. M., Jerjen H., eds., p. 73
Gerhard O., Wegg C., 2014, in Lessons from the Local Group, eds. Freeman, K. C., Elmegreen, B. G., Block, D. L., & Woolway, M. (Springer; New York), in press (arXiv:1408.0219)
Gerhard O. E., Vietri M., 1986, MNRAS, 223, 377
Jenkins A., Binney J., 1994, MNRAS, 270, 703
Klessen R. S., Glover S. C. O., 2014, ArXiv e-prints 1412.5182
Landau L. D., Lifshitz E. M., 1987, Fluid Mechanics, Vol. 6 (Course of Theoretical Physics), 2nd edn. Butterworth-Heinemann
Lee C. W., Lee H. M., Ann H. B., Kwon K. H., 1999, ApJ, 513, 242
Lin C. C., Roberts, Jr. W. W., 1981, Annual Review of Fluid Mechanics, 13, 33
Liszt H. S., Burton W. B., 1980, ApJ, 236, 779
Merrifield M. R., 2004, in Astronomical Society of the Pacific Conference Series, Vol. 317, Milky Way Surveys: The Structure and Evolution of our Galaxy, Clemens D., Shah R., Brainerd T., eds., p. 289
Mulder W. A., Liem B. T., 1986, A & A, 157, 148
Oort J. H., 1952, ApJ, 116, 233
Oort J. H., 1977a, Comments on Astrophysics, 7, 51
Oort J. H., 1977b, Annual Review of Astronomy and Astrophysics, 15, 295
Oort J. H., Kerr F. J., Westerhout G., 1958, MNRAS, 118, 379
Oort J. H., Rougoor G. W., 1959, AJ, 64, 130
Peters, III W. L., 1975, ApJ, 195, 617
Pettitt A. R., Dobbs C. L., Acreman D. M., Price D. J., 2014, MNRAS, 444, 919
Popowski P. et al., 2005, ApJ, 631, 879
Prendergast K. H., 1983, in IAU Symposium, Vol. 100, Internal Kinematics and Dynamics of Galaxies, Athanassoula E., ed., pp. 215–220
Roberts W. W., 1969, ApJ, 158, 123
Roberts W. W., 1971, in Bulletin of the American Astronomical Society, Vol. 3, Bulletin of the American Astronomical Society, p. 369
Roberts, Jr. W. W., 1979, in IAU Symposium, Vol. 84, The Large-Scale Characteristics of the Galaxy, Burton W. B., ed., pp. 175–184
Rodriguez-Fernandez N. J., Combes F., 2008, A & A, 489, 115
Sellwood J. A., Wilkinson A., 1993, Reports on Progress in Physics, 56, 173
Shore S. N., 2002, The Tapestry of Modern Astrophysics. Wiley-VCH
Shore S. N., 2007, Astrophysical Hydrodynamics: An Introduction. Wiley-VCH
Shu F. H., 1992, Physics of Astrophysics, Vol. II. University Science Books

Skrutskie M. F. et al., 2006, AJ, 131, 1163
Spitzer L., 1998, Physical Processes in the Interstellar Medium. Wiley-VCH
Stanek K. Z., Mateo M., Udalski A., Szymanski M., Kaluzny J., Kubiak M., 1994, ApJletters, 429, L73
Wegg C., Gerhard O., 2013, MNRAS, 435, 1874
Wegg C., Gerhard O., Portail M., 2015, MNRAS, 450, 4050
Weiland J. L. et al., 1994, ApJL, 425, L81
Weiner B. J., Sellwood J. A., 1999, ApJ, 524, 112
Whitney C. A., 1971, The discovery of our galaxy. Random House Inc (T)
Wyrzykowski Ł. et al., 2015, ApJS, 216, 12
Zel'dovich Y. B., Raizer Y. P., 1967, Physics of shock waves and high-temperature hydrodynamic phenomena. New York: Academic Press
Zhao H., Mao S., 1996, MNRAS, 283, 1197

CHAPTER 2

Observations of the Milky Way

In this chapter we introduce the Milky Way observational data that will be used in the remainder of the thesis. More comprehensive introductions can be found for example in Burton et al. (1992) and Binney & Merrifield (1998).

Spectral line observations are usually given as a *data cube*. For each position in the sky we measure the intensity of radiation as a function of the frequency. If the emission comes from a spectral line whose rest frequency is known, from the observed frequency it is possible to infer the line-of-sight velocity v of the emitting material. The intensity at each particular frequency will depend on the total amount of gas along the line of sight which is moving at v, the velocity corresponding to that frequency. Positions in the sky are labelled by two numbers $\{l, b\}$, where l and b are the Galactic longitude and latitude respectively.[1] The data consist in the measured line intensity for each triad $\{l, b, v\}$, hence the name data cube.

Since most of the gas in the Galaxy is confined to a thin disk, it is often convenient to average the data over some range in b. The resulting 2D maps show the line intensity in the longitude-velocity plane and are called (l, v) plots. These constitute one of the main tools to compare dynamical models to the observations. We need to learn how to interpret these plots. The upper and middle panels in Fig. 2.1 show the longitude-velocity plots for two of the most extensively studied lines: the HI 21-cm line, emitted by cold or mildly warm neutral gas, and the ^{12}CO 2.6 mm $J = 1 \rightarrow 0$ line, emitted by very dense and cold molecular gas. These plots show the intensity of the emission (i.e., the brightness temperature $T_{\rm B}$) across the whole

[1] The (l, b) coordinate system was defined in the late 1950s. The very flat distribution of 21cm HI emission was used to identify the mid-plane of the Galaxy and the position of the Galactic centre in the sky. The center determined in this way coincided well, within uncertainties of the time, with the peculiar radio source Sgr A. The center was thus defined to be at the intersection of the mid-plane determined from hydrogen and Sgr A. This radio source was later resolved into sub-sources, Sgr A East, Sgr A West and Sgr A*, and the Galactic center is currently assumed to coincide with the latter, although more precise determinations have found $l = 0^\circ$ and $b = 0^\circ$ to be slightly offset from the position of Sgr A*. Sgr A* is a compact radio source which is now believed to mark the position of the central supermassive black hole. However, it is possible that the dynamical center of the Milky Way does not coincide with Sgr A* (e.g. Blitz, 1994). It is also possible that the nuclear x_2 disk of the Milky Way (see Chapter 3) is oscillating around the center determined from the disk and spheroid stars (e.g. Fux, 1999). An interesting historical account of how the (l, v) coordinate system was defined and how we came to identify Sgr A* as the Galactic center can be found in Blitz (1994).

Galactic plane, averaged over $|b| \leqslant 4^\circ$.

The lower panel in Fig. 2.1 shows the prediction of a simple axisymmetric model. The observational aspects that are believed to be signatures of the bar are best understood as departures from the predictions of such a simple model. For this reason, in this chapter we parallel our description of the data with that of this simple model.

2.1 A simple axisymmetric model

What would we observe in the (l, v) plane if all the gas were in purely circular motion? The bottom panel in Fig. 2.1 answers this question by showing a simple model in which all gas is assumed to move on circular orbits with a completely flat rotation curve locked at $v_c = 220\,\mathrm{km\,s^{-1}}$. The Sun is assumed to be located at $R_0 = 8\,\mathrm{kpc}$ from the Galactic centre and moving on a circular orbit at the same speed v_c. The gas is distributed with uniform density throughout the plane, and is projected to the longitude-velocity plane using a simple projection law that will be described later in Sect. 3.2.

It is clear from Fig. 2.1 that, despite its simplicity, this model allows us to broadly understand observations. The overall shape reproduces the observations, with the envelope declining from the peaks near $l = 0$ to zero near $l = 90^\circ$. Several aspects of the intensity structure are also roughly captured by this simple model. Fig. 2.2 shows cuts of the plots in Fig. 2.1 through the line of sight at $l = 30^\circ$. Near $v = 0$, the intensity in both the data and the model jumps. This can also be seen directly from the 2D maps of Fig. 2.1. The simple model captures this characteristic. Another characteristic that can be seen both in HI data and in the model, although more difficult to see in the CO data, is that the intensity is higher at the envelope of the emission, i.e. where the minimum and maximum velocity at which emission is present are attained for a given longitude. In the 2D (l, v) plane this results in the envelope of the emission being bright, as if its borders were traced with a pen.

In the simple model, the gas density is uniform in the Galactic plane. Hence, any intensity variation in the model (l, v) plot is entirely due to projection effects. In general there are two reasons why a (l, v) region can be particularly bright in a model: (i) it reflects some real overdensity in the (x, y) distribution of the gas; (ii) it is the projection of a region where $\mathrm{d}v(r)/\mathrm{d}r$ is small, and so large portions of the xy plane get projected to narrow regions in the (l, v) plane. Here, r is the line-of-sight distance and v is the component of the velocity along the line of sight. Of course, it is also possible that many separated regions in the plane all possess the same v and so get projected to the same part of the (l, v) plane, increasing the brightness. When (ii) happens people speak of *velocity crowding*. One of the main difficulties in interpreting the data comes from the fact that it is very difficult to disentangle the two effects in general.

Why does our model reproduce the enhanced intensity at the envelope of the emission and the $v = 0$ jump? The envelope of the (l, v) emission coincides with regions where $\mathrm{d}v(r)/\mathrm{d}r = 0$ and therefore is bright to due velocity crowding. The value of $\mathrm{d}v(r)/\mathrm{d}r$ does not explain the jump at $v = 0$. Simple geometrical considerations show (see for example Binney & Merrifield, 1998) that on one side of the $v = 0$ line, two

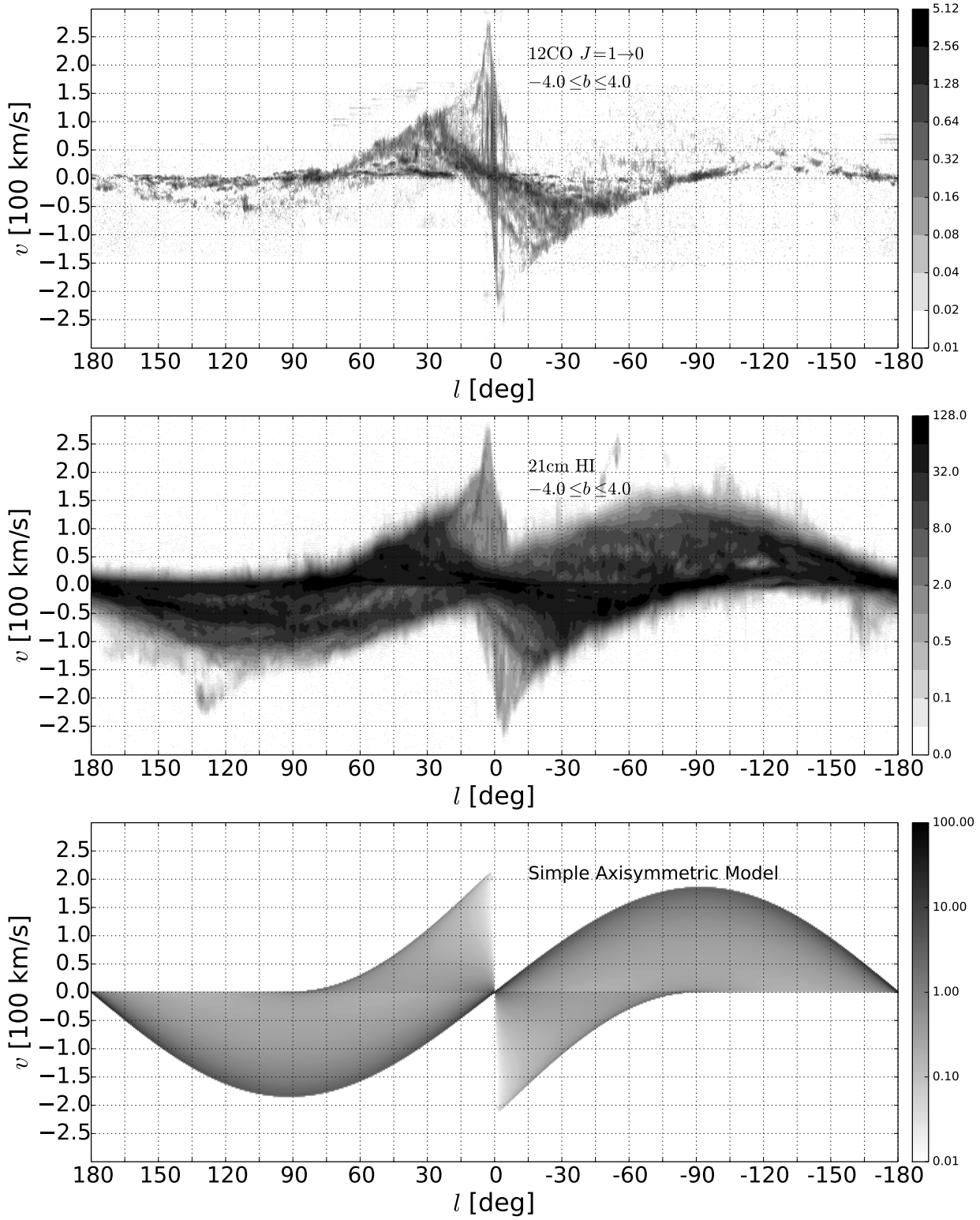


Figure 2.1: Upper panel: CO observations over the whole Galactic plane integrated over $|b| \leqslant 4°$. Middle panel: HI observations integrated over $|b| \leqslant 4°$. Lower panel: (l, v) plot for a simple model with gas moving on purely circular orbits. The rotation curve is assumed to be completely flat and the density distribution uniform throughout the plane. The colorbar is in units of K for upper and middle panel, while is in arbitrary units for the simple model. Data from Dame et al. (2001); Kalberla et al. (2005).

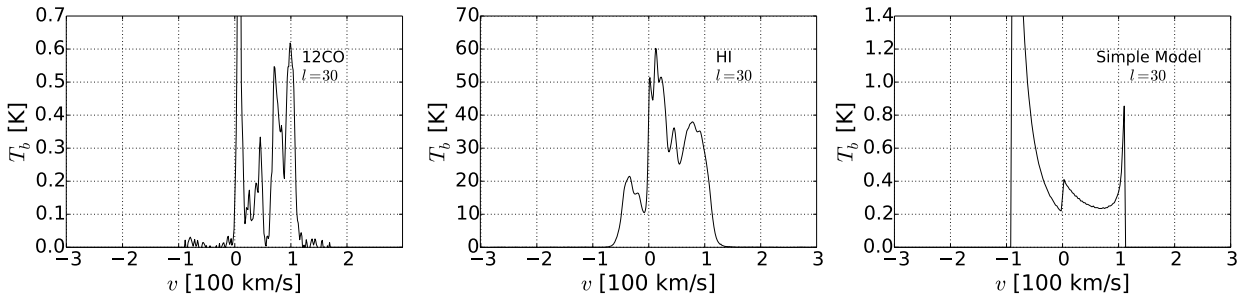


Figure 2.2: Cuts along the line of sight $l = 30°$ of the plots in Fig. 2.1.

identical but separate points along the line of sight contribute to the emission, while on the other side only one point does so. This is because the line $v = 0$ corresponds to the Solar circle in the xy plane, and lines of sight intersect a circle smaller than the Solar circle twice but a circle bigger than it only once. These geometrical considerations explain the jump at $v = 0$.

2.2 Signatures of the bar and limitations of the simple model

Fig. 2.3 shows in more detail the data in the region $|l| \leqslant 30°$, which corresponds to the region of the Galaxy inside Galactocentric radius $R \simeq 3\,\mathrm{kpc}$. Panels (a) and (b) are zooms in the central region of the upper and middle panel of Fig. 2.1 respectively. In panel (c) the grey scale shows CO data while coloured lines show features identifiable in the (l, v) distributions of CO and HI. These features will be discussed in some detail below. In panel (d) the grey scale shows HI data smoothed in velocity by convolving with a Gaussian with $\sigma = 5\,\mathrm{km\,s^{-1}}$. Contours are also shown. This will be used in Sect. 2.3.1 to determine the envelope of the emission, which is shown in panel (c) and (d) with red dots. Figs. 2.4 and 2.5 show slices at different latitudes of CO and HI data.

The simple model discussed in the previous section has many limitations, many of which are signatures of non-axisymmetric motions. Below we list some of these, but our discussion is not exhaustive and is biased by our focus on the central region. For a more complete account see for example Burton et al. (1992) and Binney & Merrifield (1998).

1. *Emission at forbidden velocities.* In the simple model, gas inside the Solar circle cannot produce emission in the two quadrants $(l > 0, v < 0)$ and $(l < 0, v > 0)$. All emission in these two quadrants comes from outside the Solar circle in the simple model. The same is true for any axisymmetric model[2] and comes from simple geometrical considerations (e.g. Binney & Merrifield, 1998, Chapter 9). However, it has long been known that observations show emission in these two quadrants that must come from material in the inner Galaxy. Such observed emission cannot be reproduced by any axisymmetric model, and for this reason is said to be at "forbidden velocities". For example, practically all emission at $|l| < 5°$, $v > 100\,\mathrm{km\,s^{-1}}$ visible in the data shown in Fig. 2.1 comes from inside the Solar circle (but the forbidden emission extends also up to $|l| \simeq 8°$). While forbidden emission cannot

[2]Unless assuming counter-rotating material; see Gerhard & Vietri (1986)

be explained by axisymmetric models, the current view is that it should be understood in terms of non-axisymmetric flows caused by the central bar (e.g. Burton et al., 1992; Binney & Merrifield, 1998). Other interpretations put forward in the early days, for example involving explosive phenomena (Oort, 1977), are now considered very unlikely.

2. *Velocity peaks.* The observations show material at very high velocities at $|v| \simeq 270\,\mathrm{km\,s^{-1}}$ in the vicinity of the Galactic centre $|l| \simeq 2^\circ$. These velocity peaks are not reproduced in the simple model, which reaches only $v \lesssim 220\,\mathrm{km\,s^{-1}}$. It is in principle possible to reproduce these peaks using an axisymmetric model, but it would imply that the circular velocity curve of the Galaxy has a high velocity peak around $R \simeq 500\,\mathrm{pc}$. This peak is believed to be an artefact of interpreting observations in terms of purely circular motion, and should not be present if we were to plot the circular velocity curve corresponding to the monopole component of the real potential in the plane of the Galaxy (e.g. Binney et al., 1991; Burton et al., 1992). In fact, the mass contained in the central parts of the Galaxy derived assuming circular motion represents a generous upper bound for the real mass of the Galaxy (Mulder & Liem, 1986; Burton & Liszt, 1978). The peaks are now generally interpreted in terms of non-axisymmetric gas streaming along the bar, although other interpretations are possible (e.g. Oort, 1977; Rodriguez-Fernandez & Combes, 2008).

3. *Internal features.* It is apparent from Figs. 2.1 and 2.3 that the internal structure of the real observations is very rich, and in particular it displays broad coherent features, while by contrast the simple model is very flat and featureless. Most internal features consist of bright ridges and often bear a name that can be confusing, as it refers to an original interpretation which may no longer be accurate. The main features that have been identified in the literature are shown schematically as dashed lines in Fig. 2.3, panel (c), and these will be discussed below in Sect. 2.3.2. More comprehensive lists of features can be found for example in Rougoor (1964), van der Kruit (1970), Cohen (1975), Bania (1977), Bally et al. (1988), Dame et al. (2001) and Dame & Thaddeus (2008). It is unlikely that most of these features can be reproduced in a model which assumes an axisymmetric gas density distribution. Can such features be reproduced in a model that assumes purely circular motion, but in which the density is not arranged axisymmetrically? Strictly speaking for most of them yes, as it is possible to arrange the xy density distribution as to reproduce any (l, v) distribution (except in the forbidden velocity region). In fact, under the assumption of purely circular motion it is possible to deduce uniquely a circular rotation curve (e.g. Binney & Merrifield, 1998, Chapter 9) from the envelope of the observations in the non forbidden quadrants. Once this is determined the xy densities can be rearranged as to reproduce the internal part of the (l, v) observations in an almost unique way. Since this deprojection is almost unique, a density map of the Milky Way can be created in this way. Indeed, this is how early maps of the Milky Way were produced (e.g. Oort et al., 1958). However, especially inside Galactocentric radius $R \simeq 3\,\mathrm{kpc}$, this is unlikely to be accurate as we know that strong circular motions are present. Velocity crowding is often more important than the real density structure in determining the (l, v)

plane intensities (e.g. Mulder & Liem, 1986; Burton et al., 1992, see also Chap. 6), although it should be noted that overdense regions, for example spiral arms, are usually also associated with velocity crowding. The final picture is complex and it is likely that most of these features should be understood as a combination of both effects. The best way to model such features is to obtain them as a result of full hydrodynamical calculations.

4. *The Envelope.* The red dots in Fig. 2.3, panel (c) and (d), trace the envelope of the observations, determined as explained below in Sect. 2.3.1. This differs from the envelope of the simple model in three main ways: through the velocity peaks, the emission at forbidden velocities and the shoulders and bumps that arise where an internal feature touches the envelope. Reconstructing the envelope using axisymmetric models is likely to produce serious mistakes for the reasons explained in items (2) and (3). The best way to model the envelope, similarly to internal features, is to use the results of full hydrodynamical calculations.

5. *Bumps on the Envelope.* When an internal feature touches the envelope, a bump is (usually) present in the observations. A mechanism by which spiral arms generate bumps on the envelope is described in §9.1.2 of Binney & Merrifield (1998). These bumps are obviously not present in the simple model. We identify as particularly interesting the *knee* marked with a grey circle in Fig. 2.3, panel (c). This knee occurs where the envelope shifts from touching the arm at *135-km/s* to touching the *connecting arm* (green dashed and light red dashed curves in Fig. 2.3). Other interesting bumps occur when the *3* kpc *arm* and its far-side counterpart touch the envelope.

6. *Tilt.* The central few kiloparsecs of the gas disc appear to be tilted with respect to the Galactic plane such that the far end of the bar lies above, and the near end below, the plane (Liszt & Burton, 1980; Burton et al., 1992). The dynamical reasons for the tilt are at present unknown. It is worth noting that the structure of the knee changes as we shift in b. In Figs. 2.1 and 2.3 the data are integrated over a range of latitudes b, but if we study slices at different latitudes (Figs. 2.4 and 2.5) we see that the connecting arm, which contributes to the envelope on the low-l side of the knee, appears only at $b < 0$, while the arm at $135\,\mathrm{km\,s^{-1}}$, which forms the envelope on the other side, appears only at $b \geqslant 0$. Since the connecting arm and the arm at $135\,\mathrm{km\,s^{-1}}$ appear at distinct latitudes, they should represent distinct dynamical features. Thus the knee, which arises from the transition between these structures, cannot reflect merely a sudden change in the circular speed.

7. *Clumpiness.* In addition to coherent features discussed in item (3), observations also show scattered clumpiness on a variety of scales. This is obviously not present in the simple model. This type of (l, v) clumpiness is believed to reflect clumpiness in real xy space. It has been modeled by Baba et al. (2010) as caused by heating and cooling processes. In general, models that do not include such effects can contain large-scale broad features, but are smooth on finer scales. Therefore, inclusion of heating and cooling effects is most likely to be necessary to explain clumpiness in detail.

8. *Asymmetry.* Approximately three-quarters of the molecular emission from $|l| \lesssim 4^\circ$ comes from positive longitudes (e.g. Burton et al., 1992). The cause of this asymmetry is a long-standing puzzle. The asymmetry is too big to be attributed solely to a perspective effect from an inclined bar (Jenkins & Binney, 1994). The only promising explanation currently available is that the asymmetry is generated by fluctuations in an unsteady flow (Chap. 3). However this idea remains embryonic and requires further investigation.

9. *Variation by species.* Different chemical species such as HI, CO, CS, etc., are distributed differently throughout the Galaxy because they probe different temperature and density environments. Consequently, each produces a different (l, v) diagram. A complete model would explain the variations in (l, v) plots. A related problem is the extent to which gas and dust are correlated (see for example Sale & Magorrian, 2014).

To improve from the simple model we must account for at least some of the items above. Progress has been made in several directions, but is difficult to account for many different aspects at the same time. For example, Mulder & Liem (1986), Englmaier & Gerhard (1999), Bissantz et al. (2003), Rodriguez-Fernandez & Combes (2008) and Pettitt et al. (2014) each reproduced some of the internal features (item 3), but none of them was able to account for the velocity peaks in a simulation (item 2). Weiner & Sellwood (1999) and Binney et al. (1991) where able to account for the velocity peaks, but not for any of the internal features (see also our model in Chapter 3). Forbidden emission has been partly reproduced by many models, but few cover the whole portion of the (l, v) plane where forbidden emission is present. The remarkable simulations of Fux (1999) were able to reproduce and give a coherent interpretation to many features and also reproduce the velocity peaks. However his simulations show transient behaviours and are very difficult to reproduce. It is very difficult to use them to constrain the properties of the bar. Baba et al. (2010) modeled the clumpiness, but they did not try to fit the Milky Way. The asymmetry has not yet a dynamical explanation, with the exception of the embryonic idea presented later in Chapter 3. Eventually, we hope to converge towards a megamodel of the Milky Way that can explain all the items above.

2.3 Envelope and internal features

In this section we analyse in more detail items (2) and (3) of the previous section.

2.3.1 Envelope

The envelope is the most obvious feature that has been used to compare dynamical models to the data. The envelope is most easily identified from HI observations, since HI represents a more diffuse and pervasive component than CO. To extract the envelope of the (l, v) plot that we will use for our fitting methods in Chapter 3 we use the HI brightness temperature measured by Kalberla et al. (2005), which has a spatial resolution of 0.5°. We average over $|b| \leqslant 4^\circ$ and convolve with a Gaussian of $\sigma = 5\ \mathrm{km\,s^{-1}}$ in velocity to

obtain the smoothed temperature map, $T_{\rm B}(l,v)$, shown in Fig. 2.3(d). For each l in the range $-5^\circ \leqslant l \leqslant 30^\circ$ we determine the upper envelope, $v_+^D(l)$, (i.e., the most positive line-of-sight velocity for each l) by examining $T_{\rm B}(v) \equiv T_{\rm B}(l,v)$ as follows.

1. Given l, find the velocities v at which $\partial T_{\rm B}(v)/\partial v$ peaks. This amounts to finding points of high gradient, which is a general definition of an edge. To avoid fitting noise, discard any edge that has $T_{\rm B}(v) < 0.125\,{\rm K}$. This is slightly larger than the RMS noise in $T_{\rm B}$ quoted by Kalberla et al. (2005). Let $v_{\rm pk}$ be the location of the highest velocity edge that remains for this l.

2. Following Shane & Bieger-Smith (1966), assign

$$v_+^D = v_{\rm pk} + \frac{1}{T_{\rm B}(v_{\rm pk})} \int_{v_{\rm pk}}^{\infty} T_{\rm B}(v)\,{\rm d}v\,. \tag{2.1}$$

   As this makes use of the full profile $T_{\rm B}(v)$, it has the advantage over other methods of being robust against systematic effects due to noise in $T_{\rm B}(v)$.

For $l < 0$, one expects that most of the $v < 0$ emission will be caused by foreground material well outside the bar (see, e.g., Figure 9.3 of Binney & Merrifield, 1998). As our interest is restricted to the bar region, after visual inspection of the CO data, we define the envelope to fall linearly from its value found using the method above at $l = -5^\circ$ to zero at $l = -12^\circ$.

The lower envelope (the most negative line-of-sight velocities), $v_-^D(l)$, is obtained in the same way, but reflecting $l \to -l$ and $v \to -v$. We make an exception for the five points $l = \{0.5,^\circ, 1.0^\circ, 1.5^\circ, 2^\circ, 2.5^\circ\}$, which we corrected manually to account for extra absorption features (the role of absorption in determining the paucity of HI emission in this region of the l-v plane is discussed in detail in Section 6.2 of Burton et al. 1992). The final result, superimposed on HI and CO data, is shown in Fig. 2.3 panels (d) and (c).

2.3.2 Internal features

The identification of large-scale features constitute the basis of the fitting method that will be presented in Chap 6. Indeed, many internal features can be identified in the (l,v) diagrams in the region $|l| < 30^\circ$. Most of them consist of bright ridges and often bear a name that can be confusing, as it refers to an original, now discredited, interpretation which may no longer be accurate. We will give our own interpretation to many features in Chap. 5. Lists of features can be found for example in Rougoor (1964), van der Kruit (1970), Cohen (1975), Bania (1977), Bally et al. (1988) and Fux (1999). Here, our goal is to focus on those that are most likely to trace the large-scale gravitational potential of the Galaxy, avoiding any that are most likely due to stellar feedback processes. Possible candidates, shown in Fig. 2.3, are the following features:

- *3-kpc arm.* This is the most apparent and coherent feature. It is a bright ridge that can be traced over a large range in longitude and crosses $l = 0$ with a velocity of $-53\ {\rm km\,s^{-1}}$. Absorption against radio continuum emission from the Galactic centre shows that it lies in front of the Galactic centre. It is probably associated with a spiral arm.

- *Far side 3-kpc arm.* This recently discovered feature (Dame & Thaddeus, 2008) lies beyond the Galactic centre and is thought to be the far-side counterpart of the 3-kpc arm, a role that was sometimes previously assigned to the 135 $\mathrm{km\,s^{-1}}$ arm. It can be seen clearly only in sufficiently high-resolution data, so it is not identifiable in our figures.

- *135 km s^{-1} arm.* This is a high-velocity arm that crosses $l = 0$ at a velocity of 135 $\mathrm{km\,s^{-1}}$. As it is not seen in absorption against Sgr A, it is most probably caused by gas that lies beyond the Galactic centre (e.g., Cohen, 1975).

- *Connecting arm.* This very bright feature lies at very high velocity and touches the positive-velocity peak emission. There is no unanimous consensus on whether it corresponds to a real arm. It has been associated with a dust lane (Marshall et al., 2009; Liszt, 2008). An alternative interpretation considers it the edge of the nuclear x_2 ring (Liszt & Burton, 1980). In the interpretation of Marshall et al. (2009) it lies in front of the Galactic Center. In Chap. 5, we interpret it as a spiral arm running alongside the bar.

- *Central molecular zone.* This refers to the off-centered concentration of dense molecular gas in the region $-1^\circ \lesssim l \lesssim 1.5^\circ$. The parallelogram bounding this structure has received particular attention (Binney et al., 1991). In Chap. 3, we will interpret two sides of the parallelogram as corresponding to the two straight shocks that connect the end of the bar with the inner nuclear disk.

- *Molecular ring.* Despite being one of the most prominent structures in the CO data, its structure is still debated. It could be either a ring-like structure, possibly associated with a resonance, or an intertwined structure of spiral arms. Sub-branches and bifurcations can be observed near the edges.

- *Vertical features.* We use this umbrella term to denote various features that span a large velocity range while being confined in a small longitude range. The vertical feature at $l \simeq 3^\circ$ is called *Clump2* after Stark & Bania (1986). These authors related it with an inner dust lane or spiral arm. Marshall et al. (2009) interpreted the vertical features at negative longitudes as the far-side counterpart of the connecting arm. Some of these vertical features might be due to magneto-hydrodynamic effects (Machida et al., 2009). In Chapter 5 we propose that all internal features are different portions of the straight shocks.

Most interesting, *all* the features listed above can be identified both in HI and CO data, though some of them (central molecular zone and Clump2) are notably fainter in HI. All of them are candidates for tracing the large-scale dynamics of the Milky Way. Some of them, like the molecular ring, the 3-kpc arm and its far side counterpart, are most likely to be independent of small scale physics, whereas others are more dubious, such as those that have been associated with dust lanes like the connecting arm and the vertical features.

References

Baba J., Saitoh T. R., Wada K., 2010, PASJ, 62, 1413
Bally J., Stark A. A., Wilson R. W., Henkel C., 1988, ApJ, 324, 223
Bania T. M., 1977, ApJ, 216, 381
Binney J., Gerhard O. E., Stark A. A., Bally J., Uchida K. I., 1991, MNRAS, 252, 210
Binney J., Merrifield M., 1998, Galactic Astronomy. Princeton University Press
Bissantz N., Englmaier P., Gerhard O., 2003, MNRAS, 340, 949
Blitz L., 1994, in Astronomical Society of the Pacific Conference Series, Vol. 66, Physics of the Gaseous and Stellar Disks of the Galaxy, King I. R., ed., p. 1
Burton W. B., Elmegreen B. G., Genzel R., eds., 1992, Saas-Fee Advanced Course 21: The Galactic Interstellar Medium
Burton W. B., Liszt H. S., 1978, ApJ, 225, 815
Cohen R. J., 1975, MNRAS, 171, 659
Dame T. M., Hartmann D., Thaddeus P., 2001, ApJ, 547, 792
Dame T. M., Thaddeus P., 2008, ApJ, 683, L143
Englmaier P., Gerhard O., 1999, MNRAS, 304, 512
Fux R., 1999, A & A, 345, 787
Gerhard O. E., Vietri M., 1986, MNRAS, 223, 377
Jenkins A., Binney J., 1994, MNRAS, 270, 703
Kalberla P. M. W., Burton W. B., Hartmann D., Arnal E. M., Bajaja E., Morras R., Pöppel W. G. L., 2005, A & A, 440, 775
Liszt H. S., 2008, A & A, 486, 467
Liszt H. S., Burton W. B., 1980, ApJ, 236, 779
Machida M. et al., 2009, PASJ, 61, 411
Marshall D. J., Fux R., Robin A. C., Reylé C., 2009, in The Evolving ISM in the Milky Way and Nearby Galaxies
Mulder W. A., Liem B. T., 1986, A & A, 157, 148
Oort J. H., 1977, Comments on Astrophysics, 7, 51
Oort J. H., Kerr F. J., Westerhout G., 1958, MNRAS, 118, 379
Pettitt A. R., Dobbs C. L., Acreman D. M., Price D. J., 2014, MNRAS, 444, 919
Rodriguez-Fernandez N. J., Combes F., 2008, A & A, 489, 115
Rougoor G. W., 1964, Bulletin of the Astronomical Institutes of the Netherlands, 17, 381
Sale S. E., Magorrian J., 2014, MNRAS, 445, 256
Shane W. W., Bieger-Smith G. P., 1966, Bulletin of the Astronomical Institutes of the Netherlands, 18, 263
Stark A. A., Bania T. M., 1986, ApJletters, 306, L17
van der Kruit P. C., 1970, A & A, 4, 462
Weiner B. J., Sellwood J. A., 1999, ApJ, 524, 112

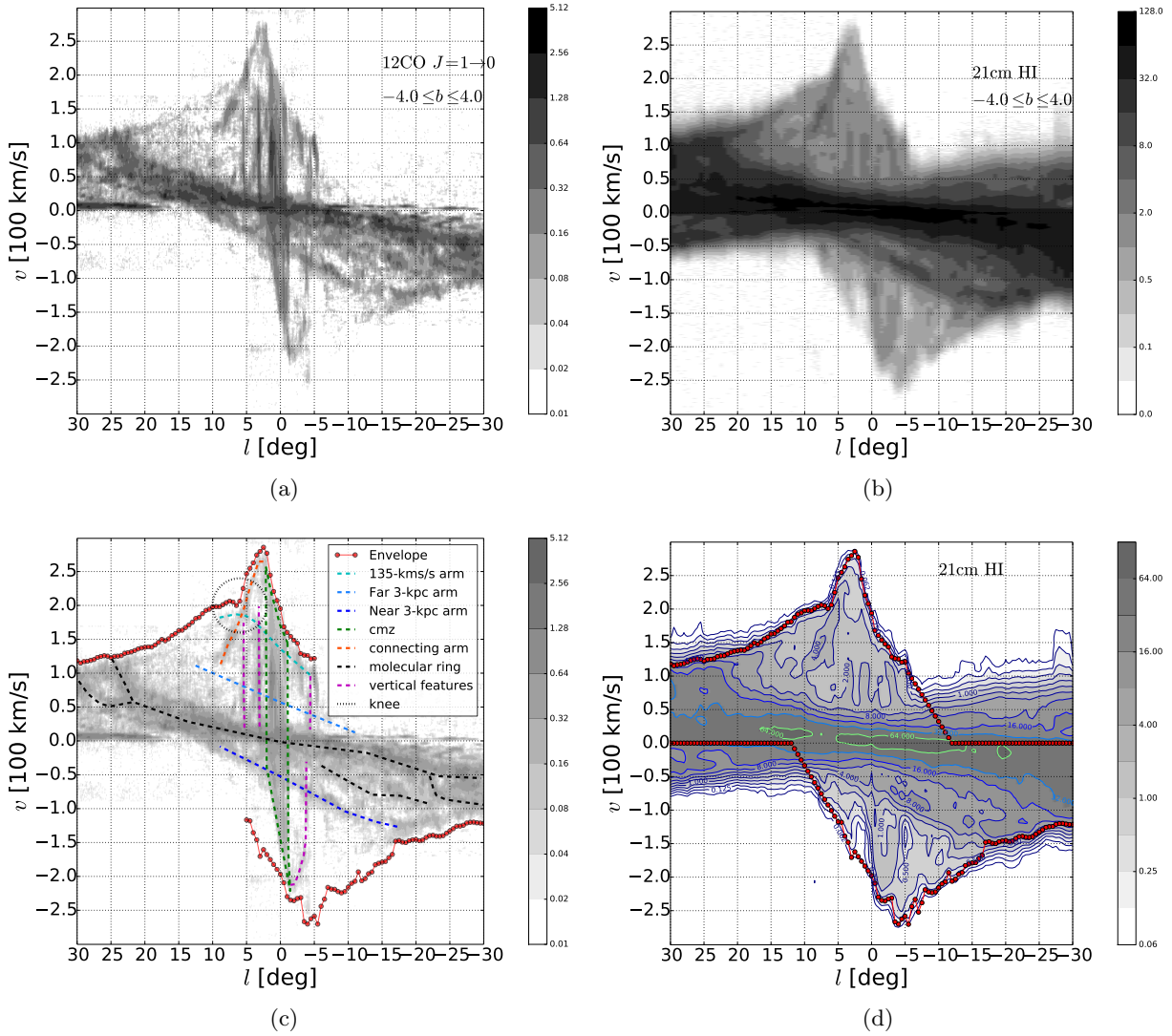


Figure 2.3: A closer look at the region $l \leqslant 30°$ of the data in Fig. 2.1. (a) CO observations integrated over $|b| \leqslant 4°$. (b) HI observations integrated over $|b| \leqslant 4°$. (c) CO observations with the envelope (determined from HI data) and internal features superimposed. (d) HI observations with superimposed envelope and brightness temperature contours indicated. Contours are spaced by factors of 2 in T_{B}. Colorbar is in units of K.

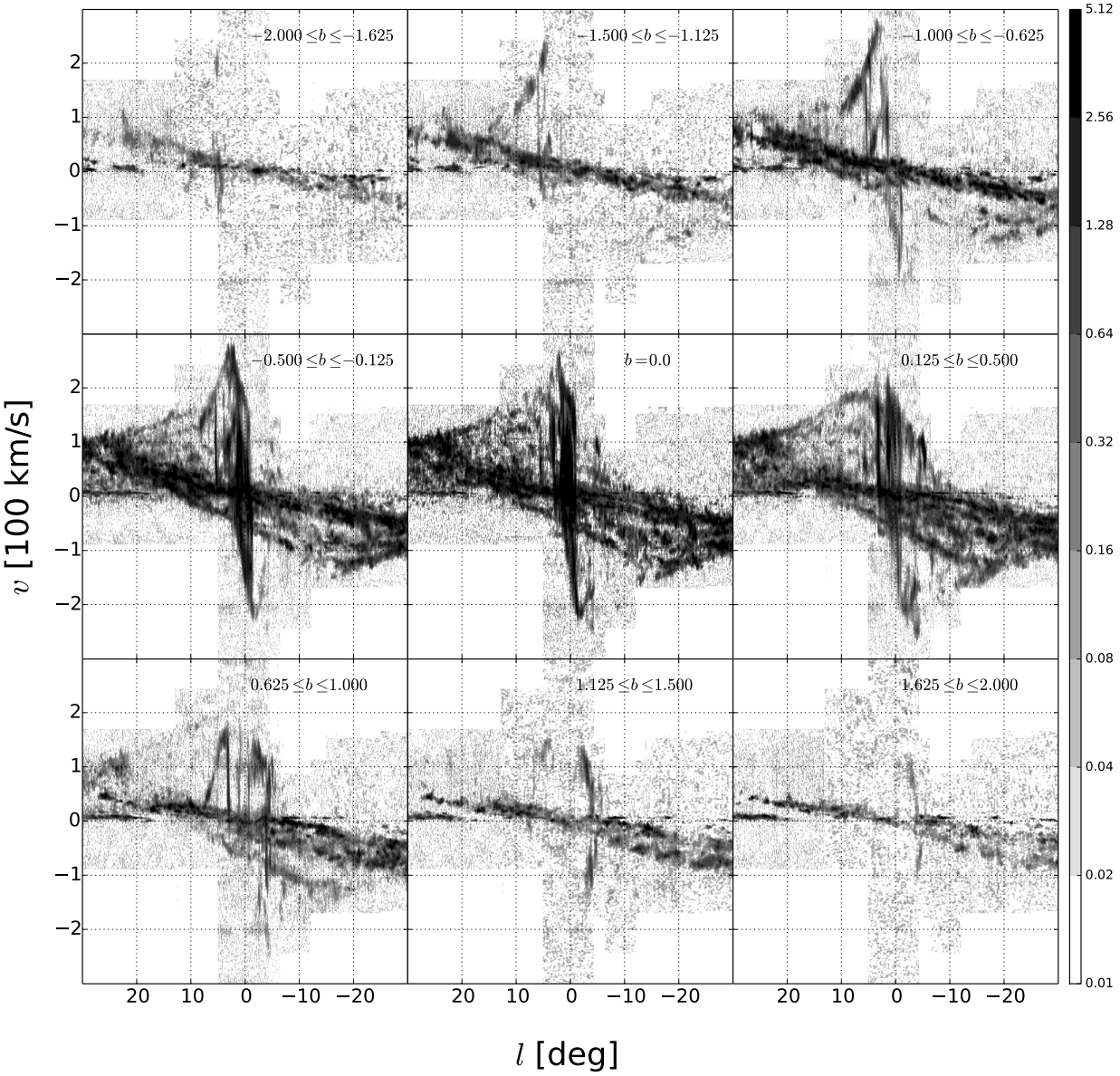


Figure 2.4: Slices at different latitudes b of the CO data in Fig. 2.3. Colorbar is in units of K.

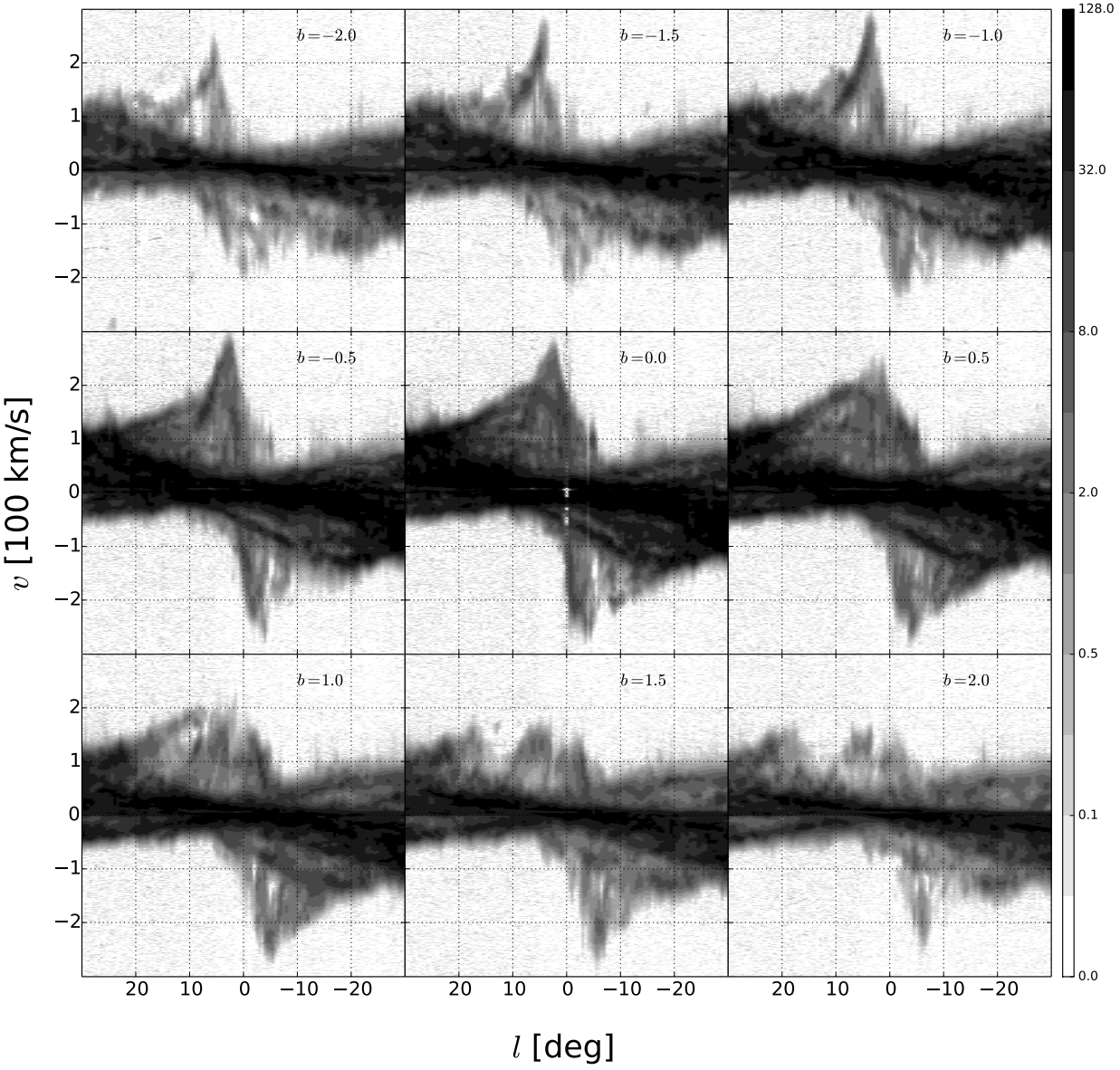


Figure 2.5: Slices at different latitudes b of the HI data in Fig. 2.3. Colorbar is in units of K.

CHAPTER 3

Gas Flow in Barred Potentials

3.1 Introduction

More than twenty years ago, Binney et al. (1991) (hereafter BGSBU) proposed a consistent picture to explain spectral line emission by different species, HI, CO and CS, in the Galactic-centre region $|l| < 10^\circ$ and $|b| < 2^\circ$. They studied the flow of gas in a particular externally imposed, rigidly rotating barred potentials and related it to the structure of the longitude-velocity, (l, v), plane that one obtains by projecting ballistic closed orbits along lines of sight through the disc. According to BGSBU, gas in the outer parts of the bar drifts slowly towards the centre while following x_1 orbits. These orbits become more and more elongated as the centre is approached, and eventually the "cusped orbit" is reached, which has a cusp at each end (see Sect. 3.3 and in particular Fig. 3.1). Interior to the cusped orbit, the orbits of the x_1 family are self-intersecting. BGSBU hypothesised that when gas reaches the cusped orbit, it encounters a shock, and then quickly plunges onto x_2 orbits. Thereafter, the gas drifts towards the centre following x_2 orbits. Using this simple representation of the gas flow, BGSBU provided an appealing interpretation of observational data: HI emission comes mainly from gas on non self-intersecting x_1 orbits, CO forms as gas is shocked on reaching the cusped orbit, which explained the characteristic parallelogram-shaped envelope of CO emission in the (l, v) plane, and CS emission comes from dense, post-shocked gas flowing on x_2 orbits.

As we have mentioned in Sect. 1.2, the BGSBU picture has enjoyed considerable success and, alongside the photometric evidence of Blitz & Spergel (1991), was among the studies that convinced the community that our Galaxy is barred. However, the BGSBU picture relied on the assumption that gas streamlines follow closed orbits and that the $x_1 \to x_2$ transfer happens at the cusped orbit, and needed validation by hydro simulations. In support of their model, BGSBU pointed to simulations by Athanassoula (1992b). However, these simulations used a different potential from BGSBU, so a natural next step was to run hydro simulations in the potential BGSBU had used. One such study appeared (Jenkins & Binney, 1994), but its

results were not encouraging: in this simulation, orbits near the cusped x_1 orbit, which played a crucial role in BGSBU picture, were found to be unoccupied by gas.

More recently (see also Sec. 1.2) numerous hydrodynamical simulations have been run with the goal of understanding the kinematics and dynamics of the cold gas in our Galaxy (Mulder & Liem, 1986; Weiner & Sellwood, 1999; Englmaier & Gerhard, 1999; Lee et al., 1999; Fux, 1999; Bissantz et al., 2003; Rodriguez-Fernandez & Combes, 2008; Baba et al., 2010; Pettitt et al., 2014). However, none of these simulations uses the potential BGSBU used. Some used a potential inferred from infrared photometric data (Englmaier & Gerhard, 1999; Bissantz et al., 2003; Rodriguez-Fernandez & Combes, 2008), while others (Fux, 1999) used the potential generated by a combined hydro and N-body simulation of the Milky Way. Notwithstanding these efforts, the interpretation of the observational data remains problematic in some aspects (see also Chapters 5 and 6).

Most authors of papers using hydro simulations mention closed orbits, in a more or less explicit connection with BGSBU, but the literature lacks a detailed examination of the extent to which good hydro simulations support the BGSBU picture. The nearest the literature comes to filling this need is the paper of Jenkins & Binney (1994), which describes simulations that are of low resolution by today's standard, and uses sticky particles rather than a conventional hydro simulation based on the Euler equations. Papers comparing closed orbits with the results of hydro simulations in bar potentials can be found (van Albada & Sanders, 1982; Athanassoula, 1992a,b; Bureau & Athanassoula, 1999; Athanassoula & Bureau, 1999), but, for lack of resolution or other reasons, none of them provides sufficient detail to show which orbits are or are not occupied by gas, especially in the vicinity of the cusped orbit, which plays a crucial role in the BGSBU picture.

In this chapter, we use high-resolution hydrodynamical simulations to re-examine the BGSBU picture. In the first part, we test the extent to which the physics of the gas flow hypothesised by BGSBU is supported by the simulations. Is the gas flow far from the shocks well approximated by closed orbits? Can the gas flow be understood as a transfer from x_1 to x_2 orbits? Does the transition happen at the cusped orbit as conjectured by BGSBU? We show that the answers to these questions depend on the spatial resolution and sound speed used in a hydro simulation. The results are likely to be valid for all barred potentials that have a general resemblance to the BGSBU potential, i.e., their orbital structure is similar to that of BGSBU potential. In the second part of the chapter, we discuss the implications of our results for the interpretation of the observational data. Can we identify in the simulations structures reminiscent of the CO parallelogram of BGSBU? Under what conditions does the size of the x_2 disc match the region covered by CS emission? Can we explain the high velocity peaks in the HI (l, v) diagrams at $|v| \simeq 270\,\mathrm{km\,s^{-1}}$ and $|l| \simeq 2°$?

This chapter is structured as follows. In Sect. 3.2 we present the numerical schemes employed in the simulations. In Sect. 3.3 we show the results of the hydro simulations. We discuss the physical interpretation of the simulations in Sect. 3.4, and in Sect. 3.5 we discuss their implications for the interpretation of observational data. We finally summarise our findings in Sect. 3.6.

3.2 Methods

3.2.1 Hydro Simulation Scheme

We assume that the gas is a two-dimensional inviscid isothermal fluid governed by the Euler equations. A justification for this set of equations is discussed in Sect. 1.3. An additional term is introduced in the continuity equation to implement the recycling law of Athanassoula (1992b). The dynamical equations in frame rotating with pattern speed $\mathbf{\Omega}_\mathrm{p}$ are

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) &= \alpha(\rho_0^2 - \rho^2), \\ \partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla)\, \mathbf{v} &= -\frac{\nabla P}{\rho} - \nabla\Phi - 2\mathbf{\Omega}_\mathrm{p} \times \mathbf{v} - \mathbf{\Omega}_\mathrm{p} \times (\mathbf{\Omega}_\mathrm{p} \times \mathbf{x})\,, \\ P &= c_\mathrm{s}^2 \rho, \end{aligned} \tag{3.1}$$

where $\mathbf{x}$ is the position vector in the rotating frame, ρ is the surface density of the gas, P is the pressure, Φ is the gravitational potential, $\mathbf{v}$ is the velocity, c_s is the sound speed, α is a constant representing the efficiency of the recycling law and ρ_0 is the initial surface density. In our simulations the gas is assumed to be isothermal. Hence c_s is a constant number.

The recycling law was originally meant to take into account in a simple way the effects of star formation and stellar mass loss. In practice, the only effect of the recycling law is to prevent too much gas from accumulating in the very centre and to replace gas lost at the boundary due to the outflow boundary conditions. It does not affect the morphology of the results, apart from a minor softening of the density contrast which does not affect our results. So our results do not change if we disable the recycling law. We adopt a recycling efficiency $\alpha = 0.3\, M_\odot\,\mathrm{pc}^{-2}\,\mathrm{Gyr}^{-1}$ and an initial density $\rho_0 = 1\, M_\odot\,\mathrm{pc}^{-2}$.

The gravitational potential Φ is assumed to be an externally imposed, rigidly rotating barred potential. It is time-independent in the frame rotating with pattern speed $\mathbf{\Omega}_\mathrm{p}$. The density distribution that gives rise to our potential is described in the next section. The gas is assumed to be non self-gravitating and therefore it does not contribute to the potential.

We use a grid-based, Eulerian code based on the second-order flux-splitting scheme developed by van Albada et al. (1982) and later used by Athanassoula (1992b), Weiner & Sellwood (1999) and others to study gas dynamics in bar potentials. The same implementation will also be used in later chapters. Our simulations are two-dimensional. The output of each simulation consists in snapshots of the velocity and surface density distributions $\rho(\mathbf{x})$ and $\mathbf{v}(\mathbf{x})$ at chosen times.

We used a grid $N \times N$ to simulate a square 10 kpc on a side. N depends on the resolution of the simulation. For example, if the grid cells are $\mathrm{d}x = 5\,\mathrm{pc}$ on a side, we have $N = 2000$. In each run the initial conditions are as follows. We start with gas in equilibrium on circular orbits in an axisymmetrized bar and, to avoid transients, turn on the non-axisymmetric part of the potential gradually during the first 150 Myr, in such a way that the total mass of the underlying potential is conserved in the process. We impose outflow

boundary conditions; that is, we allow an outward (but no inward) flux of mass and momentum across the edges of the grid. The gas at the edges of the simulated region behaves as if it were in contact with a pressureless fluid. Hence, gas can freely escape the simulated region, after which it is lost forever. The potential well is sufficiently deep, however, that very little gas escapes the regions of interest. We have experimented with alternative schemes in which we instead hold fixed the density and momentum in the outermost cells of the grid, and we find no change to our conclusions. We have checked that doubling the size of the simulated region to simulate a square 20 kpc on a side does not affect the results in the region of interest. The region of interest is the one shown in the figures of this chapter (a square 4 kpc on a side), which is already much smaller than the whole simulated area (as stated above the total simulated area is a square 10 kpc on a side).

3.2.2 The Potential

We use the same potential as Jenkins & Binney (1994). This arises from two components. The first is the bar used by BGSBU, which has the density distribution:

$$\rho_{\rm b}(a) = \rho_{\rm b0} \begin{cases} (a/a_0)^{-\alpha} & \text{if } a \leqslant a_0 \\ (a/a_0)^{-\beta} & \text{if } a > a_0 \end{cases} , \tag{3.2}$$

where $a = \sqrt{x^2 + (y^2 + z^2)/q^2}$ and the values of the parameters are $a_0 = 1.2\,\mathrm{kpc}$, $\alpha = 1.75$, $\beta = 3.5$, $\rho_{\rm b0} = 0.69\,M_\odot\,\mathrm{pc}^{-3}$ and $q = 0.75$. The major axis of the bar always lies along the x axis. The second component is a razor-thin exponential disc, which has been added to complete realistically the circular velocity curve outside $R \simeq 1\,\mathrm{kpc}$, and it has little influence inside this radius. The exponential disc is generated by a surface density distribution

$$\Sigma(R) = \Sigma_0 \mathrm{e}^{-R/R_{\rm d}}, \tag{3.3}$$

where R is the radius in cylindrical coordinates and the parameters have values $\Sigma_0 = 1300\,M_\odot\,\mathrm{pc}^{-2}$ and $R_{\rm d} = 4.5\,\mathrm{kpc}$.

The potential is assumed to be rigidly rotating with constant pattern speed $\Omega_{\rm p} = 63\,\mathrm{km\,s^{-1}\,kpc^{-1}}$. This places the inner Lindblad resonance (ILR) at $R_{\rm ILR} = 0.6\,\mathrm{kpc}$ and corotation at $R_{\rm CR} = 3.7\,\mathrm{kpc}$. The circular velocity curve for this potential and for the potential of Jenkins & Binney (1994) is shown, along with that of potentials that will be used later, in Fig. 5.1 of Chapter 5. The potential is calculated numerically using a multipole expansion (see Section 2.1 of Magorrian 1995 for details).

3.2.3 Projecting to the (l, v) plane

We adopt a very simple projection procedure to produce the predicted (l, v) distributions for each simulation snapshot $(\rho(\mathbf{x}), \mathbf{v}(\mathbf{x}))$. Throughout this chapter and in most of this thesis, we assume that the Sun is

undergoing circular motion at a radius $R_0 = 8\,\mathrm{kpc}$ with speed $v_\odot = 220\,\mathrm{km\,s^{-1}}$. Calling ϕ the angle between the major axis and the Sun–GC line, the Cartesian coordinates of the Sun are given by $x_\odot = R_0 \cos\phi$, $y_\odot = R_0 \sin\phi$. All the projections in this chapter assume $\phi = 20°$.

The resolution of our (l, v) diagrams is $\Delta l = 0.25°$ in longitude and $\Delta v = 2.5\,\mathrm{km\,s^{-1}}$ in velocity. The (l, v) intensities for a given line of sight are produced in the following way. Along the line of sight, we sample the xy density and the velocity by linearly interpolating the results of the simulations at points separated by $\delta s = 1\,\mathrm{pc}$. These points are then distributed in velocity bins of width $\Delta v = 2.5\,\mathrm{km\,s^{-1}}$. The final (l, v) intensity at the chosen longitude in each range of velocity is obtained by summing over all the relevant points along the line of sight weighted by their densities.

This procedure yields a predicted brightness temperature that is linear in column density and it is equivalent to the simplest radiative transfer calculation (i.e. the gas is optically thin). In the case of HI, the brightness temperature is linear in the column density if the gas has constant spin temperature and its optical depth is negligible. So our projection is equivalent to very simple HI radiative transfer in the constant-temperature, optically-thin case. The assumption of constant temperature is known to be a simplification for Galactic HI, which is instead often modelled as a medium made by two or more phases at different temperatures (see for example Ferrière, 2001). In the case of ^{12}CO, the brightness temperature is not linearly related to density when considering a single cloud, but a linear relationship will hold between brightness temperature and the number density of unresolved CO clouds provided the cloud density is low enough for shadowing of clouds to be unimportant (see, e.g., Binney & Merrifield, 1998, §8.1.4).

3.3 Results

Fig. 3.1 shows a selection of closed orbits in the BGSBU potential, in the frame corotating with the bar. Orbits that BGSBU believed to carry gas are shown by full lines, while those they thought empty are shown by dashed lines. In the outer region we show a nested sequence of non self-intersecting x_1 orbits that terminates in the cusped orbit, drawn in black. Inside this orbit we show dotted two self-intersecting x_1 orbits. The vertically elongated orbits belong to the x_2 family, which extends quite a bit beyond the point where the cusped x_1 orbit intercepts the vertical axis. BGSBU argued that gas transfers between the x_1 and x_2 families at the cusped orbit. A primary goal of this chapter is to test this conjecture.

Fig. 3.2 summarises the results of our simulations. It shows the density of hydro simulations for different grid spacings $\mathrm{d}x$ and sound speeds c_s. All snapshots are taken at the same time $t = 280\,\mathrm{Myr}$. Some common features of the gas flow can be identified in all panels. In the outer part, approximately corresponding to the green region, the gas follows x_1 orbits. At some point near the x axis, two thin offset shocks emerge, which connect the green region to the reddish central disc. The central disc is called the x_2 disc and is made by gas on x_2 orbits. Most of the gas plunges to the x_2 disc or "central molecular zone" through the shocks.

Thus, the gas follows the x_1 orbits in the outer part and the x_2 orbits in the central part, with a transition zone containing the shocks in between. In each simulation, we can identify an innermost occupied x_1 orbit.

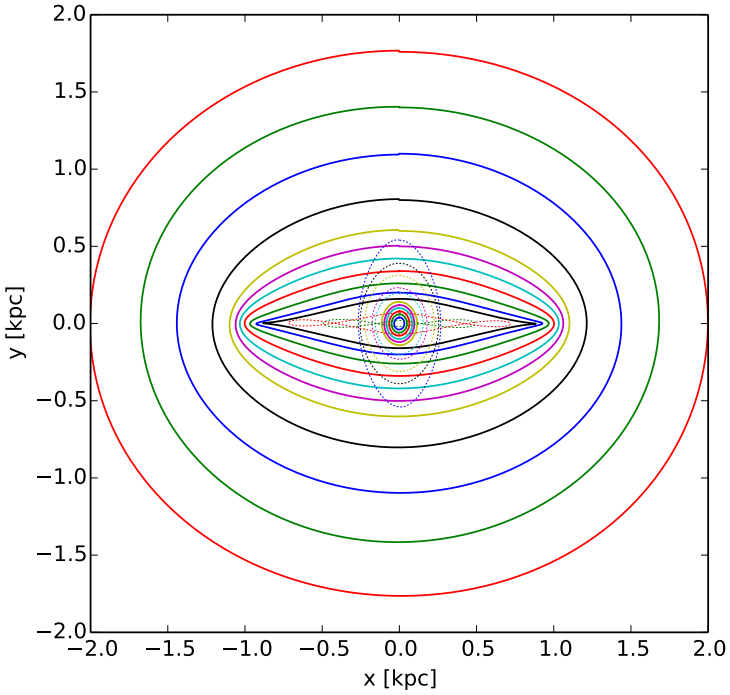


Figure 3.1: x_1 and x_2 orbits in the potential used by BGSBU. Orbits are shown in the frame that rotates (clockwise) with the bar. The horizontally elongated orbits form the x_1 family, while the vertically elongated ones are from the x_2 family. BGSBU hypothesised that the orbits drawn in full lines have gas, while those shown dashed are unoccupied. The cusped x_1 orbit is the smallest horizontally elongated orbit drawn in full lines and is shown black.

The shocks are formed just after this orbit and they induce the transition from the x_1 to the x_2 family. We call this innermost occupied x_1 orbit the transition orbit, and the transition point its position in a parametrisation of the sequence of x_1 orbits.

BGSBU assumed that the transition orbit is the cusped orbit. In our simulations, the transition point and the size of the x_2 disc depend strongly on both the resolution and the sound speed. Consider, for example, the middle column in Fig. 3.2, corresponding to $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$. As we increase the resolution, the transition point moves inwards while the x_2 disc shrinks. At the highest resolution, $\mathrm{d}x = 5\,\mathrm{pc}$, the transition orbit almost coincides with the cusped orbit, as predicted by BGSBU (see also Section 3.4.2 and 3.4.4 for discussions on numerical resolution and convergence). At lower resolution the transition happens earlier and the transition orbit is much bigger than the cusped orbit. Fig. 3.3 shows the same density snapshots as Fig. 3.2, superimposed on the closed orbits that BGSBU thought carried gas for a better comparison.

Increasing the sound speed also has the effect of postponing the transition and shrinking the x_2 disc. Consider, for example, the second row in Fig. 3.2, corresponding to $\mathrm{d}x = 20\,\mathrm{pc}$. At this resolution, the transition happens very early, for $c_\mathrm{s} = 5\,\mathrm{km\,s^{-1}}$, and the transition orbit is quite an outer x_1 orbit (approximately the yellow orbit in Fig. 3.1). As we increase the sound speed, the transition orbit moves inwards, and for $c_\mathrm{s} = 20\,\mathrm{km\,s^{-1}}$ it coincides in Fig. 3.1 with the green orbit that lies just outside the cusped orbit. We will discuss the origins of these systematics in Sect. 3.4.

The two highest-resolution simulations in the right column of Fig. 3.2 look peculiar. These correspond to $c_\mathrm{s} = 20\,\mathrm{km\,s^{-1}}$ and $\mathrm{d}x = 5, 10\,\mathrm{pc}$. At the time shown, all other simulations have already reached an

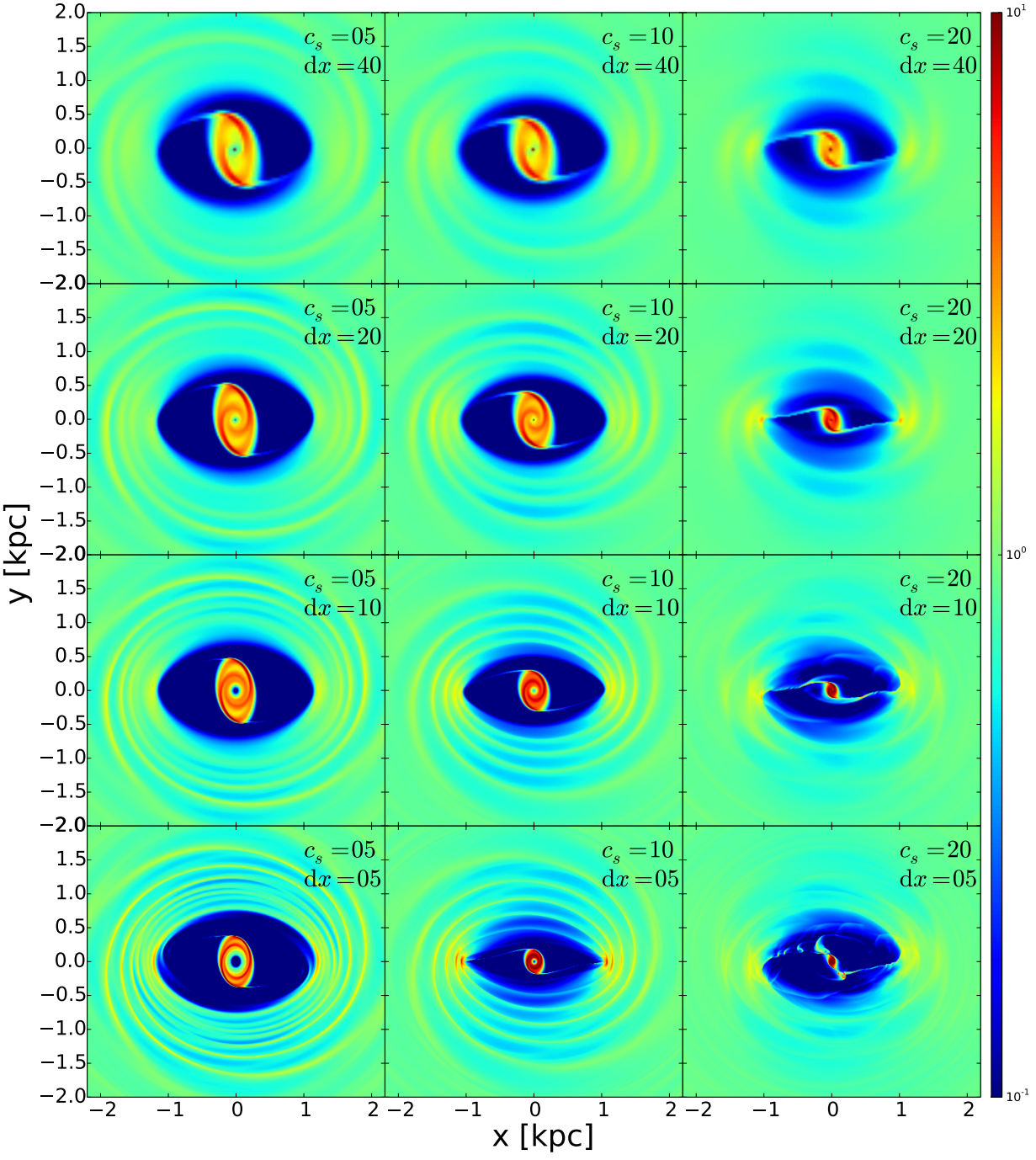


Figure 3.2: The fluid density in hydro simulations in the BGSBU potential for different spatial resolutions and sound speeds. c_s is increasing left to right taking values 5,10,20 km s^{-1}. dx is decreasing from top to bottom taking values 40,20,10,5 pc. Gas has reached an approximately steady state in the rotating frame and circulates clockwise. All snapshots are taken at $t = 280$ Myr. The colorbar is in units of $M_\odot$ pc^{-2}.

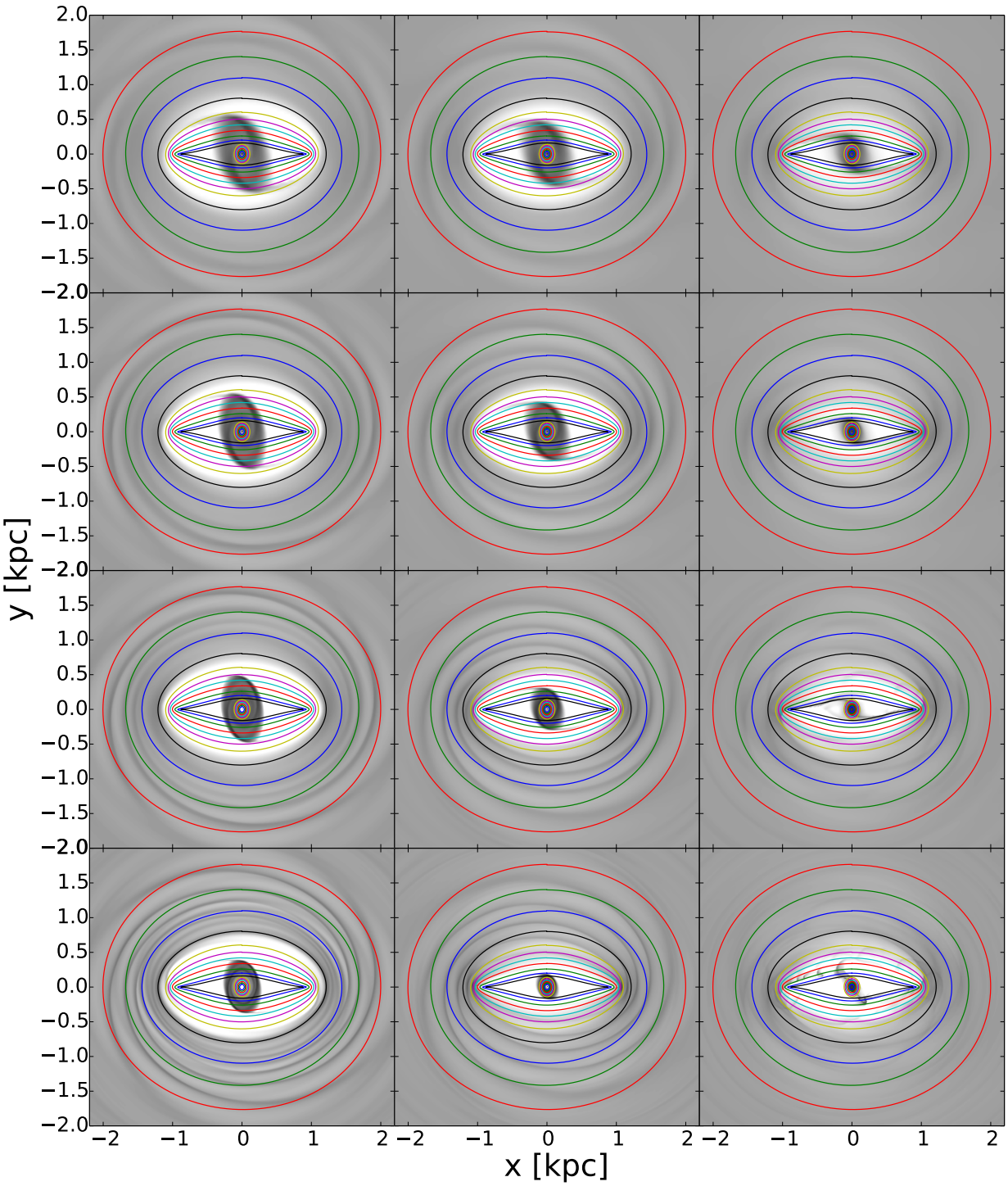


Figure 3.3: Same as Fig. 3.2 but with superimposed orbits.

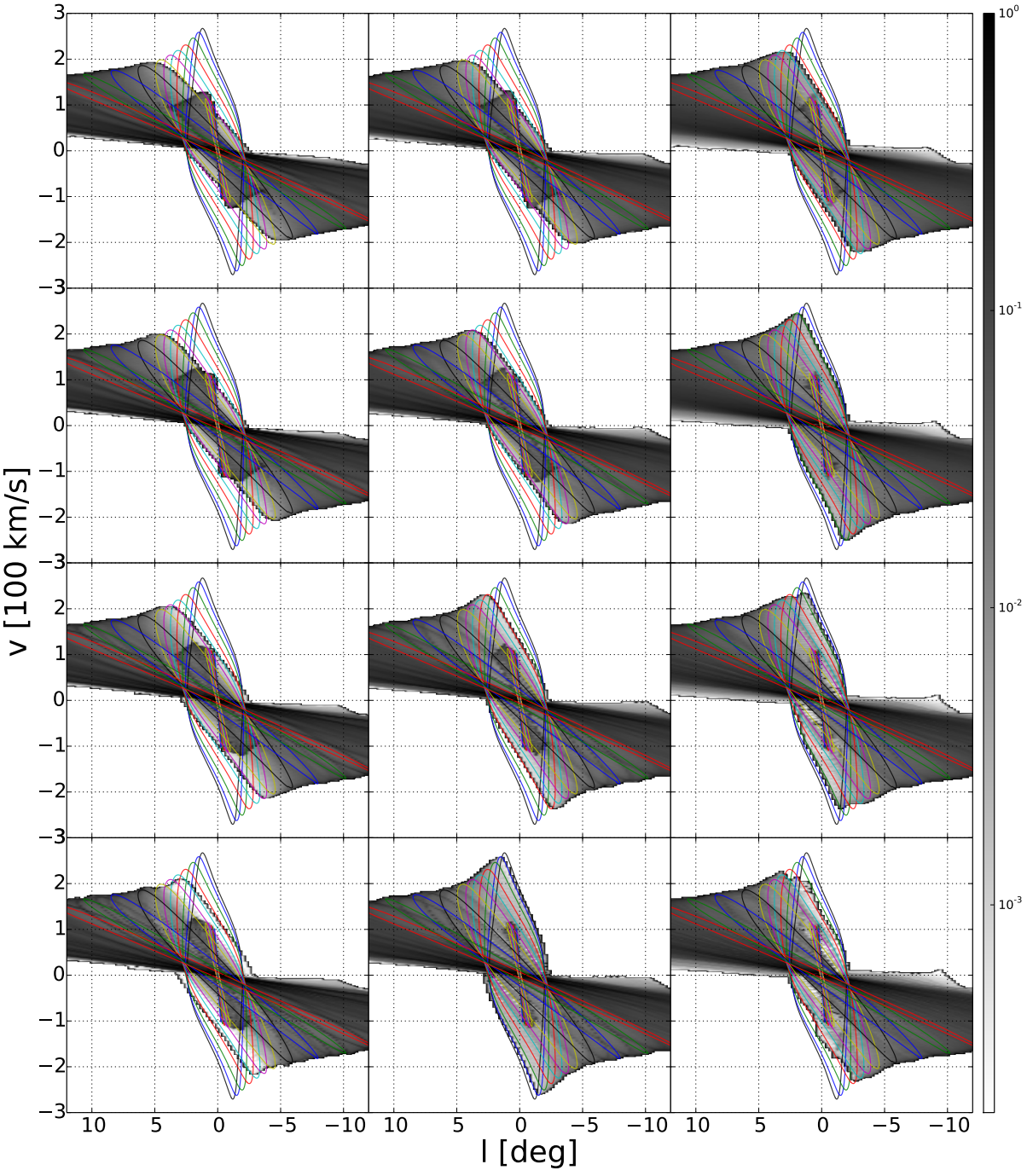


Figure 3.4: The simulations of Figs. 3.2 and 3.3 projected onto the (l, v) plane. Solid lines show the (l, v) traces of closed orbits, with colours matching other figures. The Sun is assumed to be in a circular orbit with $v = 220\,\mathrm{km\,s^{-1}}$ at $R_0 = 8\,\mathrm{kpc}$, and the bar major axis makes an angle $\phi = 20°$ with the Sun-Galactic centre line. The colorbar is in arbitrary units.

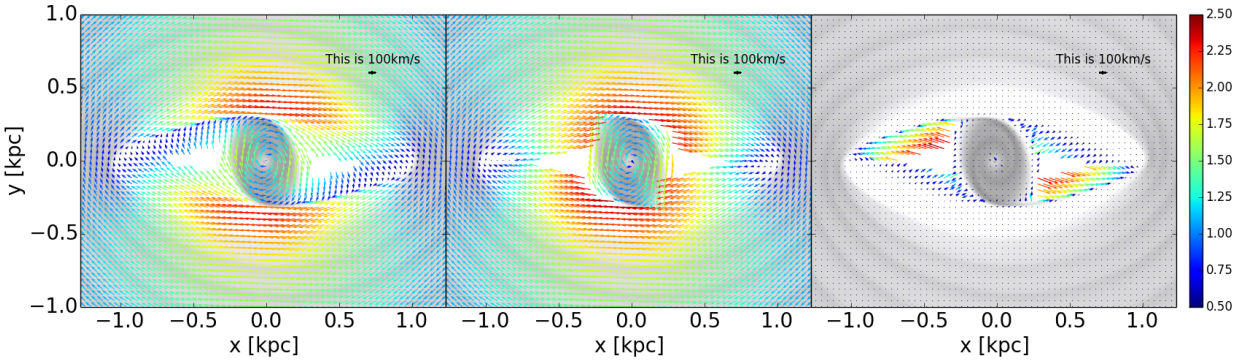


Figure 3.5: Using closed orbits to approximate the hydro velocity field for the simulation with $\mathrm{d}x = 10\,\mathrm{pc}$, $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$ shown at left. The central panel shows the best approximation to that velocity field obtained using closed orbits belonging to the x_1 or x_2 family. For each point where more than one orbit is found, the chosen velocity is the one closest to the hydro velocity. The right panel shows the vector differences between the left and middle panels. In all panels points without orbits are masked and the density of the hydro simulation is visible in the background. The colorbars show the speed at each point in units of $100\,\mathrm{km\,s^{-1}}$.

approximate steady state, and they would not appear significantly different after another 250 Myr or more. These two simulations instead manifest a complex unsteady flow interior to the transition orbit. The shocks are not stable, but keep forming and dissolving in an endless cycle. In some snapshots, they are almost completely formed and smooth (see also Sect. 3.4.3 and Fig. 3.9). At these moments, they lie very close to the x axis. A little later, vortices grow on the leading side, which move around and eventually bump on the opposite shock, creating more vorticity. Later the cycle repeats. Tests described in Sect. 3.4.3 suggest that the unsteadiness is real and not an artifact of our particular code.

So far we have not discussed the velocity structure of our simulations, but this is crucial for the interpretation of observations. Fig. 3.4 shows the projections into the (l, v) plane of the snapshots of Fig. 3.2 for an assumed angle $\phi = 20°$ between the bar's major axis and the Sun-Galactic centre line. The lines in Fig. 3.4 show the (l, v) traces of some of the closed orbits plotted in Fig. 3.3 using the same colour scheme – we plot only the orbits that BGSBU thought occupied (those shown with full lines in Fig. 3.1).

For all resolutions and sound speeds, in Fig. 3.4 the envelope of the outermost x_1 orbits matches the envelope of the hydro distribution very well. As we move towards smaller values of $|l|$, the traces of orbits sweep up towards the high-velocity peak of the cusped orbit, but at the transition orbit the hydro envelope starts to fall, and thus becomes separated from the orbit envelope. The projection of the hydro x_2 disc is clearly identifiable as a darker region near the centre. At the smallest longitudes its boundary is delineated by the traces of the x_2 orbits, but for low sound speeds or resolutions it extends far beyond the region covered by the plotted x_2 orbits.

This finding again confirms the picture of the gas flow that we delineated above. When the transition point is too early, the innermost non-self intersecting x_1 orbits, which are populated by gas in the BGSBU picture, are void of gas in the hydro simulations. Outer x_2 orbits that lack gas in the BGSBU picture are occupied by gas in the hydro simulation. As we increase the resolution at $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$ (middle column), the transition between the x_1 and x_2 families moves inwards, and the hydro envelope matches more and more closely the predictions of the BGSBU picture. At the same time, the projection of the x_2 disc shrinks,

also approaching the BGSBU picture. Eventually, for $\mathrm{d}x = 5\,\mathrm{pc}$ and $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$, the projection of the hydro x_2 disc matches very well with the BGSBU picture (see also Section 3.4.2 and 3.4.4 for discussions on numerical resolution and convergence).

At low sound speed (left column of Fig. 3.4), the transition always happens early, so orbits close to the cusped orbit are unoccupied at all resolutions. Therefore at low sound speeds the hydro simulations gives results inconsistent with the BGSBU picture. At high sound speeds the hydro simulations approach the BGSBU picture as we increase the resolution at first. But at higher resolution, unsteadiness makes the projected hydro deviate significantly from the BGSBU picture.

To explain this situation from a face-on perspective, we take as an illustrative example the case $\mathrm{d}x = 10$, $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$. Fig. 3.5 compares the velocity field of the hydro simulation with that of the closed orbits. The left panel shows the velocity field of the hydro simulation. The central panel shows the best approximation to the hydro velocity field that can be obtained from the x_1 and x_2 orbits: at each point where more than an orbit is present, we show the velocity of the orbit that best matches the hydro velocity field. In all panels the density of the hydro simulation is shown in the background. Locations through which no closed orbit passes are masked out. The right panel is the most interesting panel: it shows the vector difference between the left and middle panels, and shows clearly that the outer x_1 orbits and the inner x_2 orbits both reproduce the hydro velocity field accurately. The red orbit in Fig. 3.1 is the transition orbit in this case, the last orbit at which the velocity fields coincide, and is just outside the shocks. After this, the shocks emerge and the two velocity fields suddenly diverge.

At different resolutions and sound speeds, the situation is qualitatively very similar but the point of transition between orbit families changes. For $c_\mathrm{s} = 20\,\mathrm{km\,s^{-1}}$ and high resolution the hydro flow in the transition region is unsteady.

3.4 The physics of the gas flow

3.4.1 Dependence on the sound speed

In Sect. 3.3, we have seen that as the sound speed increases, the transition point and the shocks move inwards. At the same time, the shocks also become more horizontal and closer to the x axis. Englmaier & Gerhard (1997) and Patsis & Athanassoula (2000) found similar results when varying the sound speed. A qualitative explanation for this behaviour is as follows (for another discussion see also Englmaier & Gerhard, 1997).

The more elongated an x_1 orbit is, the larger the variation along the orbit in the speed of an orbiting particle. Since the speed decreases as particles move from the orbit's minor axis to its major axis, the density of gas that is streaming along the orbit increases as the major axis is approached. When the sound speed is high, pressure assists gravity in slowing gas as it approaches the major axis, and equally assists gravity in accelerating the flow after the major axis has been passed. In the absence of a shock, the gas is reversibly

compressed and decompressed so it can continue to keep close to one orbit for several revolutions.

If the sound speed is too low, the convergence of the gas as the major axis is approached leads to shock formation. Entropy is created in the shock, so the decompression after the major axis has been passed does not reverse what happened as the axis was approached, and the flow deviates strongly from orbits.

The key to avoiding shock formation is the ability of sound waves to carry information about fluctuations in density upstream so oncoming gas can be slowed in a timely manner when the density increases at a downstream location. The nearer the cusped orbit is approached, the larger is the velocity gradient up which sound waves have to travel if a shock is to be avoided. Hence decreasing the sound speed causes the shock to form further out.

3.4.2 Dependence on numerical resolution

Increasing the resolution moves the orbit at which shocks form inwards with important consequences for the interpretation of the gas flow in our Galaxy that will be discussed in Sect. 3.5. A likely explanation for this phenomenon is that at low resolution the innermost x_1 orbits are inadequately resolved. To resolve the cusp we need in principle infinite resolution. By increasing the resolution, we can resolve orbits closer and closer to the cusped one, and gas can settle on these orbits, delaying the formation of shocks.

Fig 3.6 tests this idea by showing the (l, v) traces of nine orbits from the cusped orbit (at the top) outwards. Each orbit is mapped four times at resolutions that increase from left to right. The faint solid lines show the traces of the orbits and as such are the same along every row. The big points scattered around this line show the (l, v) trace one obtains by associating each hydro cell through which the orbit passes with the mean velocity of its passage, and then projecting the resulting velocities of the visited cell. The small points show the projection at the visited cells of the hydro velocity field for $c_{\rm s} = 10\,{\rm km\,s^{-1}}$.

We see that orbits that are well outside the cusped orbit (lower part of the figure) are well represented by even the coarsest grid (left side of the figure), and, moreover, their velocities are accurately reproduced by the hydro code. As we move up the figure and therefore approach the cusped orbit, at low resolutions the big dots scatter widely around the orbit's curve. The scatter is widest at the largest values of $|l|$ because that is where the velocity gradient is largest and thus the finite resolution of the grid has its biggest impact. For each orbit outside the cusped orbit there is a resolution finer than which the hydro velocity field is almost the same as the orbital velocity because along that orbit the flow does not encounter a shock.

3.4.3 Unsteady flow

As mentioned in Sect. 3.3, the flow becomes unsteady at high sound speed and high resolution. To test whether this phenomenon was an artifact generated by our code, we re-ran some simulations with a completely different code, PLUTO (Mignone et al., 2007). This is free software for the numerical solution of systems of conservation laws targeting high Mach number flows in astrophysical fluid dynamics. It has a modular structure that makes it easy to change the algorithm used to simulate a flow. We use this modularity to

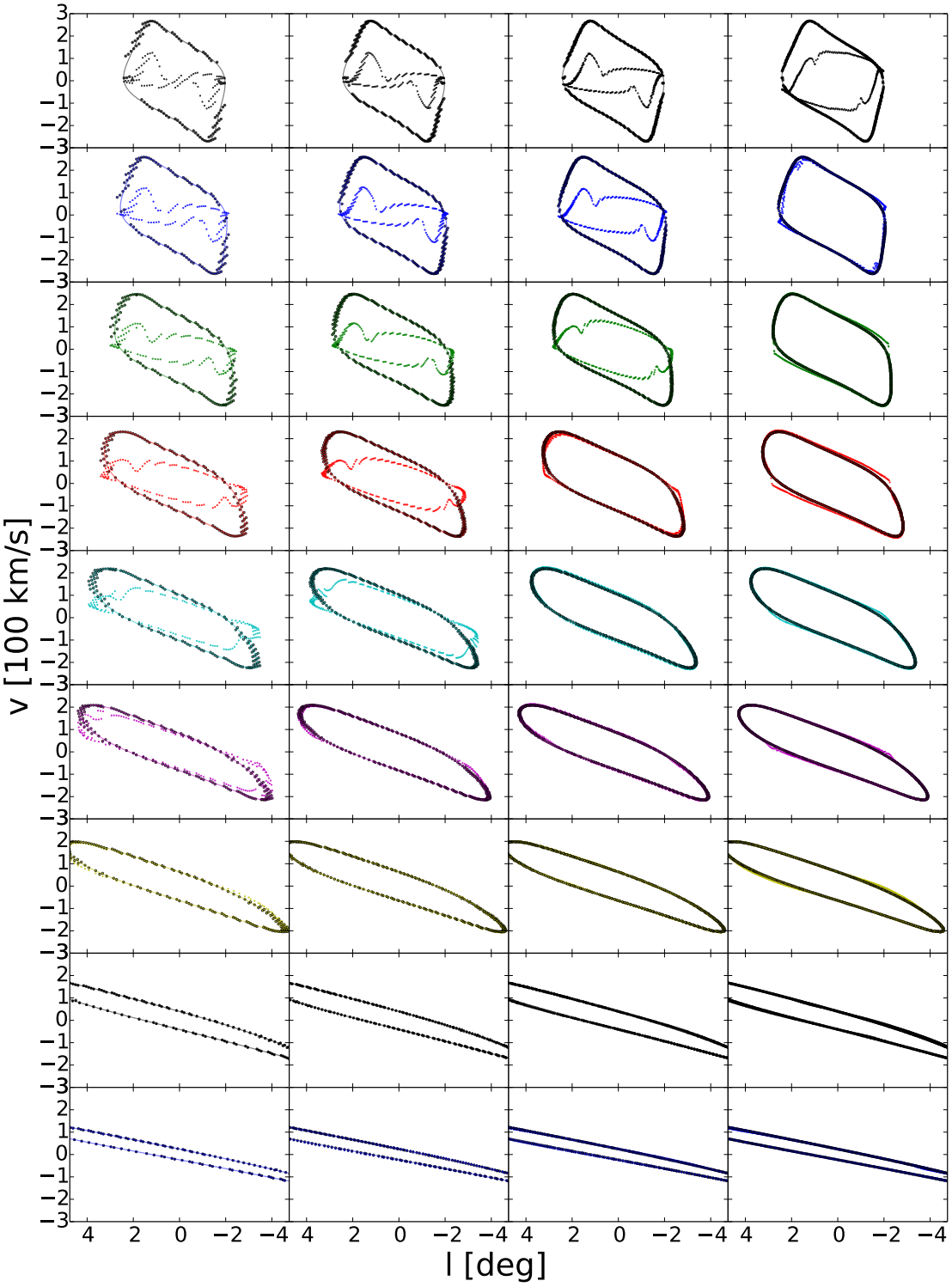


Figure 3.6: Each panel refers to a single x_1 orbit. In full lines, the trace of the orbit. The bigger, empty dots are the (l, v) projections of the velocity field of closed x_1 orbits sampled with the same resolution as the hydro simulation along the xy trajectory of the orbit. Smaller, filled dots are the projection of a hydro simulation cells along the same trajectory. Each column refers to one hydro simulation. From left to right, the resolution increases from $\mathrm{d}x = 40\,\mathrm{pc}$ to $\mathrm{d}x = 5\,\mathrm{pc}$. All simulations are for $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$.

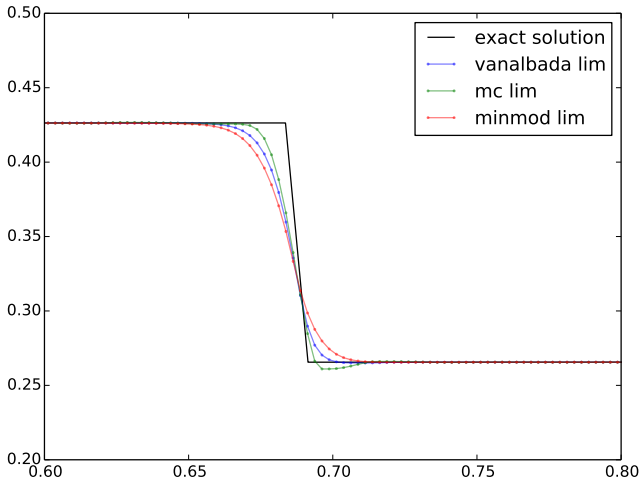


Figure 3.7: Results of SOD shock tube test problem at $t = 0.2$ for different flux limiters. This figure zooms around a shock. The resolution is $\mathrm{d}x = 0.0025$. On the vertical axis there is density, on the horizontal axis position.

investigate the effects of the choice of Riemann solver and flux limiter. A Riemann solver is a numerical scheme for solving the *Riemann problem*: given density, momentum (and energy) at the interface of two grid cells, what happens next to the gas distribution across the cell boundary? Calculating the fluxes across the boundary involves calculating spatial derivatives. When the gas distribution is smooth, that can be carried out using a second-order finite-difference scheme. In the presence of a shock, however, such a second-order scheme would produce nonsense. Hence, Riemann solvers typically employ a flux limiter that effectively switches the algorithm to a simpler, first-order upwind scheme wherever the spatial derivatives become large enough. In all of our runs we used a static Cartesian mesh and 2nd order Runge-Kutta (RK2) time-stepping. The code was modified to implement our recycling law (see Eqs. (3.1)).

We found that the Roe and Harten-Lax-van Leer (HLL) Riemann solvers (see also PLUTO's manual) produced the same results as each other, and our own code. The choice of flux limiter, by contrast, does have a significant impact on the computed flow: unsteadiness of the flow can be suppressed by changing to a more diffusive flux limiter. Flux limiters are used in numerical schemes to handle the flow close to discontinuities such as shocks. We tested three different limiters and found that unsteadiness is suppressed by choosing a more diffusive limiter, which has a higher numerical viscosity and produces thicker shocks. Fig. 3.7 shows the result of using PLUTO to perform a standard one-dimensional test problem, the SOD shock tube (Sod, 1978). It shows that the three limiters produce shocks of different thicknesses. The most diffusive (more viscosity) is the MINMOD limiter (Roe, 1986). The Van Albada limiter used in this paper (van Albada et al., 1982) has intermediate diffusivity, while the least diffusive limiter (least viscosity) is the MC limiter (van Leer, 1977).

Fig. 3.8 shows a snapshot at an intermediate time for simulations with $\mathrm{d}x = 5\,\mathrm{pc}$ and $c_\mathrm{s} = 20\,\mathrm{km\,s^{-1}}$ obtained using PLUTO. The only difference between the three simulations is the limiter used. From top to bottom the figure shows the flows obtained with limiters of increasing diffusivity. The top flow has a very

irregular and unsteady structure that is completely absent from the bottom flow. The central flow shows an intermediate level of unsteady structure. Unsteady features seem to arise at the shocks, which develop wrinkles that shed vortices.

The right panels in Fig. 3.8 show the potential vorticity, defined as

$$\boldsymbol{\Xi} \equiv \frac{\nabla \times \mathbf{v} + 2\boldsymbol{\Omega}_{\rm p}}{\rho}. \tag{3.4}$$

It can be proven that in a two-dimensional isothermal fluid governed by Eqs. (3.1) (disabling the recycling law), the potential vorticity is conserved along streamlines (but it can vary from streamline to streamline). However, at shocks the equation of motions break down. Indeed, we find that at shocks the potential vorticity is not conserved (see Binney, 1974, for an analysis of vorticity generation in shocks).

The mean gas flow is clockwise in these plots. At the shocks, swirls that rotate counter-clockwise are formed and injected into the flow. These are the blue pearls along the shocks in the top right panel. These swirls are counter-rotating as the potential vorticity inside them has the opposite sign than everywhere else. In the figures, the color associated with negative values of the potential vorticity is the same color associated with the lowest value of the potential vorticity shown in the colorbar. The swirls can be identified as they are the only regions in the figures to reach such low and negative potential vorticity. Once the swirls are produced, they cannot be destroyed due to the conservation of potential vorticity until they meet another shock. Eventually, a steady state situation is reached in which the same amount of counter-rotating vorticity is produced and destroyed at the shocks on average.

Fig. 3.9 shows the flow computed with the most diffusive limiter (MINMOD) at three different later times. It shows that at late times counter-rotating swirls flow around, cyclically formed and destroyed. Shocks keep developing wrinkles and swirls. These temporarily alter their straight shapes, which are then cyclically restored. More unstable simulations also display similar behaviour, but in these lower-diffusivity calculations the rate at which potential vorticity is created is higher, shocks do not have time to restore their full shapes and the eventual steady-state where the same amount of potential vorticity is destroyed and created at shocks is more chaotic.

An alternative way of looking at the instability is the following. Turbulence is a consequence of high rates of shear in a high Reynolds number flow. In the flows studied by engineers, for example in pipes and over aircraft wings, turbulence arises in the thin boundary layer that forms when a low-viscosity fluid meets a solid surface. Vortices formed in the boundary layer move into the bulk of the fluid, making the flow generally unsteady. In our simulations, vortices form along the shocks, where the shear rate is exceptionally high. The rate of shear is inversely proportional to the width of the shock, so it increases with the grid resolution and decreases with the diffusivity of the limiter. Thus simulations with the finest grids and the least diffusive limiters have the highest shear rates and are most likely to become turbulent.

The fact that turbulence appears only in simulations with a high sound speed can also be explained by the connection between shear rate and the onset of turbulence: the rate of shearing increases along the

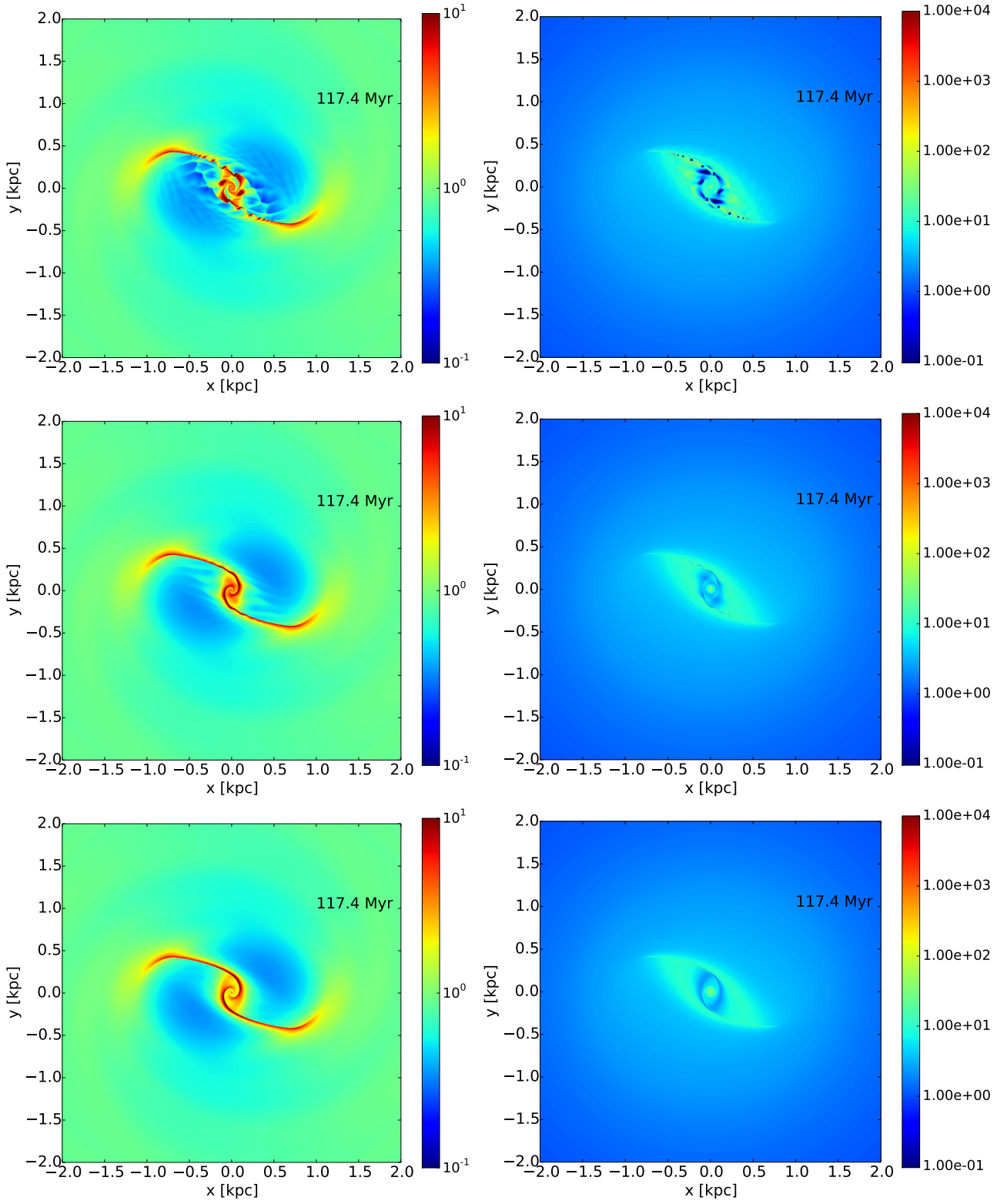


Figure 3.8: Snapshots capturing the moment when the instability is forming. The three snapshots come from three different simulations obtained with the PLUTO code. The only difference between the three is the limiter used. On top using the MC limiter, middle the VanAlbada limiter and bottom the MINMOD limiter. The left panels show the gas density, while the right panels show the potential vorticity, defined in Eq. (3.4). The colorbars on the left show the gas density in units of $M_{\odot}\,\mathrm{pc}^{-2}$. The colorbars on the right show the potential vorticity in units of $\mathrm{km\,s}^{-1} M_{\odot}^{-1}\,\mathrm{pc}$. The counter-rotating swirls are regions of the same color as the color of the lowest available value in the colorbars. Inside these swirls, the potential vorticity is negative, i.e., they are counter-rotating.

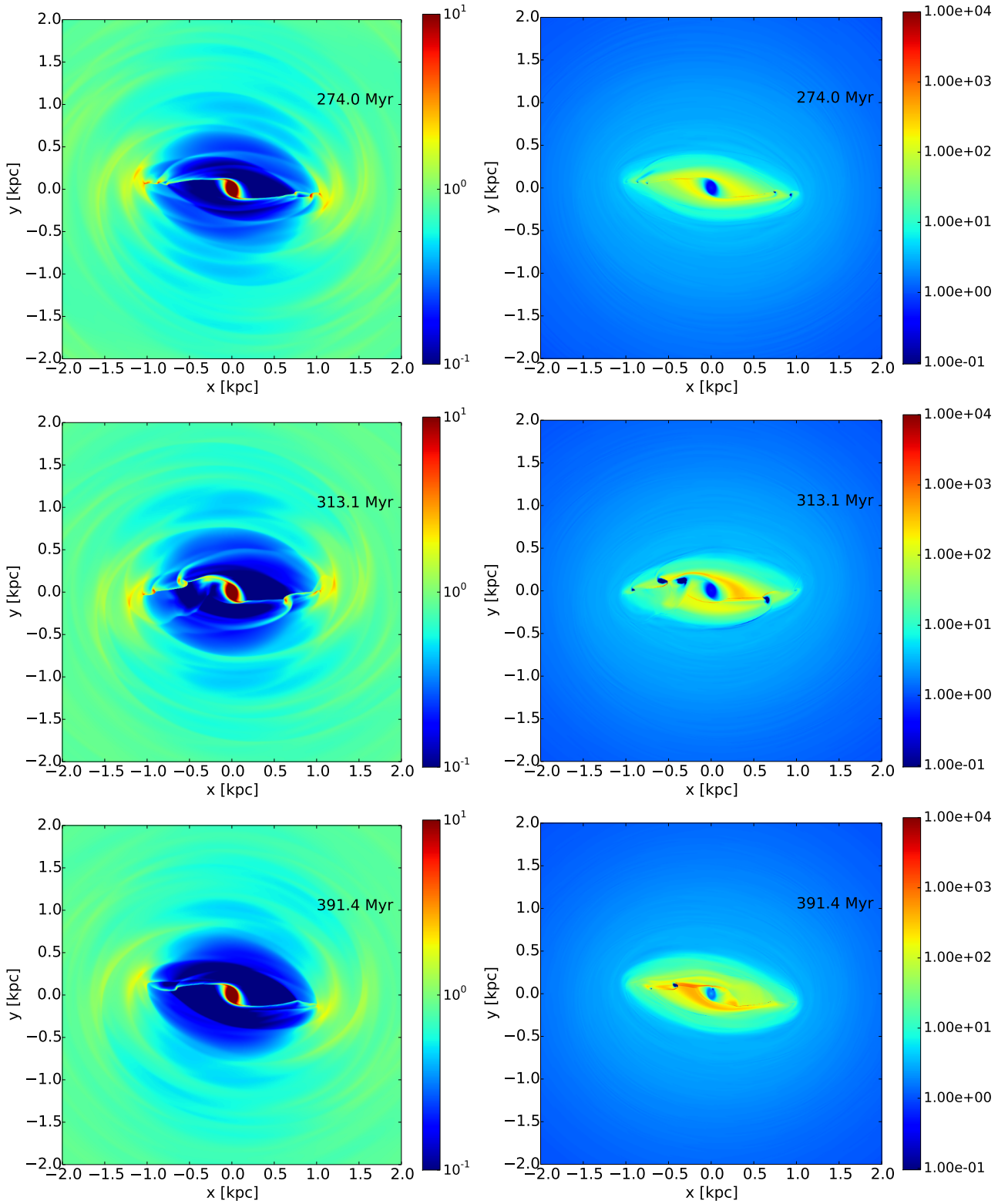


Figure 3.9: Late snapshots of a simulation using PLUTO code and the MINMOD limiter. This limiter allows the simulation to be only weakly unstable. It shows that shocks are almost formed and destroyed cyclically. Swirls produced at one shock can propagate and bump on the shock on the other side. The left panels show the gas density, while the right panels show the potential vorticity, defined in Eq. (3.4). The counter-rotating swirls are regions of the same color as the color of the lowest available value in the colorbars. Inside these swirls, the potential vorticity is negative, i.e., they are counter-rotating.

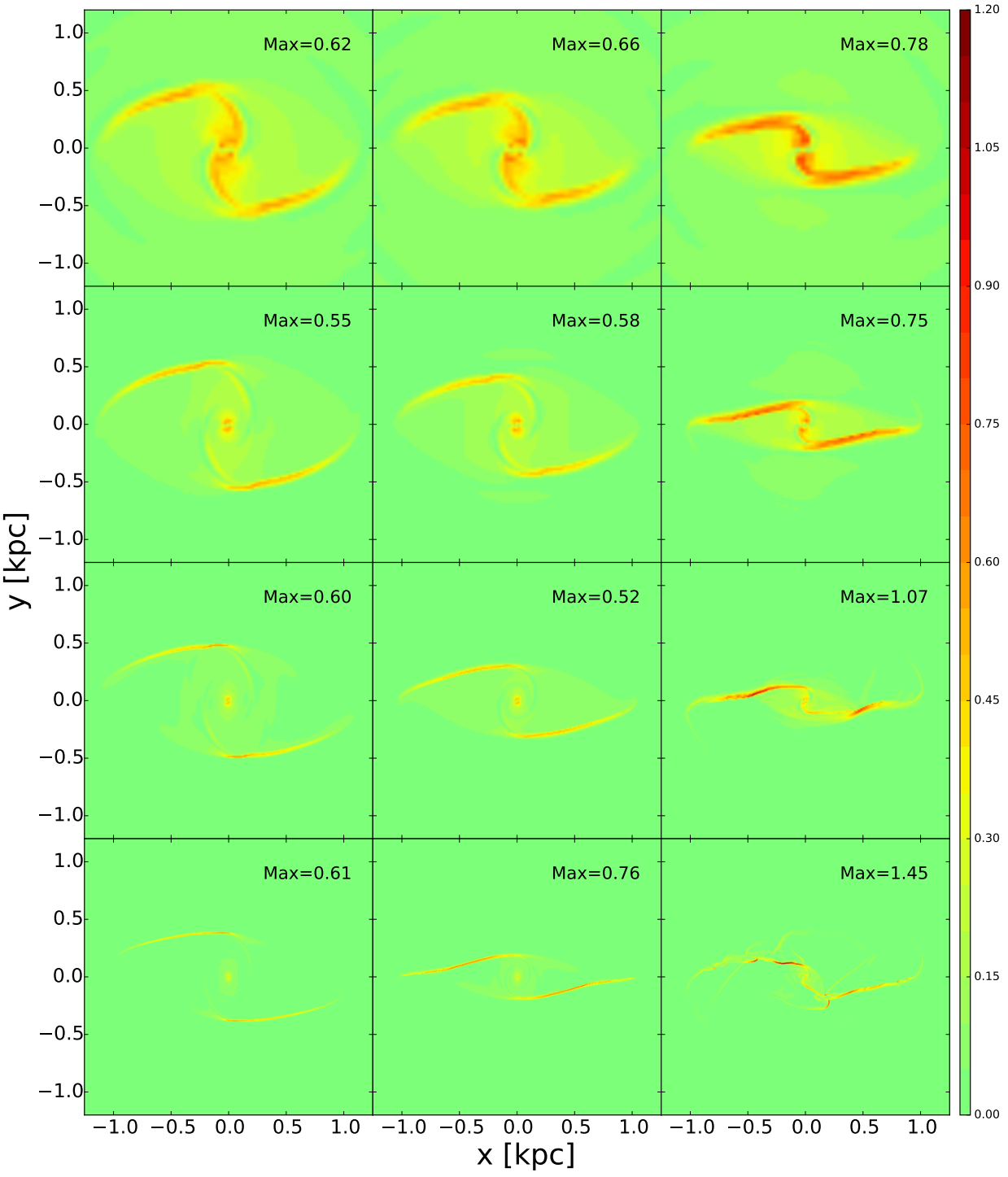


Figure 3.10: The quantity $\tau \mathrm{d}x$, where τ is an indicator of the shear and is defined in Eq. (3.5). In each panel, the maximum value reached by the quantity is shown. Panels show the same snapshots as in Fig. 3.2. Colorbar is in units of $\mathrm{km\,s^{-1}}$.

sequence of x_1 orbits as one approaches the cusped orbit, so at low sound speeds, when the shock forms far from the cusped orbit, the maximum rate of shearing is smaller than when the sound speed is larger.

A good indicator of the amount of shear in a 2D flow is the quantity (see for example Maciejewski, 2008)

$$\tau^2 = \left(\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x}\right)^2 + \left(\frac{\partial v_x}{\partial x} - \frac{\partial v_y}{\partial y}\right)^2. \tag{3.5}$$

This quantity is invariant under rotations of the coordinates, being the magnitude of the eigenvalues of the traceless shear tensor, defined by

$$D_{ij} = \frac{1}{2}\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} - \delta_{ij}(\boldsymbol{\nabla}\cdot\mathbf{v})\right). \tag{3.6}$$

Fig. 3.10 shows the quantity $\tau \mathrm{d}x$, where the shear τ is estimated by finite differences.[1] As claimed above, the shearing rate is high along the shocks and is higher for higher sound speeds.

The alert reader may have noticed that in many late snapshot where the instability is present (e.g., bottom right panels in Fig. 3.2 and all panels in Fig. 3.9) the density is not bi-symmetric (i.e., not symmetric by point reflection through the centre) even though in principle it should be, as the potential and the initial conditions are bi-symmetric. This may suggest that the instability is a numerical artefact rather than a real phenomenon. However, the fact that the bi-symmetry is not preserved is not completely unexpected: turbulence can be followed accurately over long times only by codes specifically built with this purpose in mind (e.g., see Springel & Dullemond (2012)). Neither our code nor the much more tested PLUTO (with our setup) are suited to follow the instability over such long timescales, and we should expect that small numerical errors accumulating over time eventually produce macroscopic asymmetries. Indeed, in the early stages, when the instability is forming (e.g., Fig. 3.8) the flow is, as it should, bi-symmetrical in our simulations. It only becomes not bi-symmetrical a long time after the onset of the instability.

Kim et al. (2012b) also encountered the onset of turbulence as isothermal gas moves in a rotating barred potential, and our Fig 3.8 is similar to their Fig.4, panel (c). They pointed out the process described above by which vortices generated in one shock pass to the shock on the other side of the galaxy, and are there amplified.[2] This process has been called the wiggle instability (see for example Wada & Koda, 2004; Kim et al., 2014, and references therein). Kim et al. (2014) use a shearing box analysis to show that the instability is physical. However, complete consensus has not been reached on the nature of this instability: Wada & Koda (2004) conjectured it is a Kelvin-Helmholtz (KH) type instability, but Kim et al. (2014) argued that it should be considered as a different kind of instability. Indeed, in the case of KH, we would expect the vortices to be rotating in the same sense as the rest of the flow, and not counter-rotating as we have discussed above. The instability has also been attributed to numerical noise caused by the discretisation of

[1] We show $\tau \mathrm{d}x$ rather than τ to eliminate a scaling effect. When we increase the resolution, the velocity jump near a shock stays approximately constant, while the shock width scales proportionally to $\mathrm{d}x$. As a consequence, the quantity τ estimated with finite differences scales as $1/\mathrm{d}x$. Hence, the quantity $\tau \mathrm{d}x$ avoids this scaling effect.

[2] In general, vortices can also be destroyed at shocks, but in the early stages of the instability the process of amplification is dominant.

the fluid equations (Hanawa & Kikuchi, 2012).

We conclude that most likely the unsteadiness is real, as it is not a feature of our particular hydro code, but is shared by other well tested codes as well, and, moreover, has also been confirmed by analytical calculations.

3.4.4 Numerical convergence

We have seen that as we increase the resolution at a given sound speed, the shocks are postponed and the x_2 disc shrinks. Have we converged, or would a further increase in resolution produce significant changes in the flow? In order to discuss this, let us consider how the size of the x_2 disc depends on the resolution. In each snapshot in our simulations the vertical edges of the x_2 disc can be clearly identified by a sharp variation in the gas density as we move along the vertical axis. Consequently, the vertical size of the x_2 disc is well defined.

Fig. 3.11 shows the vertical radius of the x_2 disc plotted against resolution for simulations with sound speed $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$. The data points for $\mathrm{d}x \geqslant 5\,\mathrm{pc}$ are from the simulations presented above, while the black dot at $\mathrm{d}x = 2.5\,\mathrm{pc}$ is obtained for an additional simulation. Since the computational cost of a simulation of given spatial extent rises as $(\mathrm{d}x)^{-3}$, the additional simulation was run on a grid that covered an area only $5\,\mathrm{kpc} \times 5\,\mathrm{kpc}$ in extent – we have verified that the results of lower resolution simulations for the region inside the shocks are unaffected by a reduction of the extent of the grid, so the additional simulation should provide a valid additional datum for the radius of the x_2 disc.

The simulation with $\mathrm{d}x = 2.5\,\mathrm{pc}$ was unsteady as expected, given the tendency for unsteadiness to arise at high resolutions and sound speeds. On account of the unsteadiness, the size of the x_2 disc in the $\mathrm{d}x = 2.5\,\mathrm{pc}$ simulation fluctuates. In Fig. 3.11, the point for the $\mathrm{d}x = 2.5\,\mathrm{pc}$ simulation is obtained by averaging the x_2 disc size over four different snapshots at times $t = 274, 313, 352, 391$ Myr, and the red crosses show the values in each snapshot. Such averaging would not modify the data points for the coarser, steady, simulations. Since in the simulation with $\mathrm{d}x = 2.5\,\mathrm{pc}$ the mean radius of the x_2 disc is compatible with this radius when $\mathrm{d}x = 5\,\mathrm{pc}$, we conclude that the size of the x_2 disc has converged to near its true value at $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$.

As can be seen from Fig. 3.11, the value of the vertical radius towards which we have converged for $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$ is the value of the intercept of the cusped orbit with the vertical axis. Inspection of Fig. 3.2 shows that for higher sound speeds, the x_2 disc becomes smaller. Indeed, Englmaier & Gerhard (1997) found that the x_2 disc disappears for high enough sound speed.

3.4.5 A paradox

In the limit of vanishing pressure, the characteristics of the Euler equations reduce to ballistic trajectories. Hence, we expect the ballistic approximation to fluid flow to work best when the sound speed is low. Why is the BGSBU picture, which is founded on the ballistic approximation, more accurately reproduced with $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$ than with $c_\mathrm{s} = 5\,\mathrm{km\,s^{-1}}$?

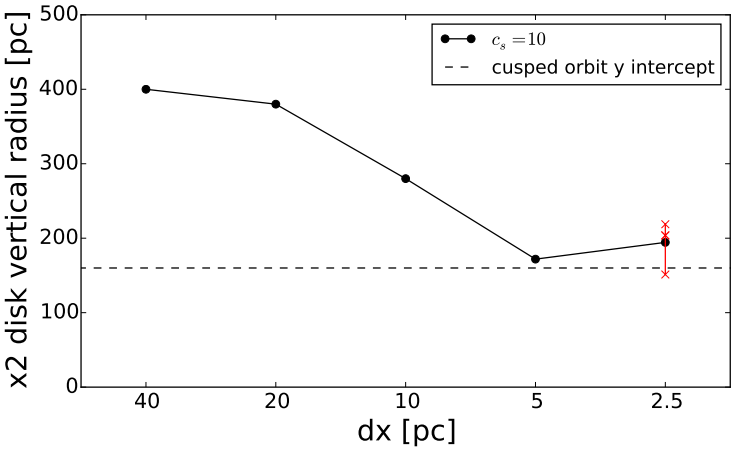


Figure 3.11: The vertical radius of the x_2 disc as a function of resolution. Each point is obtained by averaging the radius over 4 different snapshots at times $t = 274, 313, 352, 391$ Myr. For resolution $\mathrm{d}x = 2.5\,\mathrm{pc}$ the flow is unsteady and the size of the disc fluctuates, and the red crosses show the scatter in the size of the x_2 disc over the four snapshots. The horizontal dashed line shows the value of the intercept of the cusped orbit with the vertical axes.

The answer is that the BGSBU picture includes the requirement that the transition between orbit families lies close to the cusped orbit. This transition is effected by shocks, which are more rather than less likely to form in the limit of vanishing pressure. Shocks reflect the necessity in a fluid for there to be a unique velocity at each spatial point. As Riemann (1860) showed, in certain circumstances the Euler equations predict more than one velocity at a given point of a fluid in the sense that characteristics of the same family can cross. The unphysical implications of crossing characteristics are voided by a shock forming near the crossing point. In the shock the finite mean-free path of the fluid's constituent particles becomes important, and the Euler equations, which are based in the fiction that the fluid forms a continuum, cease to be valid.

In a region of the disc where there are two distinct families of closed orbits, the principle that in the limit of vanishing pressure the streamlines/characteristics will be closed orbits does not suffice to determine the flow. One needs in addition some way of selecting one orbit family for the streamlines. Since the x_1 orbits do not exist at very small radii and the x_2 orbits do not exist at very large radii, it is evident that in a barred galaxy the flow must at some point transition from x_1 to the x_2 orbits. Our simulations have shown that the location of the transition is controlled by the pressure even though the latter is too weak to be dynamically significant upstream of the transition. BGSBU argued that the transition occurs at the cusped orbit, the latest that it could occur, and with hindsight it is not surprising that this requires that the sound speed is $\gtrsim 10\,\mathrm{km\,s^{-1}}$.

3.4.6 Sound speed and the nature of the ISM

Our results can reproduce BGSBU predictions only for a particular sound speed. Does this mean we are providing a measurement of the sound speed?

In addressing this question, a major issue is that we have investigated only one potential and only a single value of the pattern speed, so we cannot exclude the possibility that in another potential the data could

be fitted with a lower value of c_s. A trivial illustration of this fact is provided by one of the axisymmetric potentials that were used to interpret (l, v) plots before BGSBU. However, the sharply peaking circular-speed curves that these potentials require to fit the HI envelope in the (l, v) plane are highly implausible, and if the mechanism proposed by BGSBU for generating this feature is to work, gas must stay close to x_1 orbits until rather close to the cusped orbit. That is, in a model with a small value of c_s, the upward rise in the HI envelope would have to be at least partly explained by inward-increasing values of the circular speed. Moreover, the cusped orbit would have to occur at smaller radii than BGSBU hypothesised.

Given the almost fractal nature of the ISM, the physical significance of the parameter c_s that plays such an important role in the simulations is far from evident. It is, however, worth noting that we have pushed the resolution so high that the cell size of our simulations is comparable to the size of a giant molecular cloud – even small GMCs are $\gtrsim 5\,\mathrm{pc}$ across.

3.5 Implications for the interpretation of observational data

3.5.1 BGSBU revisited

The three key points in BGSBU's interpretation of the observational (l, v) plots were:

1. The high velocities, reaching $v \simeq 270\,\mathrm{km\,s^{-1}}$, found in the HI (l, v) data could be explained by gas moving on x_1 orbits just outside the cusped orbit. When the (l, v) plots had been interpreted in the context of an axisymmetric Galaxy in previous studies (see for example Sofue, 2013), the high velocities observed had required the circular-speed curve to first rise, and then fall with implausible rapidity. By hypothesising a bar, BGSBU were able to explain the sharp peaks in the (l, v) envelope with a monotonically rising circular-speed curve.

2. The parallelogram they identified in the CO emission was a cut version of the longitude-velocity trace of the cusped orbit. The discrepancy between the cusped orbit and the CO parallelogram was attributed to the limits of the ballistic approximation, and it was hoped that full hydrodynamical calculations would resolve this discrepancy. In particular, BGSBU pointed to the hydrodynamical simulations of Athanassoula (1992b), which used the same numerical method that is used here, to construct a qualitative argument that qualitatively resolved the discrepancy.

3. The emission at low longitudes visible in the CS was attributed to gas on x_2 orbits. Outer x_2 orbits that can be found in the potential (see Fig. 3.1) were assumed to be unoccupied by gas.

In the previous section we confirmed that the flow of gas can be understood as a transfer from x_1 to x_2 orbits. However, the transition point is not always at the cusped orbit, as BGSBU assumed, but depends on the parameters of the gas flow, especially the sound speed. For the right sound speed, $c_s = 10\,\mathrm{km\,s^{-1}}$, and our maximum resolution, $\mathrm{d}x = 5\,\mathrm{pc}$, the gas flow is very similar to that hypothesised by BGSBU. We now

re-analyse the observational data in light of this particular simulation. Further below we take a step back and discuss interpretation of data in a broader view.

In the centre-bottom panel of Fig. 3.4 we can see the (l, v) plot corresponding to the simulation with $\mathrm{d}x = 5\,\mathrm{pc}$ and $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$. The envelope of the hydro follows that of the closed orbits very well and reproduces the high velocity peaks found in observations. Fig. 3.12 shows how the line-of-sight velocity varies in the xy plane for the assumed observing angle $\phi = 20°$. The black dots show the locations that generate the envelope of emission in the (l, v) plane. We can see that the high-velocity peaks at $|l| \simeq 2°$ arise from gas that lies very nearly along an x_1 orbit just outside the cusped orbit, as well as the portion of envelope that runs from the peak down into the zone of forbidden velocities. Thus item (1) above is nicely confirmed by this hydro simulation.

Well outside the cusped orbit the black dots in Fig. 3.12 have wiggles associated with spiral arms that run out from the extremities of the bar. These wiggles are reflected in small oscillations in the (l, v) envelope, qualitatively similar to what is observed for the envelope of real observations. We do not discuss them further here; the next chapter is entirely devoted to the study of the spiral arms that generate these wiggles.

As mentioned above, two distinct structures resembling a parallelogram play a role in the BGSBU picture: one is the CO parallelogram, presumed to be a cut version of the cusped x_1 orbit, and a second is the trace of an x_1 orbit just outside the cusped orbit, which is responsible for the velocity peaks in the envelope of HI emission. In the simulated (l, v) plots we can identify only one such parallelogram, namely the second one. How can we solve this inconsistency? Fig. 3.13 shows the shocks (blue and green) and the x_2 disc (red) in the xy plane (top) and in the (l, v) plane (below). In the lower panel the shocks closely follow the vertical sides of the cusped orbit's (black) parallelogram, but are cut almost exactly where the CO parallelogram ends. We conclude that the vertical sides of the CO parallelogram are formed by shocks. The other two sides of the CO parallelogram must be made up of gas flowing from one shock to the other. The shocks do not show up in our (l, v) projection of the gas flow, though they are quite apparent in the xy plane. In reality, we expect the shocked gas to be brighter than our simple minded projection law suggests, because a lot of atomic gas is converted to molecular as it is compressed at the shocks. When this effect is taken into account the second parallelogram should appear in the (l, v) projection of the model. Thus it is likely that item (2) above of BGSBU's interpretation can be confirmed, although the physical mechanism that causes the parallelogram to be cut was not exactly described by BGSBU. Finally, Fig. 3.13 clearly shows that in the (l, v) plane the central molecular zone (red) occupies the region of the inner x_2 orbits, as conjectured by BGSBU [item (3) above].

When the x_1/x_2 transition happens well before the cusped orbit because either the sound speed or the resolution is low, there is no gas able to explain the high velocity peaks as in item (1). Moreover, the shocks project to $|l| > 2°$ so they cannot be associated with part of the CO parallelogram, and the central molecular zone is predicted to extend to higher $|l|$ than where significant CS emission is found, compromising the interpretation of item (3). Thus several different aspects of the data point to gas remaining on x_1 orbits right up to the cusped orbit.

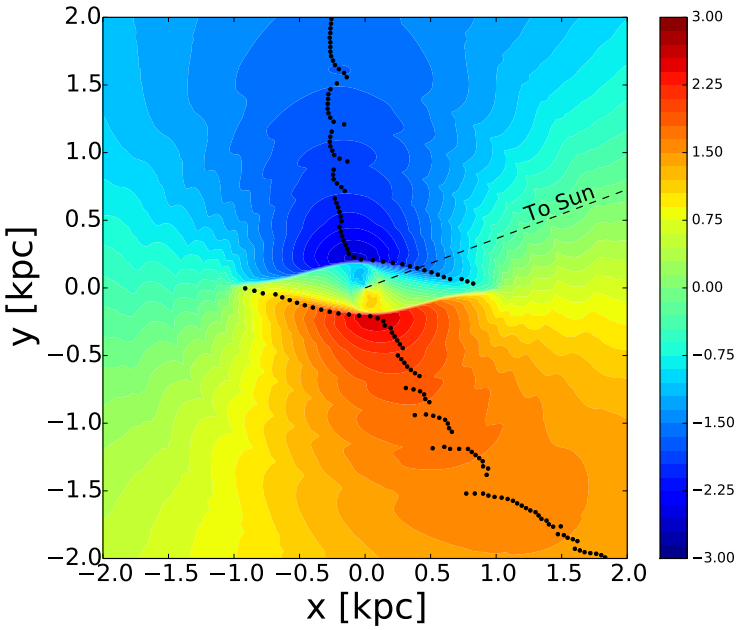


Figure 3.12: The distribution of the projected line-of-sight velocity in the xy plane for the model $\mathrm{d}x = 05\,\mathrm{pc}$, $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$. The black dots show the points corresponding to the envelope of gas in the (l, v) plane. The angle between the Sun-Galactic centre line and the bar major axis is assumed to be $\phi = 20°$. The colorbar is in units of $100\,\mathrm{km\,s^{-1}}$.

3.5.2 The asymmetry

It is well known that the molecular emission in the central molecular zone is highly asymmetric: three quarters of the ^{13}CO and CS emission comes from positive longitudes. Perspective effects cannot account for this asymmetry (Jenkins & Binney, 1994), so the observed asymmetry must be transient: observations made tens of megayears in the past or future would often show asymmetry in the opposite sense. Thus the likely explanation of the asymmetry is unsteady flow, and the principal motivation of the work reported in Jenkins & Binney (1994) was to find evidence of unsteady flow. Although low-amplitude unsteadiness was generated in their simulations, the present study implies that their simulations were too crude to probe the physically interesting regime.

We saw above that the greater prominence of the CO rectangle in the observed (l, v) plot than in the theoretical (l, v) plot suggests that the shocks are important sites for the conversion of atomic to molecular gas. This being so, unsteady flow through the shocks will give rise to unsteady conversion of atomic to molecular gas, so the atomic/molecular ratio on each side of the Centre could well fluctuate as widely as the observations imply. However, a full explanation of the observed asymmetries in molecular-gas emission must await high-resolution simulations that keep track of the chemistry of the ISM.[3]

[3]A note on the use of the term "chemistry". We use chemistry to refer to how matter is affected by the way the electrons around atoms are disposed, which includes processes such as ionisation and photo-dissociation of molecules. Here, chemical changes are meant to be principally the formation and dissociation of molecules, especially H_2 and CO. Of course, these changes will be controlled by the physical conditions of the ISM, such as the temperature, density and opacity. Thus, keeping track of the chemistry means keeping track of all the changes in chemical composition of the ISM, including those driven by a change in its physical conditions. Changes in the chemical composition of the ISM will be the dominant cause that determines the observed CO emission. However, even if we assume that the abundances of each species in the ISM are fixed, the observed emission could still vary between different regions purely because the physical conditions are different, but this is not the dominant effect here.

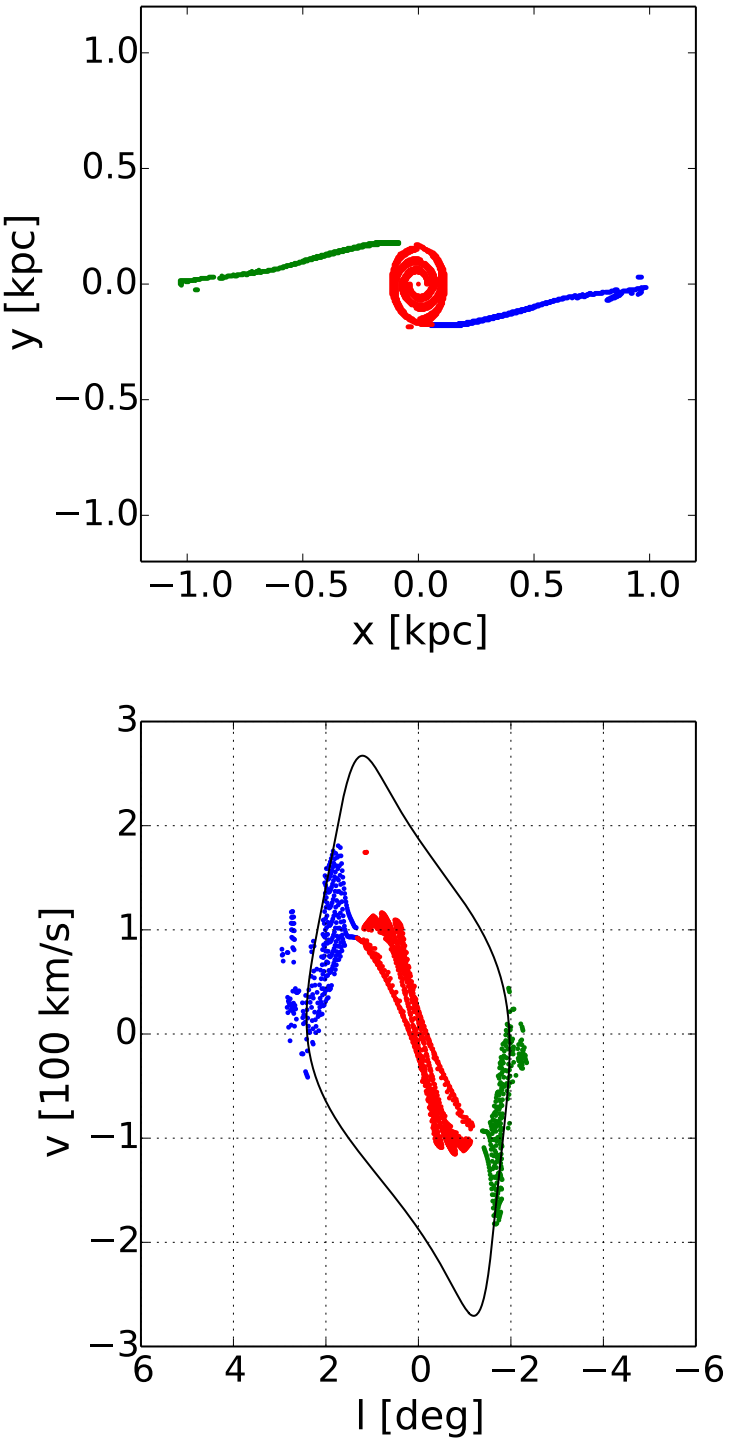


Figure 3.13: Top: points on the narrow shocks and the x_2 disc in the xy plane. Bottom: their projected position in the (l, v) plane. In black the trace of the cusped orbit is shown.

3.5.3 What do we still need to explain in the (l, v) diagram?

Two main aspects of observations in particular remain inadequately explained by the model presented in this chapter:

- Coherent broad features like the 3 kpc arm and its counterpart on the far side of the Galaxy (Dame & Thaddeus, 2008). These are not produced in our simulations. They *are* produced in other hydro simulations (Mulder & Liem, 1986; Rodriguez-Fernandez & Combes, 2008), but these simulations did not reproduce the high peaks in the envelope of HI emission in the (l, v) plane, probably for lack of resolution.

- Forbidden emission at large longitudes. The portion of the (l, v) diagram covered by forbidden emission in our simulations is smaller than the region in which coherent forbidden emission is seen in the data. A higher quadrupole moment is probably needed to reproduce this.

In addition to the above, other important elements need to be explained eventually, in particular the sharp vertical features and items 5-9 (bumps on the envelope, tilt, clumpiness, asymmetry and variation by species) in the list of Section 2.2. However, these are different from the two items above, in the sense that they probably cannot be explained simply by taking the models of this chapter and changing the underlying potential, as we will do for the two items above in Chapter 5, but they require new physics to be implemented into the models. For example, the tilt clearly cannot be explained by two-dimensional models. For the vertical features, see also Section 5.4.

The essential elements of the BGSBU picture are that streamlines coincide with x_1 and x_2 orbits, and that the shocks responsible for the transition lie near the cusped orbit. BGSBU illustrated these principles with one particular, very simple potential. Better fits to the data could surely be obtained with other, similar potentials. Indeed, in Chapter 5 we show models that keep the essential elements of the BGSBU picture while remedying the above shortcomings.

3.5.4 Relation to prior work

The question of what physical mechanism determines the size of the x_2 disc is relevant for the interpretation of observations in our and external galaxies (see for example Combes, 1996; Kim et al., 2012a). Our results suggest that some previous studies may be biased by not taking into account the effects of varying the resolution. For example, Cole et al. (2014) in their simulations of galaxy formation found that the main mismatch between their models and the observations was that the nuclear discs of their models were too big relative to their bars. Since nuclear discs are x_2 discs, our results suggest that the mismatch would be resolved by an increase in resolution.

Finally, we mention that our findings do support the hopes of Bissantz et al. (2003). These authors found, as in our low-resolution simulations, that the innermost non-self intersecting x_1 orbits near the cusped orbit were unoccupied, and attributed this fact to details of the SPH scheme, which do not apply for our

grid-based simulations. The work presented here gives strong support to their view that in higher-resolution, grid-based simulations their inner x_1 orbits would be occupied by gas, giving rise to peaks in their (l, v) projections.

3.6 Conclusion

Binney et al. (1991) (BGSBU) constructed a picture of the flow of gas through the central few kiloparsecs of our Galaxy. Their picture was based on the idea that gas follows closed orbits, and it involved a particular choice of orbit at which the gas transitions from the x_1 orbit family to the x_2 family. Their orbit-based picture required validation by hydrodynamical simulations of gas flow. Early efforts in this direction did not provide the necessary validation, in part because they did not adopt the same Galactic potential as BGSBU, but largely because in them gas occupied some x_2 orbits that BGSBU required to be empty, and left empty some x_1 orbits that BGSBU required to be occupied. We have run high-resolution, grid based hydro simulations of gas flow in the potential of BGSBU and validated their picture in the case that the effective sound speed in the ISM is $c_\mathrm{s} \gtrsim 10\,\mathrm{km\,s^{-1}}$. This is the first time the BGSBU picture has been validated by hydrodynamical simulations.

The simulations confirm that, regardless of the sound speed adopted and the grid resolution employed, gas streamlines closely coincide with closed orbits everywhere outside a shock-dominated transition region that divides the outer region, in which gas follows x_1 orbits, from an inner region, in which gas follows x_2 orbits. Previous simulations by other authors already found that gas flow in hydro simulations is following x_1 and x_2 orbits, but to our knowledge none showed precisely to what extent the velocity fields predicted by the orbits and the hydro actually differ and which streamlines were occupied by gas (see also Chapter 4, in particular Fig. 4.11). Doing this with precision is necessary in connection with the observed high velocity peaks, the explanation of which is highly sensitive to which orbits are populated in the vicinity of the cusped orbit. Previous work by other authors did not have the necessary level of detail.

The orbit at which the shock arises, and the transfer commences, depends on both the sound speed and the grid's resolution. The dependence on the sound speed was already discussed by Englmaier & Gerhard (1997), while the dependence on the resolution is a new effect in the light of which some conclusions from previous authors should be revised. Shock formation is favoured by both low sound speeds and low grid resolutions, so increasing the sound speed and/or the grid resolution moves inward the shock that causes gas to plunge from x_1 to x_2 orbits. The BGSBU picture calls for the shock to occur as close to the Galactic centre as it logically can, namely at the cusped orbit, interior to which x_1 orbits become self-intersecting. Consequently, a flow consistent with the BGSBU picture cannot be obtained with either a low sound speed or poor spatial resolution. It seems that previous simulations lacked the requisite resolution. We find that a consistent flow can be obtained for $c_\mathrm{s} \simeq 10\,\mathrm{km\,s^{-1}}$ and grid spacing $\mathrm{d}x \lesssim 5\,\mathrm{pc}$.

BGSBU did not provide a satisfactory explanation of the parallelogram-like structure of CO emission in the (l, v) plane. We find that the shocks form two sides of the CO parallelogram and conjecture that

the prominence of the CO parallelogram is due to efficient conversion of atomic gas into molecular gas. Unfortunately, we do not follow the ISM's chemistry.

In our highest-resolution simulations the flow in the transition region between the x_1 and x_2 orbits is unsteady. A similar unsteadiness was found by other authors. We have run further tests that, combined with other authors findings, lead us to believe that this unsteadiness is probably a real physical phenomenon rather than a computational artefact, and is essentially turbulence due to the production of vorticity at the shocks, which are regions of high shear. We propose that this unsteadiness, in conjunction with efficient conversion of atomic gas to molecular form in the shocks, provides a promising explanation of the observed asymmetry in the distribution of CO emission either side of the Galactic Centre.

While our simulations do provide strong support for the BGSBU picture, they do not explain all aspects of the observed HI and CO emission. There is, however, every prospect that further high-resolution simulations of flows in potentials similar to that used by BGSBU will explain all significant features. In this connection, two very worthwhile upgrades of our simulations would be an increase in the quadrupole moment of the bar, and inclusion of the conversion of gas between atomic and molecular forms. In Chapter 5 we perform the first of these upgrades.

References

Athanassoula E., 1992a, MNRAS, 259, 328
Athanassoula E., 1992b, MNRAS, 259, 345
Athanassoula E., Bureau M., 1999, ApJ, 522, 699
Baba J., Saitoh T. R., Wada K., 2010, PASJ, 62, 1413
Binney J., 1974, MNRAS, 168, 73
Binney J., Gerhard O. E., Stark A. A., Bally J., Uchida K. I., 1991, MNRAS, 252, 210
Binney J., Merrifield M., 1998, Galactic Astronomy. Princeton University Press
Bissantz N., Englmaier P., Gerhard O., 2003, MNRAS, 340, 949
Blitz L., Spergel D. N., 1991, ApJ, 379, 631
Bureau M., Athanassoula E., 1999, ApJ, 522, 686
Cole D. R., Debattista V. P., Erwin P., Earp S. W. F., Roškar R., 2014, MNRAS, 445, 3352
Combes F., 1996, in Astronomical Society of the Pacific Conference Series, Vol. 91, IAU Colloq. 157: Barred Galaxies, Buta R., Crocker D. A., Elmegreen B. G., eds., p. 286
Dame T. M., Thaddeus P., 2008, ApJ, 683, L143
Englmaier P., Gerhard O., 1997, MNRAS, 287, 57
Englmaier P., Gerhard O., 1999, MNRAS, 304, 512
Ferrière K. M., 2001, Reviews of Modern Physics, 73, 1031
Fux R., 1999, A & A, 345, 787
Hanawa T., Kikuchi D., 2012, in Astronomical Society of the Pacific Conference Series, Vol. 459, Numerical Modeling of Space Plasma Slows (ASTRONUM 2011), Pogorelov N. V., Font J. A., Audit E., Zank G. P., eds., p. 310
Jenkins A., Binney J., 1994, MNRAS, 270, 703
Kim W.-T., Kim Y., Kim J.-G., 2014, ApJ, 789, 68
Kim W.-T., Seo W.-Y., Kim Y., 2012a, ApJ, 758, 14
Kim W.-T., Seo W.-Y., Stone J. M., Yoon D., Teuben P. J., 2012b, ApJ, 747, 60
Lee C. W., Lee H. M., Ann H. B., Kwon K. H., 1999, ApJ, 513, 242
Maciejewski W., 2008, in IAU Symposium, Bureau M., Athanassoula E., Barbuy B., eds., Vol. 245, pp. 161–164
Magorrian J., 1995, Stellar dynamical models of disky elliptical galaxies. D.Phil thesis, University of Oxford
Mignone A., Bodo G., Massaglia S., Matsakos T., Tesileanu O., Zanni C., Ferrari A., 2007, ApJS, 170, 228

Mulder W. A., Liem B. T., 1986, A & A, 157, 148
Patsis P. A., Athanassoula E., 2000, A & A, 358, 45
Pettitt A. R., Dobbs C. L., Acreman D. M., Price D. J., 2014, MNRAS, 444, 919
Riemann G. F. B., 1860, Abh. Konigl. Gesell. Wiss. Göttingen., 43-65, 192
Rodriguez-Fernandez N. J., Combes F., 2008, A & A, 489, 115
Roe P. L., 1986, Annual Review of Fluid Mechanics, 18, 337
Sod G. A., 1978, Journal of Computational Physics, 27, 1
Sofue Y., 2013, Mass Distribution and Rotation Curve in the Galaxy, Oswalt T. D., Gilmore G., eds., p. 985
Springel V., Dullemond C., 2012, Numerical fluid mechanics lecture notes (`http://www.ita.uni-heidelberg.de/~dullemond/lectures/num_fluid_2012`)
van Albada G. D., van Leer B., Roberts, Jr. W. W., 1982, A & A, 108, 76
van Albada T. S., Sanders R. H., 1982, MNRAS, 201, 303
van Leer B., 1977, Journal of Computational Physics, 23, 263
Wada K., Koda J., 2004, MNRAS, 349, 270
Weiner B. J., Sellwood J. A., 1999, ApJ, 524, 112

CHAPTER 4

Bar Driven Spiral Arms

4.1 Introduction

Many grand design barred spirals exhibit spiral arms starting at and extending out from the end of the bar (e.g. Dobbs & Baba, 2014, Section 2.3). The intuitive interpretation of this observational fact is that the bar and the spiral arms rotate at the same pattern speed, but it is still unclear whether this is really the case and there are other common interpretations in which the bar and the spiral arms have different pattern speed (Dobbs & Baba, 2014). This type of spiral arms is relevant for the interpretation of observations of both our Galaxy and external galaxies. For example, it is likely that some features present in spectral-line emission (see Chapter 2) of atomic and molecular gas in the inner Galaxy, for instance the "3-kpc arm", are produced by such arms (see Chapter 5). It is therefore important to have a physical understanding of why such spiral arms arise. In this chapter, we focus on the intuitive interpretation that bar and spiral arms have a common pattern speed.

How are these spiral arms generated? The natural interpretation is that they are driven by the bar.[1] On the theoretical side, the possibility of bar-driven spiral arms is now well established. The first investigations used analytical methods and were carried out by Feldman & Lin (1973) and Lin & Lau (1975). Then, in the following years, Sanders & Huntley (1976), Sanders (1977) and Huntley et al. (1978) performed pioneering numerical experiments and discussed the physical mechanism responsible for the generation of the spiral arms in terms of closed orbits. Their work was later extended independently by Wada (1994) and Lindblad & Lindblad (1994), and more recently by Piñol-Ferrer et al. (2012). These authors constructed phenomenological analytical models under the epicyclic approximation with the purpose of understanding the spiral arms. In their models gas parcels follow weakly oval orbits, whose major axes orientation changes

[1] We note that, while in this chapter we follow the nomenclature of Dobbs & Baba (2014), Section 2.3, Sanders & Huntley (1976), Huntley et al. (1978), Wada (1994) and use the term bar-driven spirals to also refer to spirals inside corotation, it is common in the astrophysical community to reserve the term bar-driven spirals only for spiral arms located outside corotation.

with radius giving rise to kinematic density waves à-la Lindblad, and hence to spiral arms.

An alternative viewpoint which does not consider the spiral arms as density waves has also been discussed in a series of papers by Romero-Gómez et al. (2006, 2007) and Athanassoula et al. (2009a,b, 2010). Their theory, which focuses more on stellar arms rather than gaseous ones, is based on the observation that orbits in the vicinity of unstable Lagrangian points can be trapped into invariant manifolds whose morphology can reproduce the spiral arms.

Many authors have noted the presence of bar-driven spiral arms in simulations, and discussed them in a more or less descriptive way (see for example Athanassoula, 1992; Englmaier & Gerhard, 1999; Bissantz et al., 2003; Rodriguez-Fernandez & Combes, 2008). For a recent review see also Section 2.3 of Dobbs & Baba (2014). Note that in all works cited above the gas is driven by an external potential generated by the bar and is not self-gravitating, and therefore the spiral arms are not spiral density waves in the sense of Lin & Shu (1964). We do not discuss self-gravity in this work.

The mainstream viewpoint today is that spiral arms can be understood as density waves, but this has never been tested with sufficient detail. On modern computers, it is very cheap to run simulations able to test the predictions and the limits of the phenomenological models cited above. Moreover, the literature only addresses the weak bar regime under the epicyclic approximation, and does not discuss how the picture should be extended to the strong bar case.

In this chapter we investigate the physical mechanism responsible for the generation of the spiral arms and in particular we resume the discussion of how they can be understood as kinematic density waves. The structure is as follows. In Sec. 4.2 we briefly review previous work aimed at understanding in a phenomenological fashion the spiral arms in the epicyclic approximation. Then we run grid-based, isothermal, non-self gravitating 2D hydrodynamic simulations in an externally imposed rigidly rotating barred potential, addressing both the weak and the strong bar case. The numerical methods employed are similar to those used in Chapter 3, and the small differences from there are explained in Section 4.3. In Sec. 4.4 we compare the results of the simulations with the phenomenological models available in the literature. In Sec. 4.5 we discuss in more detail the weak bar case, and in Sec. 4.6 how the results for the weak bar case should be extended to the strong bar case. Finally, we summarise our findings in Sect. 4.7.

4.2 Review of previous work

In this section we briefly review previous work aimed at understanding the bar-driven spiral arms in the epicycle approximation. In Sec. 4.2.1 we discuss the equations describing ballistic closed orbits in the epicycle approximation (see for example Sect. 3.3.3 in Binney & Tremaine, 2008). In Sec. 4.2.2 we discuss how various authors have modified the ballistic equations in order to describe gaseous parcels, and the phenomenological models based on these equations. These phenomenological models will be in later sections compared with the results of hydro simulations.

4.2.1 Ballistic closed orbits in the epicycle approximation

Consider a rigidly rotating external potential of the following form:

$$\Phi(R,\theta) = \Phi_0(R) + \Phi_1(R,\theta), \tag{4.1}$$

where Φ_0 is an axisymmetric potential and Φ_1 is a small but otherwise arbitrary perturbation. The potential is assumed to be rigidly rotating at pattern speed Ω_p. R, θ are polar coordinates in the rotating frame, with $\theta = 0$ corresponding to the positive horizontal axis, and θ increasing clockwise. The equation of motion for a ballistic particle in this potential is

$$\ddot{\mathbf{x}} = -\nabla\Phi + \Omega_\mathrm{p}^2\mathbf{x} - 2\Omega_\mathrm{p}\left(\hat{\mathbf{e}}_z \times \dot{\mathbf{x}}\right), \tag{4.2}$$

where $\mathbf{x}$ is the position vector of the particle in the rotating frame and $\hat{\mathbf{e}}_z$ is the unit vector perpendicular to the plane. The first term on the RHS represents gravitational forces, the second term is the centrifugal force and the third is the Coriolis force.

We want to find closed orbits in the above potential under the epicycle approximation. In the absence of the perturbation Φ_1 the potential is axisymmetric and the only stable closed orbits are circular orbits. These circular orbits will be described in polar form as

$$R(t) = R_0 \tag{4.3}$$

$$\theta(t) = \theta_0(t) = \Omega_\mathrm{f} t, \tag{4.4}$$

where

$$\Omega_\mathrm{f} = \Omega - \Omega_\mathrm{p}, \tag{4.5}$$

and $\Omega = \Omega(R_0)$ is the angular velocity for circular motion at radius R_0 in the potential Φ_0. Now consider the situation in the presence of the small perturbation Φ_1. We want orbits that are closed in the rotating frame.[2] We expect that far from resonances closed orbits will be weak oval deformations of the circular orbits found in Φ_0. We look for closed orbits of the form:

$$R(t) = R_0 + R_1(t) \tag{4.6}$$

$$\theta(t) = \theta_0(t) + \theta_1(t), \tag{4.7}$$

where all quantities with subscript 1 are to be considered small. By expressing Eq. (4.2) in polar components, then substituting Eqs. (4.6), (4.7) and (4.1) into it and approximating to first order in quantities with

[2] Recall that the notion of "closureness" of an orbit is frame dependent. Orbits that are closed in the inertial frame are in general not closed in the rotating frame and vice-versa.

subscript 1, the equations of motion take the following form:

$$\ddot{R}_1 + \left(\frac{\mathrm{d}^2\Phi_0}{\mathrm{d}R^2} - \Omega^2\right)_{R_0} R_1 - 2R_0\Omega\dot{\theta}_1 = -\left(\frac{\partial\Phi_1}{\partial R}\right)_{R_0,\theta_0(t)} \tag{4.8}$$

$$\ddot{\theta}_1 + 2\Omega\frac{\dot{R}_1}{R_0} = -\frac{1}{R_0^2}\left(\frac{\partial\Phi_1}{\partial\theta}\right)_{R_0,\theta_0(t)}, \tag{4.9}$$

where the derivatives of Φ_1 are evaluated along the unperturbed trajectory $[R_0, \theta_0(t)]$ and are therefore to be considered given functions of time. Exploiting the fact that $\theta_0(t)$ is a linear function of time, Eq. (4.9) can be integrated immediately to obtain $\dot{\theta}_1$, which can then be substituted into Eq. (4.8). This gives:

$$\boxed{\ddot{R}_1 + \kappa_0^2 R_1 = f_1(t)}, \tag{4.10}$$

where

$$\kappa_0^2 = \left(R\frac{\mathrm{d}\Omega^2}{\mathrm{d}R} + 4\Omega^2\right)_{R_0} \tag{4.11}$$

$$f_1(t) = -\left[\left(\frac{\partial\Phi_1}{\partial R}\right) + \frac{2\Omega}{\Omega_\mathrm{f} R_0}\Phi_1\right]_{R_0,\theta_0(t)}. \tag{4.12}$$

Eq. 4.10 is the equation of a forced harmonic oscillator. The natural frequency of this oscillator is κ_0, the usual epicycle frequency that can be calculated from Φ_0. The driving force is $f_1(t)$, and as one would expect it is due to the perturbing potential Φ_1 and reduces to zero when this is turned off. $f_1(t)$ is a periodic function of time. If we expand it in its Fourier components with respect to time, we will find the same components contained in the Fourier expansion of Φ_1 with respect to θ. This is not surprising, as the perturbed potential encountered by the guiding centre of the ballistic particle is given by $\Phi_1(R_0, \theta_0(t))$ where $\theta_0(t)$ increases linearly with time.

The general solution of Eq. (4.10) could in principle be easily written down. For each solution, we could eliminate t from Eqs. (4.6), (4.7) to find the orbit in the form $R = R(\theta)$. Not all solutions of Eq. (4.10) correspond in general to closed orbits in the rotating frame. Some particular solutions describe closed orbits, while the others in general describe non closed loop orbits. The first type of solution are the ballistic closed orbits in the epicycle approximation.

For concreteness, let us now consider the particular form of the potential used by Sanders (1977) and Wada (1994):

$$\Phi_0(R) = -\frac{av_0^2}{\sqrt{R^2 + a^2}} \tag{4.13}$$

$$\Phi_1(R,\theta) = \Phi_\mathrm{b}(R)\cos(2\theta) \qquad \Phi_\mathrm{b}(R) = -\epsilon\frac{(av_0R)^2}{(R^2+a^2)^2}, \tag{4.14}$$

where a, v_0 and ϵ are constant parameters. The axisymmetric part Φ_0 is a Kuzmin-Toomre potential, and the perturbation Φ_1 was introduced by Sanders (1977). Following Wada (1994), we choose $a = 1\,\mathrm{kpc}$ and

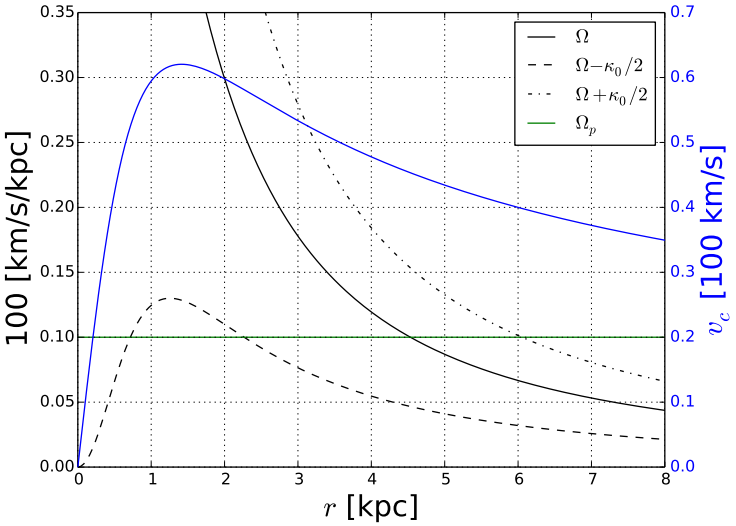


Figure 4.1: Blue curve: circular speed curve for the potential in Eq. (4.13) with values of parameters indicated in the text. Black curves: behaviour of $\Omega - n\kappa_0/2$ for the same potential. The horizontal green line indicates the assumed value of the pattern speed.

$v_0 = 100\,\mathrm{km\,s^{-1}}$ and a pattern speed of $\Omega_\mathrm{p} = 10\,\mathrm{km\,s^{-1}\,kpc^{-1}}$. Different values of ϵ will be considered in the following. Fig. 4.1 shows the circular speed curve and the behaviour of $\Omega - n\kappa_0/2$, from which we can determine the location of the resonances, for this choice of values of the parameters. For this potential, Eq. (4.10) becomes:

$$\boxed{\ddot{R}_1 + \kappa_0^2 R_1 = f_\mathrm{w}\cos(2\Omega_\mathrm{f} t)}, \tag{4.15}$$

where

$$f_\mathrm{w}(R_0) = -\left[\left(\frac{\mathrm{d}\Phi_\mathrm{b}}{\mathrm{d}R}\right) + \frac{2\Omega}{\Omega_\mathrm{f} R_0}\Phi_\mathrm{b}\right]_{R_0}. \tag{4.16}$$

The forcing is now a periodic function of time with frequency $2\Omega_\mathrm{f}$. By solving (4.15), it is found that ballistic closed orbits have the major axis always either perpendicular or parallel to the major axis of the bar.[3] The orientation changes abruptly at each Inner Lindblad Resonance (ILR), at Corotation (CR), and at each Outer Lindblad Resonance (OLR). Note however that the above analysis is not valid in the vicinity of these points: at all Lindblad resonances the natural frequency of the oscillator κ_0 tends to the forcing frequency $2\Omega_\mathrm{f}$, while at CR the forcing f_w becomes infinite. The orientation of the ballistic closed orbits will be important in the next subsection.

[3]These correspond to orbits obtained by taking the limit $\lambda \to 0$ of the solution that neglect transients of Eq. (4.17) below, which are also the orbits with $C_1 = 0$ in the notation of Sect. 3.3.3 in Binney & Tremaine (2008). It should be noted that other closed orbits solutions are possible at special radii, i.e. when the ratio $\kappa_0/(\Omega - \Omega_\mathrm{p})$ happens to be a rational number, and correspond to orbits with $C_1 \neq 0$. However these are isolated cases that are not part of a family of closed orbits parenting non closed orbits, and therefore are not of interest here.

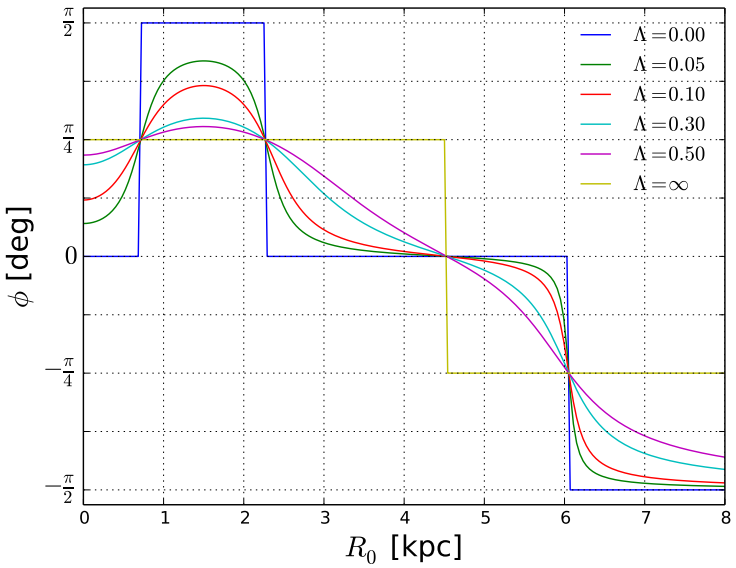


Figure 4.2: Predictions of Wada's model for the phase ϕ as function of guiding-centre radius R_0 for the potential in Eq. (4.13). ϕ represents the orientation of the major axis of a closed orbit. Curves corresponding to different values of the damping parameter Λ are shown. The possible values allowed by the phenomenological model are enclosed between the blue and yellow lines.

4.2.2 Gaseous closed orbits in the epicycle approximation: phenomenological models

In the previous section we have seen that for the particular potential given by Eqs. (4.13) and (4.14) the significant ballistic closed orbits are always oriented either perpendicularly or parallel to the major axis of the bar, with abrupt changes of orientation at each resonance. However, the discussion was only valid for ballistic particles. How should we modify it to describe the motion of a gaseous parcel which is part of a continuous fluid? Sanders & Huntley (1976) suggested that in the gaseous case the orientation shift between horizontally and vertically aligned orbits happens gradually, and that this gives rise to kinematic spiral arms à-la Lindblad. Moreover, they noted that the addition of a dissipation term[4] to Eq. (4.15),

$$\boxed{\ddot{R}_1 + 2\lambda\dot{R}_1 + \kappa_0^2 R_1 = f_\mathrm{w}\cos\left(2\Omega_\mathrm{f} t\right)}, \tag{4.17}$$

can remove the divergence of the radial oscillation amplitude at the Lindblad resonaces, and that the solution of Eq. (4.17) where transients[5] are neglected always describes a closed orbit. They also noted that at each ILR, the closed orbit solution so obtained has a major axis inclined at $\pi/4$ independently of the value of λ. Since $\pi/4$ is halfway between horizontally and vertical elongated orbits, this corroborated the conjecture that in the gaseous case we expect a gradual shift from vertically elongated to horizontally elongated when we cross a Lindblad resonance, which is the key to the generation of spiral arms in their model.

[4] In the work of Sanders & Huntley (1976) λ is a constant, but in subsequent development by other authors discussed later in this section $\lambda = \lambda(R_0)$.

[5] By transients we mean the part of the solution that vanishes in the limit $t \to \infty$. In this case we have a driven and damped harmonic oscillator. The general solution is the sum of a decaying exponential and an oscillatory term, and the decaying exponential is the transient.

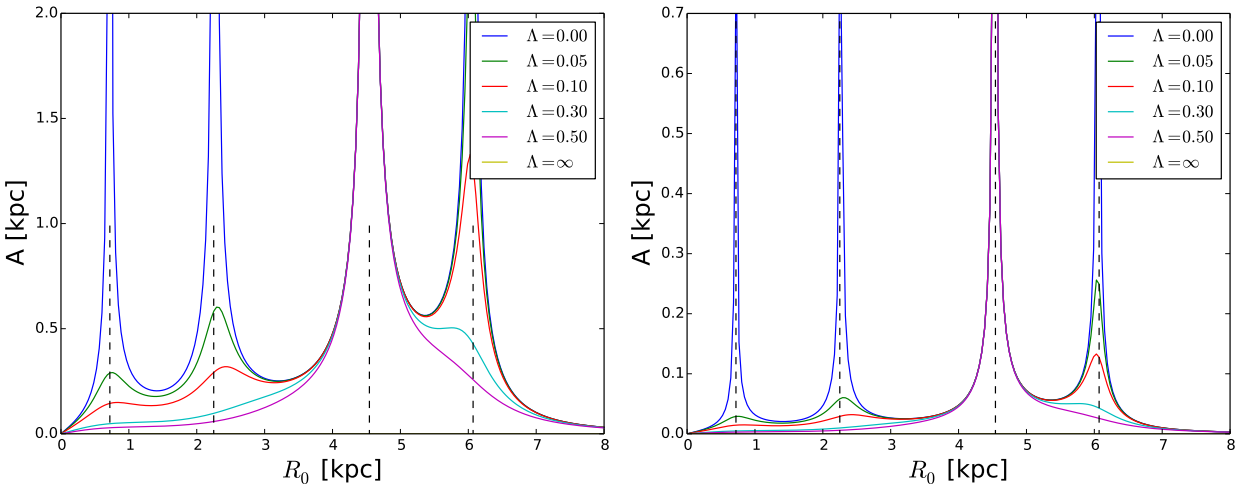


Figure 4.3: Amplitude of the radial oscillations predicted by Wada's model for two different values of the bar strength ϵ. The left panel is for $\epsilon = 0.05$, and the right panel is for $\epsilon = 0.005$. In each panel, curves corresponding to different values of the damping parameter Λ are shown. In absence of a damping term, $\Lambda = 0$, the amplitude diverges at all Lindblad resonances and at corotation. The presence of a damping term makes the amplitude finite at the Lindlblad resonances, but not at corotation.

Wada (1994) and Lindblad & Lindblad (1994) (see also Piñol-Ferrer et al., 2012) independently developed further this idea by taking more seriously the dissipation term, considering it as a phenomenological model that describes the trajectory of gaseous particles over extended regions. The approaches of Wada (1994) and Lindblad & Lindblad (1994) differ slightly in the way they implement the dissipation term. The phenomenological model of Wada (1994) is based exactly on Eq. (4.17). This amounts to add a dissipation term for the motion of the particle in the radial direction only. Lindblad & Lindblad (1994) added a dissipation term that is proportional to the difference between the velocity of the particle and the local circular speed in the underlying axisymmetric potential. Here, we consider in more detail the variant of Wada (1994), but the results of the Lindblad & Lindblad (1994) variant are qualitatively similar, and a more detailed account of the results using this variant can be found in Piñol-Ferrer et al. (2012).

Let us now briefly review Wada's model. The solution of Eq. (4.17) that excludes transient terms yields closed orbits of the following form:

$$R(\theta) = R_0 + A\cos\left(2\theta - 2\phi\right), \tag{4.18}$$

where

$$A = \frac{|f_{\rm w}|}{F} \tag{4.19}$$

$$\begin{aligned} \sin 2\phi &= \frac{4\lambda|\Omega_{\rm f}|}{F}\mathrm{sign}(f_{\rm w}) \\ \cos 2\phi &= \frac{\kappa_0^2 - (2\Omega_{\rm f})^2}{F} \end{aligned} \qquad -\pi \leqslant 2\phi \leqslant \pi. \tag{4.20}$$

$$F = \sqrt{\left[\kappa_0^2 - (2\Omega_{\rm f})^2\right]^2 + (4\lambda\Omega_{\rm f})^2}. \tag{4.21}$$

The phase ϕ encodes the information about the orientation of the major axis of the closed orbit, and is the inclination of the major axis of the orbit with respect to the horizontal axis. From Eq. (4.20) we see that the quadrant 2ϕ belongs to is determined by the sign of $f_{\rm w}$, which flips at CR, and by the sign of $\kappa_0^2 - 2\Omega_{\rm f}^2$, which flips at each Lindblad resonance. The phase ϕ in Eq. (4.18) is determined for all orbits once $\lambda = \lambda(R_0)$ is given at each radius. Since λ has the dimension of a frequency, it is convenient to express it as a multiple of a characteristic frequency at that radius. Following Wada (1994), we express λ as a multiple of the epicycle frequency at that radius,

$$\lambda \equiv \Lambda\kappa_0. \tag{4.22}$$

Fig. 4.2 reproduces Fig. 2 of Wada (1994) and shows his model predictions for the phase ϕ for different constant values of Λ. In the limit $\Lambda \to 0$ we recover the ballistic case, while in the limit $\Lambda \to \infty$ all orbits are circular. The curves $\Lambda = 0$ and $\Lambda = \infty$ bound the possible values that the ϕ can assume in its model. Note that this plot does not depend on the strength of the bar potential ϵ, because the phase ϕ does not depend on the magnitude of $f_{\rm w}$ but only on its sign.

A is the amplitude of radial oscillations. Fig. 4.3 shows the amplitude A of the radial oscillations predicted by the model for various values of Λ and two different values of ϵ. As can be seen from this figure, A always diverges at CR, regardless of the value of Λ, but diverges at the Lindblad resonances only in the absence of the damping term. Note also that, for a given ϵ, away from resonances the value of A is limited from above: the maximum value that A can reach at a given R_0 is limited by its value for $\Lambda = 0$. Finally, note that for a given potential, this theory has only one adjustable parameter at each radius, the damping λ. Both ϕ and A at each radius depend on a single phenomenological parameter in this theory.

Using the theory outlined above, it is possible to construct explicit models of the spiral arms. Fig. 4.4, top panels, reproduces the model shown in Fig. 4 of Wada (1994). It shows a nested sequence of closed orbits for values of the parameters $\epsilon = 0.05$ and $\Lambda = 0.05$. The bottom panels show another model, for a weaker bar with $\epsilon = 0.005$ and $\Lambda = 0.02$.

4.3 Numerical methods

In our simulations, we use the same methods described in Sect. 3.2. We very briefly review the main points here. We assume that the gas is a fluid governed by the Euler equations complemented by the equation of state of an isothermal ideal gas. Then we run two-dimensional hydrodynamical simulations in an externally imposed, rigidly rotating barred potential. The grid-based, Eulerian code is based on the second-order flux-splitting scheme developed by van Albada et al. (1982) and later used by Athanassoula (1992), Weiner & Sellwood (1999) and others to study gas dynamics in bar potentials.

We also implement the same recycling law as in Sect. 3.2, with the same values of the parameters. As we have seen, the recycling law introduces a term in the continuity equation that was originally meant to take

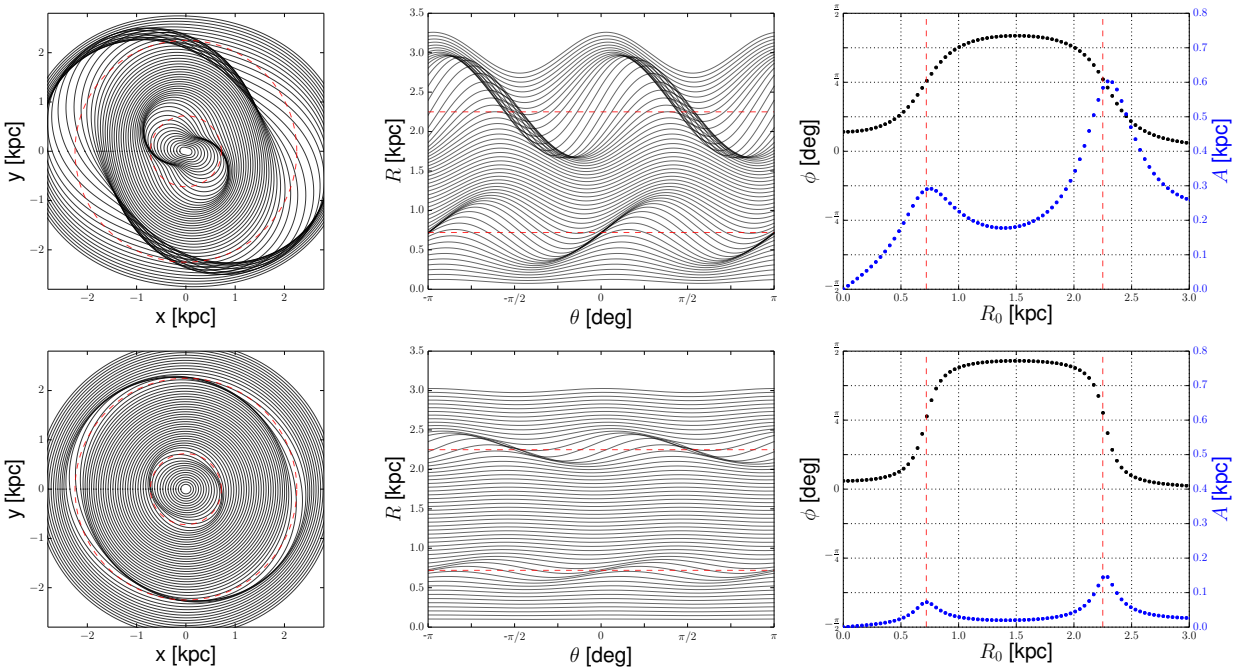


Figure 4.4: Predictions of the theory of Wada (1994) for two different pairs of the parameters ϵ and Λ. The top row is for a bar strength of $\epsilon = 0.05$ and a damping parameter of $\Lambda = 0.05$. The bottom row is for a model with $\epsilon = 0.005$ and $\Lambda = 0.02$. The left panels show a sequence of nested orbits as predicted by Wada (1994) in the xy plane. The central panels show the same orbits represented in the $R\theta$ plane. The right panel shows the values of ϕ (in black) and A (in blue) predicted by Wada (1994) as a function of R_0. Each dot in the right panels corresponds to an orbit.

into account in a simple way the effects of star formation and stellar mass loss, but in practice, the only effect of the recycling law is to prevent too much gas accumulating in the very centre and to replace gas lost at the boundary due to the outflow boundary conditions. It does not affect the morphology of the results, so the results of this chapter do not change if we disable the recycling law.

We used a grid $N \times N$ to simulate a square 16 kpc on a side. N depends on the resolution of the simulation. The initial conditions of the runs are as follows. We start with gas in equilibrium on circular orbits in an axisymmetrized potential and, to avoid transients, turn on the non-axisymmetric part of the potential gradually, i.e., the bar strength ϵ increases linearly with time until it reaches its final value, and is then frozen. We use outflow boundary conditions: gas can freely escape the simulated region, after which it is lost forever. The potential well is sufficiently deep, however, to prevent excessive quantities of material to escape.

4.4 Phenomenological models vs hydro simulations

What happens if we run a simulation in the potential of Eqs. (4.13) and (4.14)? Do the results of a hydro simulation resemble the predictions of the phenomenological models? Fig. 4.5 answers this question. It shows the results of a simulation in which $\epsilon(t)$ is a slowly changing function of time. At $t = 0$ we have $\epsilon = 0$, then ϵ increases linearly with time until it reaches Wada's value of $\epsilon = 0.05$ at $t = 12.3\,\mathrm{Gyr}$, and for later times it is kept constant at this value. The potential evolves so slowly that the gas flow configuration evolves

almost adiabatically, adjusted to the instantaneous underlying potential. In other words, at each time t the gas configuration is almost the same as the steady state configuration that would be obtained by freezing the potential and then waiting for the gas to settle down into a steady state.

At early times, when ϵ is very small, the results of the hydro simulation do qualitatively resemble the predictions of the phenomenological models, as can be seen by comparing the top row in Fig. 4.5 with Fig. 4.4. At later times, when ϵ is greater, this is not true: for example panels in the bottom row do not look like Wada's predictions. Inspection of streamlines shows that for very small values of ϵ the gas is flowing on almost circular orbits, as the epicycle approximation requires, but when ϵ becomes larger the gas flows instead on very horizontally elongated orbits, and the epicycle approximation is not valid anymore. How can we explain this behaviour? Why does the gas flow on very elongated orbits despite the fact that the perturbation is apparently very small? As we shall now see, the key to answer these questions lies in the orbital structure of the underlying potential.

Fig. 4.6 shows how the orbital structure of the underlying potential changes as ϵ is varied. Each row refers to a particular value of ϵ. Let us first consider the top row, which shows the case $\epsilon = 0$, when the bar perturbation is turned off and the potential is exactly axisymmetric. Consequently, orbits that close in our rotating frame are either circular or they are orbits for which the precession frequency happens to coincide with our chosen "pattern speed", which is actually of no physical significance when $\epsilon = 0$. The top left panel of Fig. 4.6 shows several families of closed orbits: each dot represents a closed orbit in terms of its value of the Jacobi constant E_j and the coordinate at which the orbit cuts the vertical axis. All the orbits shown in this diagram (known as characteristic diagram, see Contopoulos & Grosbol, 1989) have the property of cutting the vertical axis at right angles. The line of blue dots shows the circular orbits. Since for $\epsilon = 0$ the potential is axisymmetric, these are the only stable closed orbits.

The red crosses in the top left panel of Fig. 4.6 show eccentric orbits that close in the rotating frame. Such orbits exist only between the two ILRs. They spring out from the sequence of nearly circular orbits at the radius of the innermost inner Lindblad resonance ILR1, and then merge into it again at the bigger radius corresponding to the other inner Lindblad resonance, ILR2. The red crosses that are above the blue line are elongated perpendicularly to the bar, while those below it are elongated along the bar. In fact, since the potential is axisymmetric, the red crosses above the blue line and those vertically below it represent orbits that differ only in a 90° rotation of the major axis. Since the potential is axisymmetric, it is actually possible to find equivalent orbits for any orientation, but these are not shown in the diagram as they do not cut the vertical axis at right angles. The central panel of Fig. 4.6 shows how the orbits appear in the xy plane. These orbits are coloured differently according to their value of the Jacobi constant E_j. In this chapter, E_j always will always have units of $100^2\,\mathrm{km}^2\,\mathrm{s}^{-2}$.

We now consider what happens when a small perturbation is turned on, so the potential is no longer exactly axisymmetric. The left panel of the second row of Fig. 4.6 shows that for $\epsilon = 0.005$, a bifurcation is present approximately at the location of ILR1. The line of blue points that marks the circular orbits in the axisymmetric case is now split. The part inside ILR1 merged with what used to be the orbits elongated

along the bar in the axisymmetric case (which have become stable) to form a continuous line. In fact, this continuous line constitutes the x_1 family in the notation of Contopoulos & Grosbol (1989). The part that used to be circular orbits between the two ILRs now forms a closed loop with the orbits elongated perpendicularly to the bar. In fact, the former circular orbits between the ILRs have become the stable x_2 family, while the orbits elongated perpendicularly to the bar have become the x_3 family and have remained unstable, as opposed to the x_1 orbits that have become stable. Note also that the x_3 orbits are not the x_1 orbits rotated as in the axisymmetric case, because the potential is no longer axisymmetric.

When in the bottom two rows of Fig. 4.6 we further increase ϵ, the x_2-x_3 loop shrinks, and has almost disappeared when $\epsilon = 0.05$. Note that the shape of the orbits, visible in the central column, does change as we increase ϵ, but not dramatically. The x_1 orbits are already very elongated in the axisymmetric case $\epsilon = 0$. What changes significantly is the fraction of the volume of phase space that is occupied by non-closed orbits that librate around x_1 orbits rather than around x_2 orbits. The right column of Fig. 4.6 illustrates this point by showing surfaces of section[6] for a value of Jacobi constant $E_j = -0.5$ (for a definition of surfaces of section see for example Chapter 3 in Binney & Tremaine, 2008). When $\epsilon = 0$, all non closed orbits are parented by the circular orbit, which corresponds to the centre of the "eye" in the top-right panel. When $\epsilon = 0.005$, two eyes are present in the surfaces of section; the centre of the left one corresponds to the x_1 orbit, the centre of the right one to the x_2 orbit (which replaces the circular orbit). Some orbits are now parented by the x_1 orbit, and some others by the x_2. As we increase ϵ, the x_1 orbits becomes predominant in the surface of section, and the fraction of orbits parented by it increases until the x_2 orbit disappears for this energy.

It is now easy to go back to Fig. 4.5 and interpret the results. When ϵ is small, the gas is circulating on x_2 orbits in the outer parts and on weakly elongated x_1 orbits in the inner part (inside the ILR1).[7,8] When ϵ is increased, the volume of phase space occupied by orbits parented by x_2 orbits is reduced. x_2 orbits start disappearing at small radii, and as they disappear the gas has no other choice than to settle onto x_1 orbits. When $\epsilon = 0.05$ the x_2 orbits are gone almost everywhere, and the gas is flowing everywhere on x_1 orbits,

[6] However, one must be careful not to confuse the full phase-space volume occupied by a group of orbits with the area in a surface of section filled by the same orbits, see Binney et al. (1985).

[7] Note that there must be a transition region in between, and spiral arms are generated as the orientation of orbits changes to make this transition. When the bar perturbation is stronger and orbits are more elongated, as we have seen in Chapter 3, the transition can be mediated by shocks rather than by a soft spiral arm overdensity as in this case. Note also that in the present case the transition is $x_2 \rightarrow x_1$ outwards-in, while the transition discussed in Chapter 3 is $x_1 \rightarrow x_2$. This is to be expected given the orbital structure of the potential: whenever an orbit family disappears and another appears the gas, which follows the available orbit families, must transit from one family to the other. The orbital structure of the potential in Chapter 3 was such that x_2 orbits appear at smaller radii than x_1 orbits, and thus the transition is in the sense $x_1 \rightarrow x_2$. The orbital structure of the potential considered here is such that x_1 orbits appear inside x_2 orbits, and thus the transition is the other way around.

[8] Note also that in the top panels in Fig. 4.5, as well as in Fig. 4.6, there are two sets of spiral arms, one leading and the other trailing. In general, whenever a family of orbits appears and another disappears, which usually happens close to resonances, we expect spiral arms to appear. The gas follows the available orbits families and has to make a transition, either through a spiral arm or a shock. Indeed, the x_2 family approximately extends from ILR1 to ILR2. Thus in the potential considered here gas must transit to the x_2 family when coming from outside the ILR2, generating a set of arm, and must transit to another family when going inside the ILR1, generating the other set of arms. See also Wada (1994). In cases with no ILR, the only family present is the x_1 family and there is no x_2 family. In these cases, there are no spiral arms due to the transition of gas from x_1 to x_2 orbits, as the latter are not present. Gas simply flows on x_1 orbits until it reaches the centre. Spiral arms can be generated by librations, and not by a transition between orbits, in the same way as for the model in Section 4.6.

including the region where the x_1 orbits are highly elongated. In this regime, the epicycle approximation obviously cannot work as the x_1 orbits are not weakly deformed circular orbits. If we call weak a bar that can be well described under the epicycle approximation, then the $\epsilon = 0.05$ case should be classified as a strong bar, despite the fact that the non-axisymmetric part of the potential might naively appear small compared to the axisymmetric part.

We have now two questions.

1. How do the phenomenological theories reviewed above compare with the hydro simulations in the weak bar case, that is, when ϵ is extremely small and the epicycle approximation is valid?

2. What can we say in the strong bar case, when the gas is flowing on very elongated orbits and the epicycle approximation is not valid?

Sections 4.5 and 4.6 investigate these two questions respectively.

4.5 Weak bar case

In this section we want to investigate how the phenomenological models reviewed above compare in detail with the results of a hydro simulation in the weak bar case, when the epicycle approximation is valid. To do this, we study the case $\epsilon = 0.005$. Note that this is a value ten times smaller than the value $\epsilon = 0.05$ considered by Wada (1994) and Piñol-Ferrer et al. (2012). As we have seen in the last section, that should actually be considered a strong bar case, because of its orbital structure.

When the sound speed is not too high, the morphology of the density distribution qualitatively resembles the predictions of Wada (1994) shown in Fig. 4.4. This can be seen from Fig. 4.7, which shows results of hydro simulations with $\epsilon = 0.005$ and different values of the sound speed c_s. In these simulations the bar strength is frozen after the value $\epsilon = 0.005$ is reached, and the snapshots considered are all at $t = 2.9$Gyr, long after the bar is fully grown to this value (which happens at $t = 1.2$Gyr) and the gas has long settled down into a steady-state configuration. The left column shows the density distribution. The density distributions for $c_\mathrm{s} = 0.625\,\mathrm{km\,s^{-1}}$ and $c_\mathrm{s} = 1.25\,\mathrm{km\,s^{-1}}$ are very similar: it appears as if the gas configuration is tending towards some limiting configuration as $c_\mathrm{s} \to 0$ at fixed spatial resolution, and that this limiting configuration has much in common[9] with the predictions of the phenomenological models (compare Fig. 4.7 with the bottom row in Fig. 4.4).

A second important prediction of the phenomenological model that is confirmed by the hydro simulations is that the spiral arms can be understood as kinematic density waves. Fig. 4.8 investigates the instantaneous streamlines, that is, the streamlines calculated from a frozen snapshot of the velocity field. Since the gas is in a steady state, these are not much different from the real streamlines followed by gas parcels during the

[9]Note however that the contrast is not accurately reproduced: as discussed below the reason is that in the phenomenological models the prediction for the amplitude of the oscillations A is not accurate, and is weaker than in the hydro, producing weaker density contrasts.

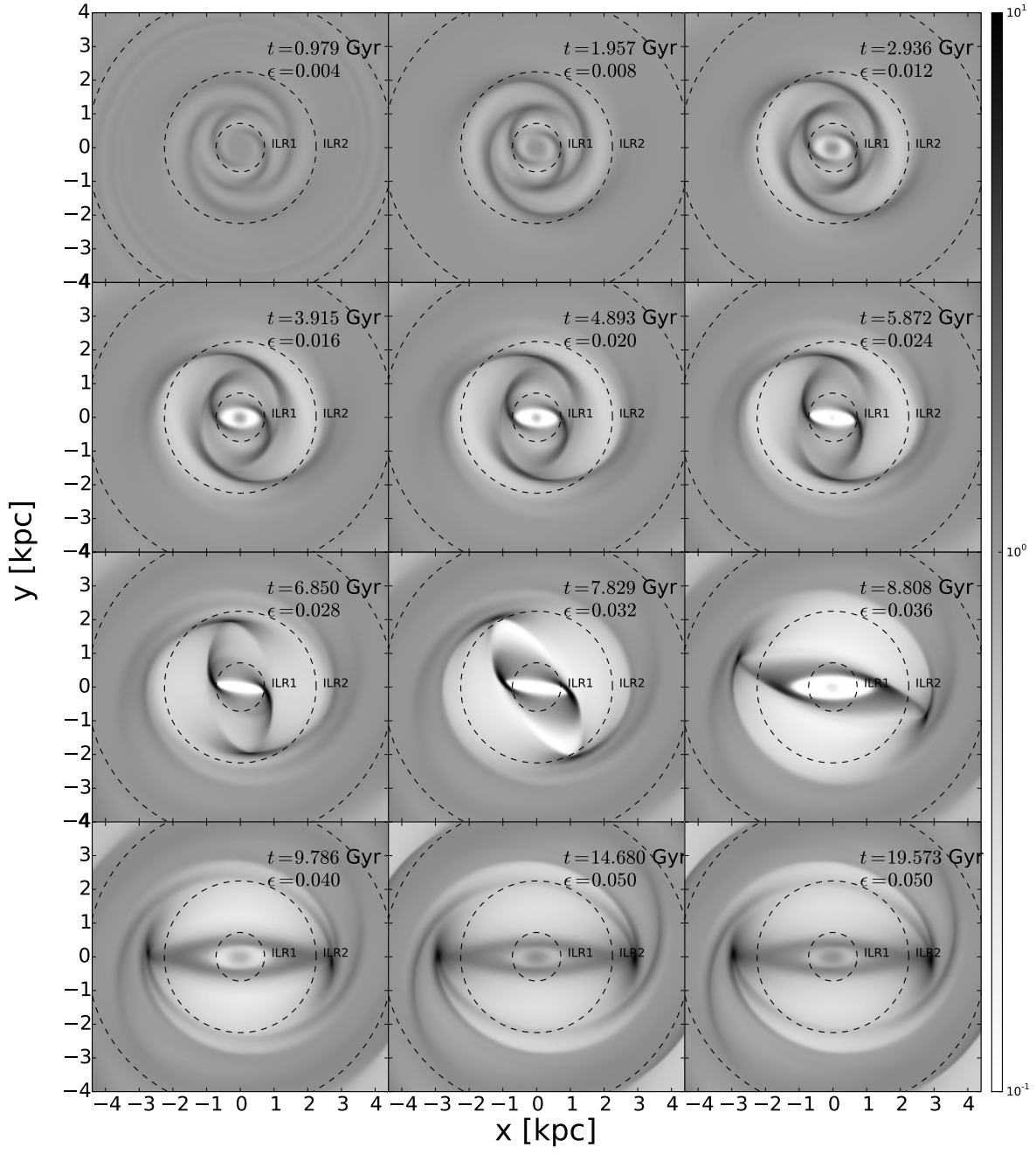


Figure 4.5: Time sequence for a simulation in the potential defined by Eqs. (4.13) and (4.14). Each panel shows the density map at a particular time. The bar strength ϵ increases slowly and linearly with time until it reaches the value of $\epsilon = 0.05$, and is then frozen. The gas adjusts almost adiabatically to the instantaneous underlying potential. The sound speed is $c_{\rm s} = 2.5\,{\rm km\,s^{-1}}$ and the spatial resolution is ${\rm d}x = 20\,{\rm pc}$. The dotted circles indicate the two ILRs and the CR. In each panel, the evolutionary time t and the instantaneous value of ϵ are shown. The colorbar is in units of $M_\odot\,{\rm pc}^{-2}$.

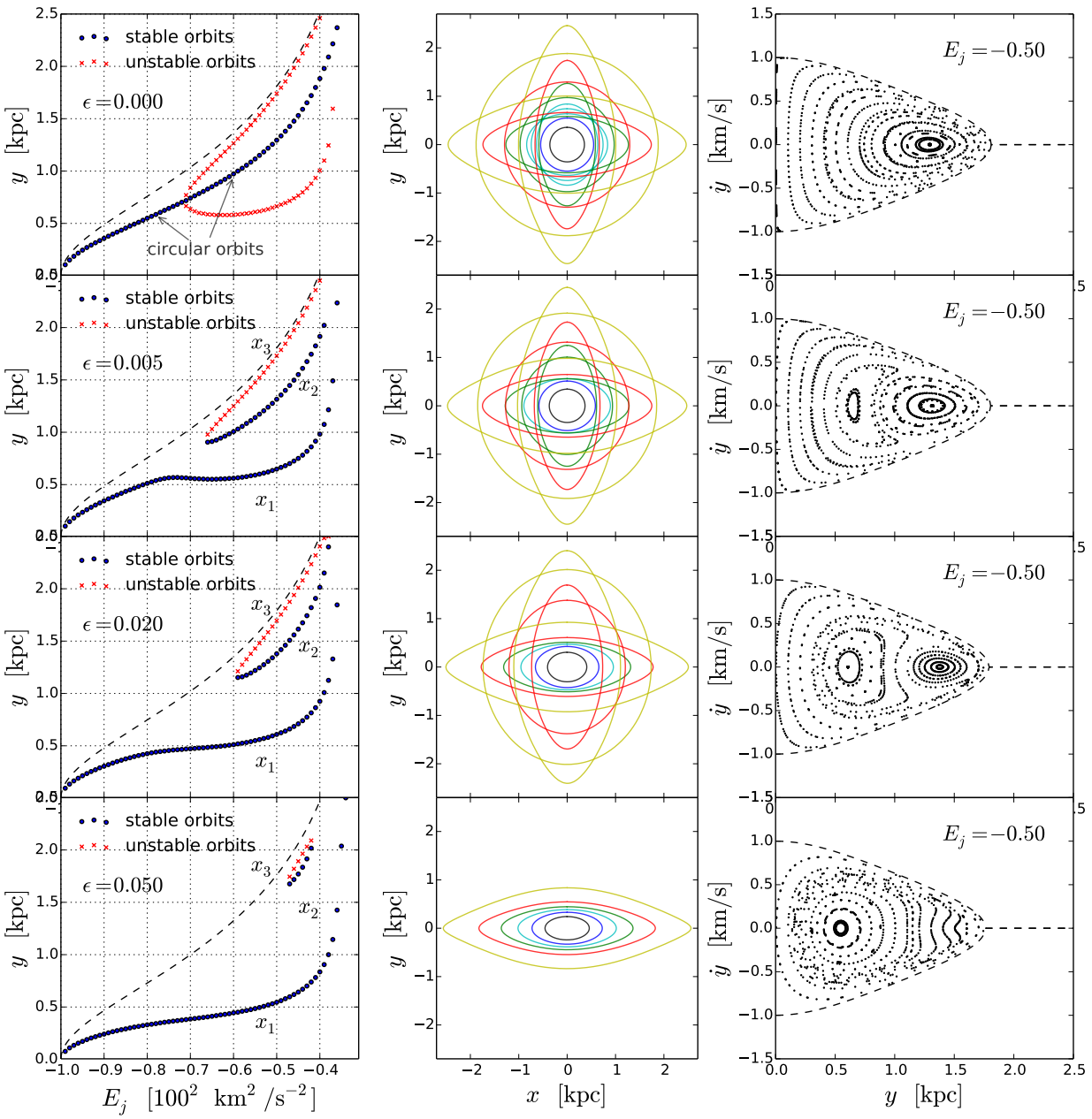


Figure 4.6: Orbital structure of the potential given by Eq. (4.13) for different values of ϵ. Rows from top to bottom show the cases $\epsilon = 0.0, 0.005, 0.02$ and 0.05, respectively. In the left column, the Jacobi constant E_j of a closed orbit is plotted against the value at which the orbit cuts the vertical axis in potentials with different values of ϵ. Each dot represents a closed orbit, and all closed orbits cut the vertical axis with purely horizontal velocity. The middle column shows the closed orbits in xy space, with orbits of the same colour having the same energy. The orbits shown are those corresponding to values of E_j equal to $\{-0.9, -0.8, -0.7, -0.6, -0.5, -0.4\}$. The vertical slices of the grid in the left panels are drawn at these energies. The correspondence between energy and colours can be identified from the bottom panel in the central column: here greater energy (i.e., less negative) corresponds to bigger orbits, and the same colour coding applies to other panels. The right column shows surfaces of section for $E_j = -0.5$.

time evolution. We see that in all cases streamlines are almost closed curves, in the sense that in general the mismatch after a single loop is much less than the extent of a radial oscillation during the loop.

Another verified prediction is that streamlines are well described by closed orbits of the form Eq. (4.18). The middle column in Fig. 4.8 shows in full lines the same streamlines as the left column in the $R\theta$ plane, and in dashed lines a simple fit to each streamline using Eq. (4.18) where A and ϕ are considered independent free parameters. The streamlines are well fitted by a function of this form, and if we were to reproduce a nested sequence of streamlines as in Fig. 4.4 using the best fitting values of these two parameters, we would obtain kinematic density waves that reproduce those obtained in the simulation.

Let us know discuss the limits of the phenomenological models. In the phenomenological models, A and ϕ both depend on a single parameter, the dissipation λ, while in the fitting procedure above A and ϕ are treated as independent parameters. If the phenomenological models were correctly predicting the outcome of the simulations, it would be possible by appropriately choosing the value of λ for each orbit to reproduce both A and ϕ obtained from the simulation simultaneously. We can see that this is not the case from the right column of Fig. 4.8. The phenomenological models can reproduce at best one among A and ϕ by carefully tuning λ. This is a weakness of the phenomenological models. Moreover, in the phenomenological models some values of A and ϕ are forbidden: ϕ must stay within the blue and yellow line in Fig. 4.2, and the amplitude A away from resonance has an upper bound. Therefore in some cases the phenomenological models cannot even reproduce either A or ϕ, as these lie outside the regions obtainable with the phenomenological models for any value of λ. This is clear from the right column of Fig. 4.8: in the hydro simulation A and ϕ are often outside the regions allowed by the phenomenological models for the same potential. In fact, this is the reason why we chose to fit the streamlines using A and ϕ as two independent parameters instead of using directly the dissipation λ as the only free parameter: A and ϕ lie outside the allowed regions too often. We conclude that not all hydro simulations can be reproduced by an appropriate gauging of the dissipation, even in the weak bar case. However, in the limit of very low sound speed, the results of the hydro simulations for A and ϕ are not far from the predictions of the phenomenological model in Fig. 4.4. The values of ϕ in the top two rows of Fig. 4.8 are more similar than those in other rows to the black dots in the bottom right panel of Fig. 4.4, although some significant differences remain, and the values of A have also much in common with the blue dots in the same panel, although many are still forbidden because they lie just above the curve $A = 0$ in the right panel of Fig. 4.3. Another interesting prediction that is verified in this limit is that the orientation of the major axis of the closed orbits at the two ILRs is $\pi/4$.

Why are the results of the phenomenological models well reproduced in the limit of vanishing sound speed, and less well when the sound speed is higher? To understand this, consider the equations of motion of a gaseous parcel flowing in the hydro simulation. If the fluid were really an ideal isothermal fluid, these would be:

$$\frac{D\mathbf{v}}{Dt} = -\nabla\Phi - c_{\mathrm{s}}^2 \frac{\nabla\rho}{\rho}, \tag{4.23}$$

where D indicates the material derivative, the first term on the right hand side is the gravitational force and the second is the pressure force. However, in a real hydrodynamical simulation at fixed spatial resolution, some unavoidable amount of numerical viscosity, which is not included in Eq. (4.23), will always be present. This numerical viscosity decreases as we increase the resolution, and tends to zero as the resolution goes to infinity. The equations of motion of a parcel of gas are therefore something like

$$\frac{D\mathbf{v}}{Dt} = -\nabla\Phi - c_{\rm s}^2\frac{\nabla\rho}{\rho} - F_{{\rm d}x}, \tag{4.24}$$

where $F_{{\rm d}x}$ indicates the force arising from the finite resolution, which we will loosely refer to as the viscous force. Hence, a parcel in a hydro simulation is subject to three different forces: gravitational, pressure and viscosity.

The phenomenological models completely neglect pressure forces. Instead, they only account for the gravitational forces and, phenomenologically, for the viscous forces $F_{{\rm d}x}$. Note that the pressure force can be derived from a "pressure potential"

$$\Phi_P = c_{\rm s}^2\log(\rho/\rho_0), \tag{4.25}$$

where ρ_0 is an arbitrary number. In a steady-state configuration, the density does not change with time, and the pressure forces can be derived from a static potential. Hence, a gaseous parcel can be considered to move in a potential that is the sum of the gravitational plus the pressure potential. This static pressure potential is not easily derived a priori, and we can only calculate it once we are given the steady-state configuration of the gas after solving the hydrodynamical equations.

The central and right columns in Fig. 4.7 show the pressure and perturbation potential forces along the horizontal axis. The central column shows forces in the y direction; the gravitational forces in this direction are zero for the chosen form of the perturbation potential. The right column shows forces in the horizontal direction. We see that at high sound speed, when the phenomenological models are less accurate, the magnitude of the pressure forces is in general greater than the magnitude of the perturbation potential forces. At $c_{\rm s} = 1.25$ the pressure contribution becomes smaller than the bar perturbation contribution, and at $c_{\rm s} = 0.625$ the pressure forces are negligible compared to the bar perturbation contribution. This is why the phenomenological models above work in this limit. The phenomenological models have terms that take into account both the effects of f and the gravitational potential, but they do not take into account the effects of pressure. In the limit where pressure is negligible, they work. If in the phenomenological models we could replace the gravitational potential with the sum of the gravitational plus the pressure one, then they would describe the results of the simulations at finite pressure. The problem is that in general the pressure potential is not easily obtained a priori.

Before moving on to discuss the strong bar case, we have a couple of remarks left. From a theoretical point of view, when the sound speed tends to zero the gas is always supersonic and we expect all small

perturbations to cause shocks. Hence, we expect that when c_s tends to zero, we find shocks anywhere there is a small velocity gradient. At a shock, the gradient of the density diverges but the forces due to pressure could remain finite as they depend on the product of the sound speed (which goes to zero) with the gradient of the pressure (which goes to infinity). The reason why shocks are not formed in our simulations when the sound speed tends to zero is that at finite resolution the numerical viscosity[10] prevents them from forming. In the limit of vanishing sound speed, the numerical viscosity becomes more important than the pressure force. Indeed, we have checked that increasing the numerical resolution leads to steady states in which there are sharper density variations – see Fig. 4.9. Note that the pressure forces in Fig. 4.9 are stronger than the pressure forces for a simulation in the same potential, with the same value of the sound speed but at lower resolution – top row of Fig. 4.7. Hence, we conclude that the phenomenological models work in reproducing the result of a simulation in the limit of low sound speed and of finite numerical viscosity, which means finite numerical resolution.

4.6 Strong bar case

In the previous section, we have seen that in the weak bar case the spiral arms can be understood as kinematic density waves, generated by weakly oval closed orbits. What can we say about the strong bar case? In order to investigate a realistic example of a strong bar we consider a simulation we used in Chapter 3, performed in the Binney et al. (1991) barred potential. The simulation used here is the one with sound speed $c_s = 10\,\mathrm{km\,s^{-1}}$ and spatial resolution $\mathrm{d}x = 10\,\mathrm{pc}$.

In Chap. 3, we noted the presence of spiral arms, but we postponed their detailed discussion to this chapter. We have shown in Chap, 3 that in the region where spiral arms are present the velocity field was very well approximated by x_1 orbits. To discuss the spiral arms, we will need to look at the tiny deviations of the hydro velocity field from the ballistic x_1 orbits field.

The left panel in Fig. 4.10 shows the same density snapshot as in Fig 3.2 ($c_s = 10\,\mathrm{km\,s^{-1}}$ and $\mathrm{d}x = 10\,\mathrm{km\,s^{-1}}$), taken after the bar is fully grown and the gas has reached an approximate steady state. Spiral arms emerging from the end of the bar are clearly identifiable. These spiral arms are stationary in the frame corotating with the bar. The central panel of Fig. 4.10 shows what happens when we nest together many instantaneous streamlines. We see that, to a very good approximation, streamlines are closed curves that when nested together generate the spiral arms. In the right panel ballistic closed orbits are shown for comparison. Fig. 4.11 analyses in more detail a selection of instantaneous streamlines in this snapshot. In the top panel, five streamlines are shown in dashed lines. These are very nearly closed. In the same panel, the ballistic closed orbits that cut the y axis at the same value of the streamlines are shown in full lines. It can be seen that streamlines are librations around underlying closed x_1 orbits. In the middle panel, the

[10]The numerical viscosity is an unavoidable numerical issue associated with hydro simulations. The property of the real interstellar medium that corresponds to it is a finite real viscosity of the interstellar medium, due most likely to the exchange (by diffusion) of particles between areas of different average momentum. See also Klessen & Glover (2014) and references therein.

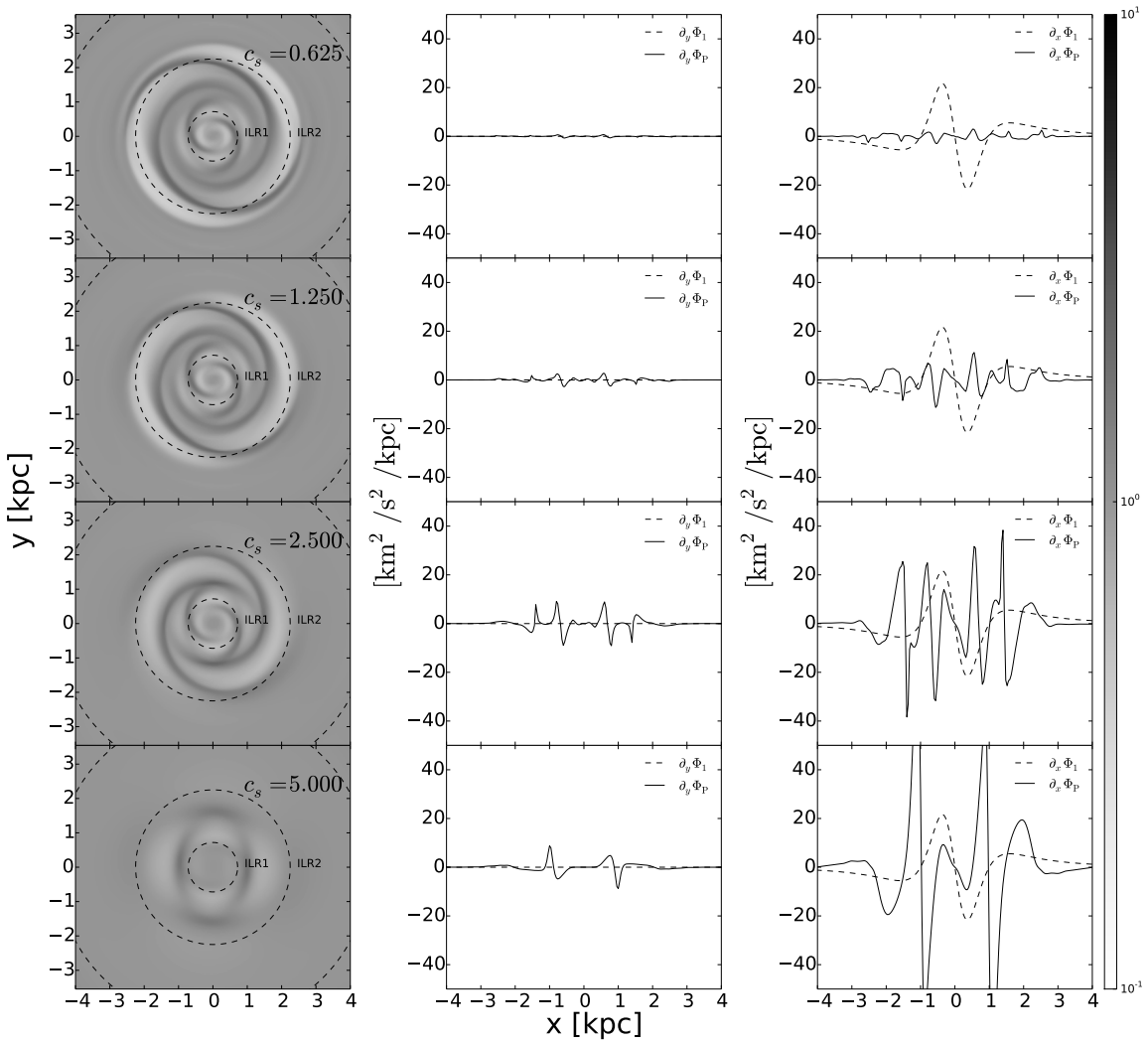


Figure 4.7: Results of a simulation in Wada's potential in the case of a weak bar ($\epsilon = 0.005$) for different values of the sound speed c_s. Each panel in the left column shows the steady-state density long after the bar has reached the final value of $\epsilon = 0.005$ – compare with Fig. 4.4. The middle column shows the y components of pressure forces (full lines) and bar perturbation gravitational forces (dashed lines) from the potential in Eq. (4.13) with $\epsilon = 0.005$ calculated along the horizontal axis. The right column shows the x components of the same forces, again calculated along the horizontal axis. Note that pressure forces can be derived from a pressure potential Φ_P as described in the text. The colorbar is in units of $M_\odot\,\mathrm{pc}^{-2}$.

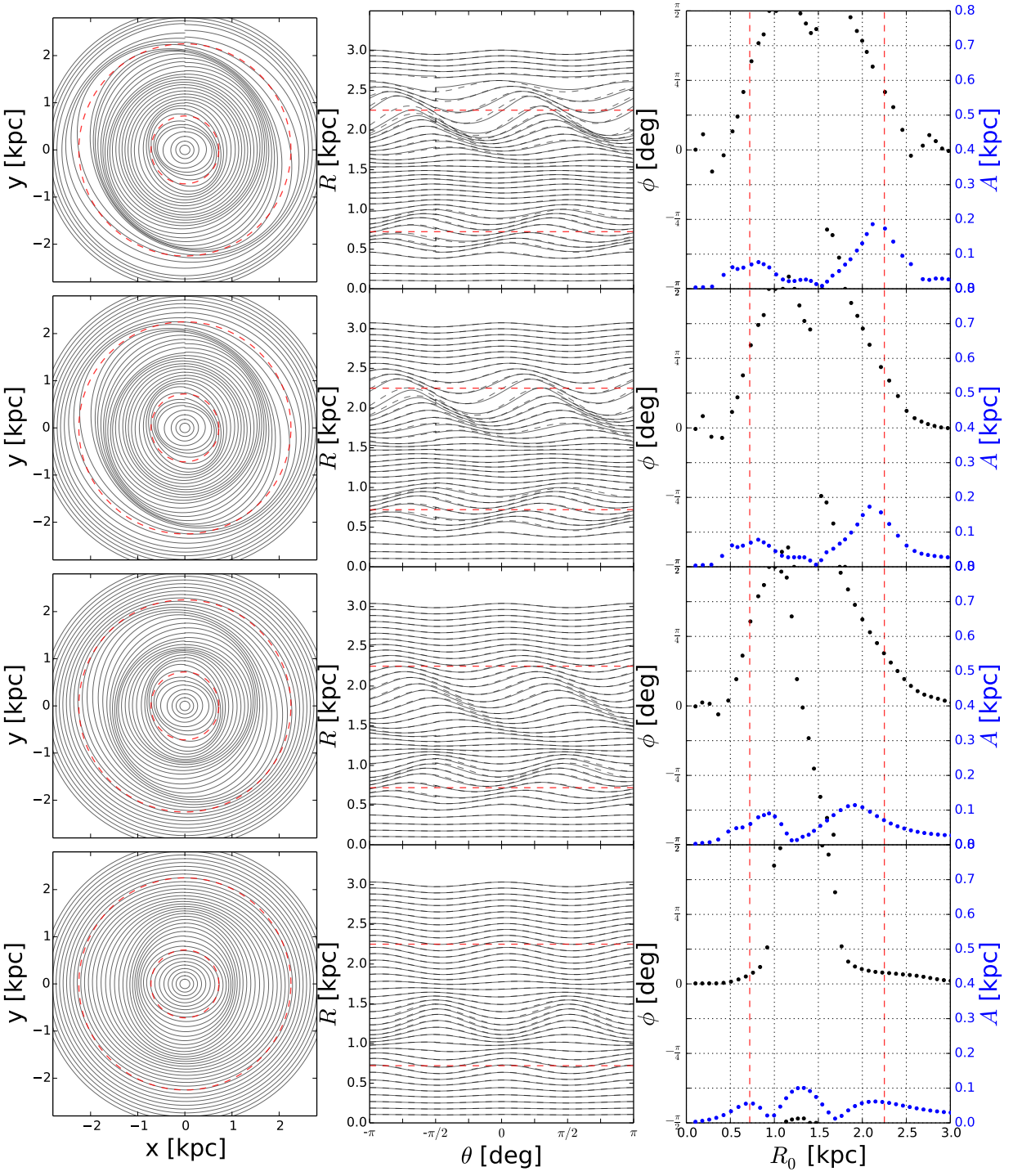


Figure 4.8: Analysis of the instantaneous streamlines for the snapshots shown in Fig. 4.7. The left column shows a sequence of nested streamlines in the xy plane. The middle column shows in full lines the same streamlines in the $R\theta$ plane, and in dashed lines the result of a simple fit using Eq. (4.18), where A and ϕ are independent parameters. The right panel shows the best fit values of A and ϕ as a function of radius. In each panel the dashed red lines mark the two inner Lindblad resonances.

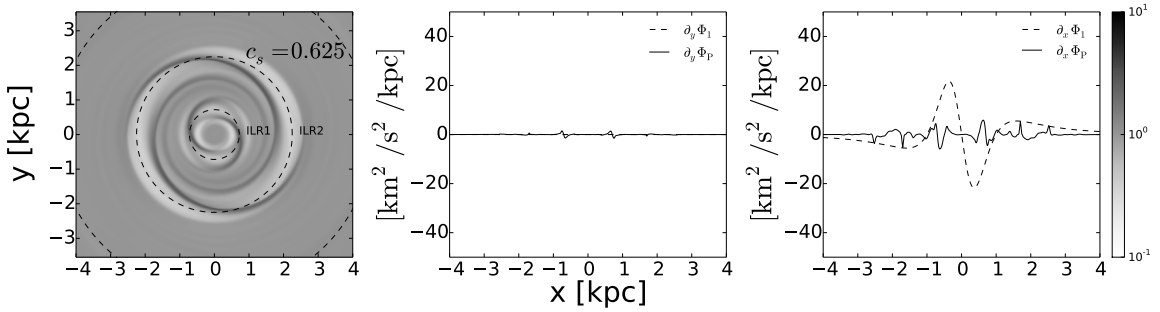


Figure 4.9: Same as the top row of Fig. 4.7, but for a simulation with twice the spatial resolution. Here, $\mathrm{d}x = 10\,\mathrm{pc}$. The colorbar is in units of $M_\odot\,\mathrm{pc}^{-2}$.

same streamlines and orbits are shown in the $R\theta$ plane. In the bottom panel, the radial difference between the two (orbits and streamlines) is shown in the $R\theta$ plane. The librations do not have the simple $\cos(2\theta)$ structure that we encountered in Sect. 4.5.

The above considerations strongly suggest that also in the strong bar case the spiral arms can be understood as kinematic density waves, generated by small librations around underlying closed orbits. In the epicycle approximation, the librations are around circular orbits. Here, the gas parcels librate around x_1 orbits. It is natural to ask whether the phenomenological models can be extended to the strong bar case, where now the perturbations should be made around the x_1 orbits and not around circular orbits. In order to investigate this, let us go back to the equations of motion in the rotating frame. For a gaseous parcel in a hydro simulation, Eq. (4.2) is modified to

$$\ddot{\mathbf{x}} = -\nabla(\Phi + \Phi_P) + \Omega_\mathrm{p}^2 \mathbf{x} - 2\Omega_\mathrm{p}\,(\hat{\mathbf{e}}_z \times \dot{\mathbf{x}}) + F_{\mathrm{d}x}\ , \tag{4.26}$$

where Φ_P is the "potential" that gives the pressure force introduced above and $F_{\mathrm{d}x}$ is the viscous force. We can do a Floquet analysis to investigate librations around ballistic closed orbits (see for example Binney & Tremaine, 1987). Let us write $\Phi = \Phi_0 + \Phi_1$, where this time Φ_0 is not necessarily axisymmetric. Let $\mathbf{x}_\mathrm{c}(t)$ be a ballistic closed orbit in the potential Φ_0. We can write a libration around this closed orbit as $\mathbf{x}(t) = \mathbf{x}_\mathrm{c}(t) + \mathbf{x}_1(t)$. By substituting this last equation into (4.26), expanding to first order in quantities with subscript 1 and cancelling the zeroth order terms, we obtain that the equation of motion for the libration is

$$\begin{aligned}\ddot{\mathbf{x}}_1 = &-\left[(\mathbf{x}_1 \cdot \nabla)\nabla\Phi_0 + \nabla\Phi_1 + \nabla\Phi_P\right]_{x_\mathrm{c}(t)} + \Omega_\mathrm{p}^2 \mathbf{x}_1 \\ &- 2\Omega_\mathrm{p}\,(\hat{\mathbf{e}}_z \times \dot{\mathbf{x}}_1) + F_{\mathrm{d}x}\ .\end{aligned} \tag{4.27}$$

In this equation, all the derivatives of the potentials are to be calculated along the unperturbed closed orbit and are therefore known functions of time. Equations (4.9) for the epicycle approximation can be considered as a particular case of Eq. (4.27), obtained assuming that Φ_0 is axisymmetric, $\Phi_P = 0$ and there are no viscosity forces (which are then added heuristically in the phenomenological models). In trying to generalise the analysis of Sect. 4.2 to the case when Φ is a strongly barred potential, we are faced with the difficulty that there is in general no natural way of splitting Φ into a Φ_0 where we can easily calculate closed orbits

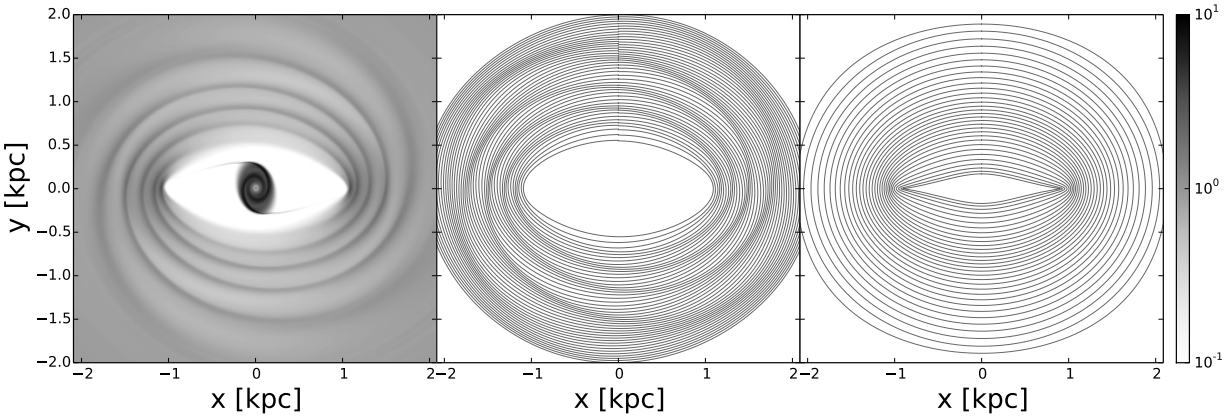


Figure 4.10: In the left panel, the same density distribution shown in Fig. 3.2 for the simulation with $c_{\rm s} = 10\,{\rm km\,s^{-1}}$ and spatial resolution ${\rm d}x = 10\,{\rm pc}$. This snapshot is taken after the bar is fully grown and the final steady state is reached. The middle panel shows a nested sequence of instantaneous streamlines. The right panel shows ballistic closed x_1 orbits in the same underlying potential. The colorbar is in units of $M_\odot\,{\rm pc}^{-2}$.

and a perturbation Φ_1 that only weakly deforms closed orbits. The only natural choice is to set $\Phi_1 = 0$. If this choice is made, the whole perturbing potential is given by Φ_P, which is not known a priori. In the phenomenological models in the epicycle approximation we were able to neglect Φ_P compared to Φ_1 in the low pressure case, but here the same cannot be done if $\Phi_1 = 0$.

Suppose that Φ_P is somehow known. If we assume $\Phi_1 = 0$ and define a suitable phenomenological viscous term $F_{{\rm d}x}$ in Eq. (4.27), would the resulting equation correctly describe librations that give rise to spiral arms? To investigate this, we can do as follows. We first extract the density from the snapshot shown in the left panel of Fig. 4.10. From this density, we derive the pressure potential using Eq. (4.25) and $c_{\rm s} = 10\,{\rm km\,s^{-1}}$. Then we pretend we do not know how Φ_P was obtained and find librations using the following equation:

$$\boxed{\ddot{\mathbf{x}}_1 = -\left[(\mathbf{x}_1\cdot\nabla)\nabla\Phi_0 + \nabla\Phi_P\right]_{x_c(t)} + \Omega_{\rm p}^2\mathbf{x}_1 - 2\Omega_{\rm p}\left(\hat{\mathbf{e}}_z\times\dot{\mathbf{x}}_1\right) - 2\lambda\dot{\mathbf{x}}_1}\,. \tag{4.28}$$

In this last equation, we have introduced a phenomenological dissipation term analogous to the models in the epicycle approximation. The dissipation is proportional to the difference between the total velocity and the local x_1 velocity field. Φ_0, Φ_P and $\Omega_{\rm p}$ are now known, and given a value of λ we can solve Eq. (4.28) to find the librations. Among all possible solutions to this equation, we want the solutions $\mathbf{x}_1(t)$ that are periodic with the same period of the underlying closed orbits. This amounts to finding the solution where transients are gone, dissipated away. The method that we used to solve the differential equation (4.28) is described in detail in Appendix A.

Fig. 4.12 shows a nested sequence of librations around x_1 orbits calculated using Eq. (4.28) and a value of $\lambda = 50\,{\rm km\,s^{-1}\,kpc^{-1}}$. Spiral arms identical to those present in the hydro simulation of Fig. (4.10) are recovered. This confirms that the spiral arms can be understood as kinematic density waves also in the strong bar case. Eq. (4.28) well describes the librations if the correct Φ_P is provided.

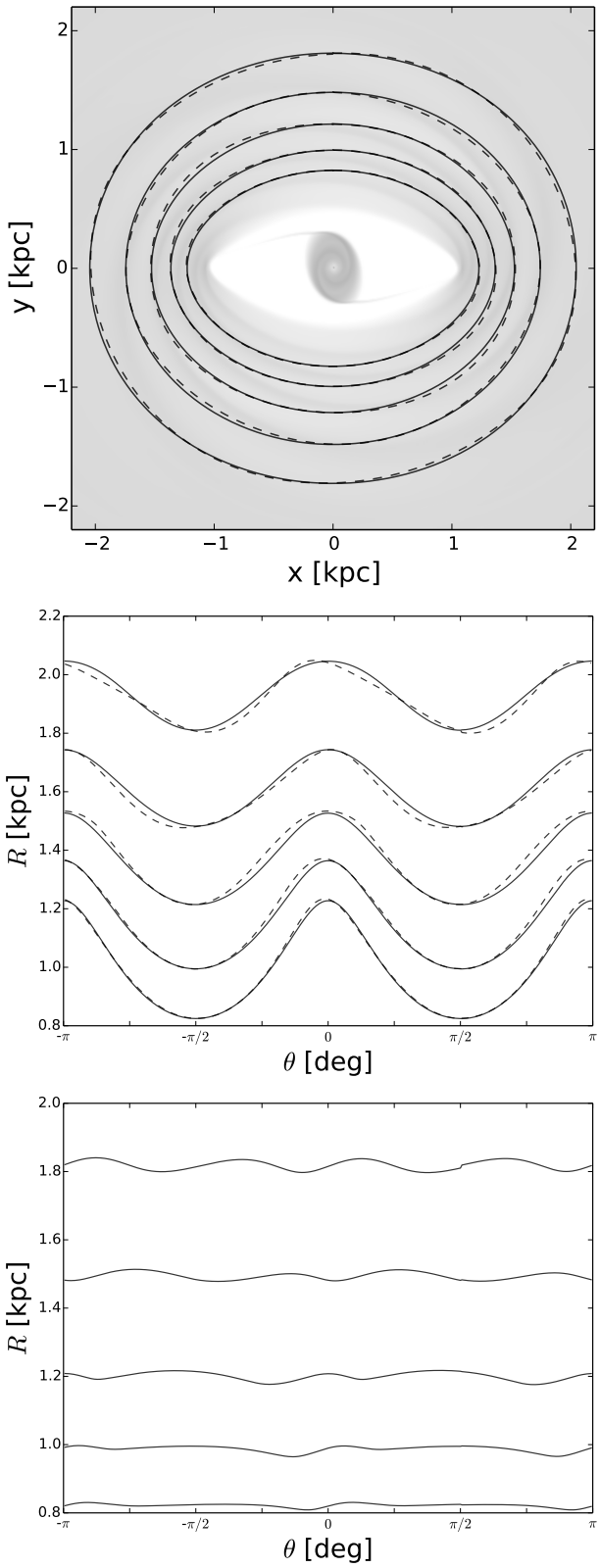


Figure 4.11: A Closer look at some of the streamlines of Fig. 4.10. In the top panel, streamlines (dashed lines) and ballistic x_1 orbits that cut the y axis at the same value (full lines). In the background is visible the density distribution of Fig. 4.10 for ease of comparison. The middle panel shows the same streamlines and orbits in the $R\theta$ plane. The bottom panel shows the difference $R_{\rm streamline}(\theta) - R_{x_1}(\theta)$, where the zero is shifted for clarity.

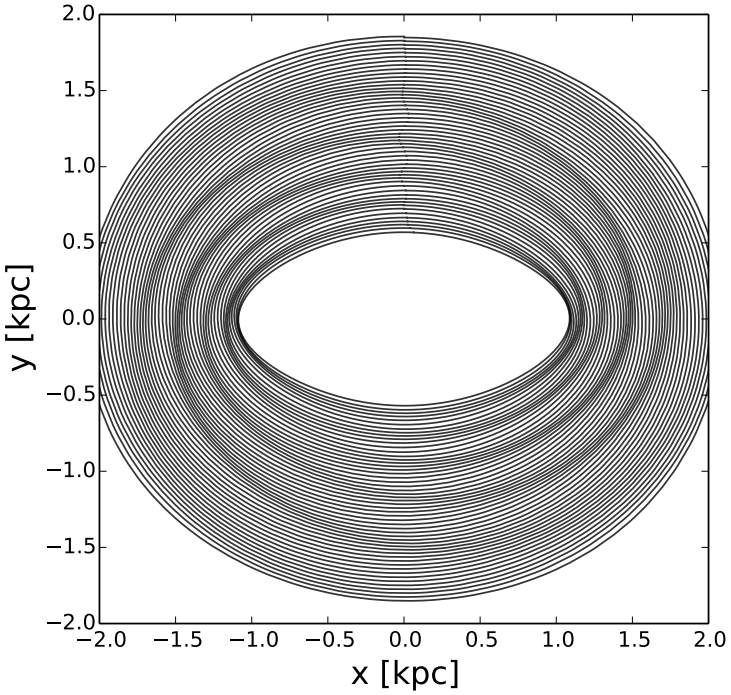


Figure 4.12: Nested librations around closed x_1 orbits generated using Eq. (4.28). The potential Φ_0 is the potential used in Chap. 3, the pressure potential Φ_P is derived from the density map in the left panel of Fig. 4.10, and $\lambda = 50\,\mathrm{km\,s^{-1}\,kpc^{-1}}$.

However, in the above analysis we have cheated in the sense that we have obtained Φ_P from the result of a full hydrodynamical calculation. Is there a simple way of deriving a suitable Φ_P, without solving the full hydrodynaical problem? We haven't found a better way. We considered the possibility of obtaining Φ_P from x_1 closed orbits. Since the x_1 orbits are highly elongated, we have tried to infer a density map from orbit crowding, and from this the pressure potential. However, the pressure forces obtained in this way were much weaker than the ones obtained in the hydro simulation for a value of the sound speed of $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$, and were not capable of reproducing spiral arms even for a small value of the dissipation. For a higher value of the sound speed we obtained segments of spiral arms that were not coherent and did not produce any "grand design". Hence, we conclude that while the spiral arms in the strong bar case can still be explained by kinematic density waves generated by librations around the appropriate closed orbits, calculation of these librations require knowledge of the pressure forces and hence solution of the full hydrodynamical problem. Indeed, the morphology of the spiral arms found in Chapter 3 depends significantly on the value of pressure, and a phenomenological model that wants to reproduce them must take pressure explicitly into account.

4.7 Conclusion

In this chapter, we have investigated bar-driven spiral arms in the absence of self-gravity. Our focus was on understanding the physical mechanism involved. We concluded that, both in the weak and strong bar cases, the spiral arms can be understood as kinematic density waves generated by librations around underlying ballistic closed orbits. In the weak bar case, the librations can be considered deviations from circular orbits.

In the strong bar case, the librations are to be considered deviations from the appropriate underlying closed orbits, which can be highly elongated. In the strong bar case the epicycle approximation is not valid. In fact, whether a bar can be considered weak or strong is determined by the validity of the epicycle approximation. Indeed in Sect. 4.4 we argued that a bar is weak or strong if the orbital structure is or is not obtainable from the epicycle approximation. Bars that might naively be considered weak, for example the case $\epsilon = 0.05$, should be considered strong.

A parcel in a hydro simulation is subject to three different forces: gravitational, pressure and viscosity. We have tested the phenomenological models available in the literature aimed at explaining the spiral arms in the weak bar case, when the epicycle approximation is valid. We found that the key ingredient not taken into account by these models is pressure. Therefore they work well in regimes where pressure can be neglected, such as simulations in a weak bar at finite resolution and vanishing sound speed. We have also discussed how the phenomenological models should be extended to the strong bar case. When the pressure forces are known, these extensions work very well in explaining the spiral arms in the strong bar case. Unfortunately, the pressure forces are in general known only after solving the full hydrodynamical problem and are not known a priori. Thus, while the phenomenological models provide insight into the physical mechanism that generates the spiral arms, they appear to be of little practical use.

References

Athanassoula E., 1992, MNRAS, 259, 345
Athanassoula E., Romero-Gómez M., Bosma A., Masdemont J. J., 2009a, MNRAS, 400, 1706
Athanassoula E., Romero-Gómez M., Bosma A., Masdemont J. J., 2010, MNRAS, 407, 1433
Athanassoula E., Romero-Gómez M., Masdemont J. J., 2009b, MNRAS, 394, 67
Binney J., Gerhard O. E., Hut P., 1985, MNRAS, 215, 59
Binney J., Gerhard O. E., Stark A. A., Bally J., Uchida K. I., 1991, MNRAS, 252, 210
Binney J., Tremaine S., 1987, Galactic dynamics
Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition. Princeton University Press
Bissantz N., Englmaier P., Gerhard O., 2003, MNRAS, 340, 949
Contopoulos G., Grosbol P., 1989, Astronomy and Astrophysics Reviews, 1, 261
Dobbs C., Baba J., 2014, PASA, 31, 35
Englmaier P., Gerhard O., 1999, MNRAS, 304, 512
Feldman S. I., Lin C. C., 1973, Studies in Applied Mathematics, 52, 1
Huntley J. M., Sanders R. H., Roberts, Jr. W. W., 1978, ApJ, 221, 521
Klessen R. S., Glover S. C. O., 2014, ArXiv e-prints 1412.5182
Lin C. C., Lau Y. Y., 1975, SIAM Journal of Applied Mathematics, 29, 352
Lin C. C., Shu F. H., 1964, ApJ, 140, 646
Lindblad P. O., Lindblad P. A. B., 1994, in Astronomical Society of the Pacific Conference Series, Vol. 66, Physics of the Gaseous and Stellar Disks of the Galaxy, King I. R., ed., p. 29
Piñol-Ferrer N., Lindblad P. O., Fathi K., 2012, MNRAS, 421, 1089
Rodriguez-Fernandez N. J., Combes F., 2008, A & A, 489, 115
Romero-Gómez M., Athanassoula E., Masdemont J. J., Garc´ia-Gómez C., 2007, A & A, 472, 63
Romero-Gómez M., Masdemont J. J., Athanassoula E., Garc´ia-Gómez C., 2006, A & A, 453, 39
Sanders R. H., 1977, ApJ, 217, 916
Sanders R. H., Huntley J. M., 1976, ApJ, 209, 53
van Albada G. D., van Leer B., Roberts, Jr. W. W., 1982, A & A, 108, 76
Wada K., 1994, PASJ, 46, 165
Weiner B. J., Sellwood J. A., 1999, ApJ, 524, 112

CHAPTER 5

Effects of Varying the Quadrupole

5.1 Introduction

In the two previous chapters we investigated the dynamics of gas flow in barred potentials. We studied in detail the relation between the flow of a 2D isothermal non self-gravitating inviscid gas and closed ballistic orbits in the same externally imposed rigidly-rotating barred potential. We found that hydro streamlines closely follow ballistic closed orbits at large and small radii, and tiny deviations of the hydro streamlines from ballistic closed orbits generate bar-driven spiral arms as kinematic density waves. At intermediate radii shocks arise and the streamlines shift between two families of closed orbits. We showed that the sound speed in the gas and the spatial resolution of the grid both affect the gas flow significantly.

In Chapter 3 we discussed the implications of our simulations for the interpretation of longitude-velocity diagrams (hereafter (l, v) diagrams) of atomic (HI) and molecular (CO, CS) gas in the Milky Way. The simulations presented there were based on the Binney et al. (1991) barred potential, which was originally used to construct a picture of the flow of gas through the central few kiloparsecs of our Galaxy on the assumption that gas follows closed orbits. Our simulations provided strong support for this assumption, but refined the Binney et al. (1991) picture of gas flow in several respects. Specifically: (i) in Binney et al. (1991) the parallelogram in the (l, v) plot for CO was interpreted as the trace of the cusped orbit, while we found that the shocks form two sides of the CO parallelogram, and conjectured that the prominence of the CO parallelogram is due to efficient conversion of atomic gas into molecular gas at the shocks; (ii) Binney et al. (1991) did not have an explanation for the observed asymmetry in the distribution of molecular emission near the Galactic Centre, while we argued that a promising explanation for the asymmetry is provided by the way the *wiggle instability* makes the flow through the shocks unstable. The large fluctuations generated by this instability might cause the conversion efficiency to fluctuate wildly and give rise to gross asymmetry in the distribution of molecular gas.

However, since we did not keep track of the chemistry of the ISM, items (i) and (ii) remained merely promising conjectures.

In Chapter 3 we identified two key features of the observed (l, v) diagrams that were still unexplained after their revision of the Binney et al. (1991) model and that were likely to be explained simply by a better choice of the potential: (i) coherent broad features like the 3 kpc arm and its counterpart on the far side of the Galaxy (Dame & Thaddeus, 2008) – these where absent from the simulations in Chapter 3; (ii) forbidden emission at large longitudes – in the models in Chapter 3 forbidden emission covers a significantly smaller portion of the (l, v) diagram than in the data.

In Chapter 3 we suggested two main directions for improving the models: first, inclusion of a law for the conversion of gas between atomic and molecular forms, so that, when gas is compressed at a shock, much of it is converted to molecular gas; second, modification of the quadrupole moment of the bar, since a higher quadrupole moment should generate stronger spiral arms and stronger non-circular motions, which are the likely explanations for internal features and forbidden velocity emission in the observed (l, v) diagrams. In Chapter 4 we explored the mechanism by which the bar generates the spirals that are responsible for internal structure in the (l, v) plane. Consequently, we here implement the second upgrade recommended in Chapter 3 by systematically exploring the effects on the gas flow of a variation of the quadrupole component of the potential, and discuss the consequence of this variation for the interpretation of Milky Way (l, v) diagrams.

In Section 5.2 we explain our numerical methods. In Section 5.3 we enumerate the bar's signatures in the (l, v) plane. In Section 5.4 we describe a reference model, which we use to relate structures in the Galactic plane to those in the (l, v) plot. In Section 5.5 we study how the structure of the (l, v) plane changes as we vary the parameters defining the bar. In Section 5.6 we outline the extent to which the Galactic bar is constrained by the observed (l, v) diagrams, and in Section 5.7 we sum up and consider promising directions for future work.

5.2 Methods

5.2.1 The potential

We assume that the gas flows in a simple externally-imposed two-dimensional barred potential that rotates at constant pattern speed $\Omega_{\rm p}$. The potential can be expanded in multipoles

$$\Phi(R, \phi) = \Phi_0(R) + \sum_{m=1}^{\infty} \Phi_m(R) \cos(m\phi + \phi_m)\,, \tag{5.1}$$

where $\{R, \phi\}$ are planar polar coordinates, the phases ϕ_m are constants and the amplitudes Φ_m are functions of R only. We assume that the potential comprises only monopole and quadrupole terms, so

$$\Phi(R, \phi) = \Phi_0(R) + \Phi_2(R) \cos(2\phi)\,. \tag{5.2}$$

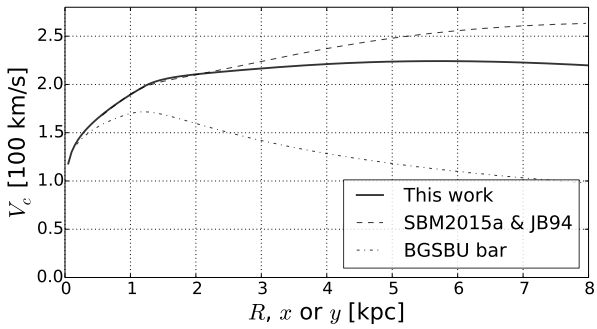


Figure 5.1: Circular speeds. Solid Line: potential used in this chapter. Dashed line: potential used in Chapter 3 and Jenkins & Binney (1994). Dot-dashed line: potential used in Binney et al. (1991). All the curves in this picture are calculated from the monopole component of each potential.

The solid line in Fig. 5.1 shows the circular speed v_c implied by the monopole Φ_0 used in this chapter, while the dashed line shows v_c for the monopole used in Chapter 3 and by Jenkins & Binney (1994). The latter monopole is unrealistic at large R because its circular speed becomes excessive, whereas the monopole used here coincides with the monopole of Chapter 3 at small radii but at greater radii generates a circular speed that plateaus at $220\,\mathrm{km\,s^{-1}}$. The dot-dashed line shows for comparison the circular speed of the Binney et al. (1991) potential. The potential used in Chapter 3 differs from the potential used in Binney et al. (1991) only by the addition of an axisymmetric disc component.

Our quadrupole Φ_2 is generated by the density distribution

$$\rho_2(r,\phi,\theta) = \frac{KA}{r_\mathrm{q}^2}\exp\left(-\frac{2r}{r_\mathrm{q}}\right)\sin^2\theta\,\cos(2\phi)\;, \tag{5.3}$$

where $\{r,\theta,\phi\}$ are spherical coordinates with the $\theta=0$ axis pointing towards the north Galactic pole. The constants are

$$K \equiv \frac{v_0}{4\pi G}\exp(2)\;, \tag{5.4}$$

where $v_0 = 220\ \mathrm{km\,s^{-1}}$ and G is the gravitational constant. The two main free parameters of ρ_2 are the quadrupole strength A and the quadrupole length r_q. We have chosen the form of the quadrupole density distribution to be exponential as recent infrared photometry has found that the Milky Way bar density profile is roughly exponential (Wegg & Gerhard, 2013).

Since ρ_2 is proportional to the real part of the spherical harmonic Y_2^2, which is an eigenfunction of the Laplacian operator, the density distribution (5.3) gives rise to the potential

$$\Phi_2(r,\phi,\theta) = \Phi_2(R)\sin^2\theta\,\cos(2\phi)\;. \tag{5.5}$$

We calculated $\Phi_2(r,\phi,\theta)$ numerically using the same code and methods used in Chapter 3. Our simulations are two-dimensional, so we only evaluate Φ_2 in the plane $\theta=\pi/2$. Thus, while the 3D density distribution given by Eq. (5.3) should not be considered a realistic density distribution for our Galaxy, the resulting

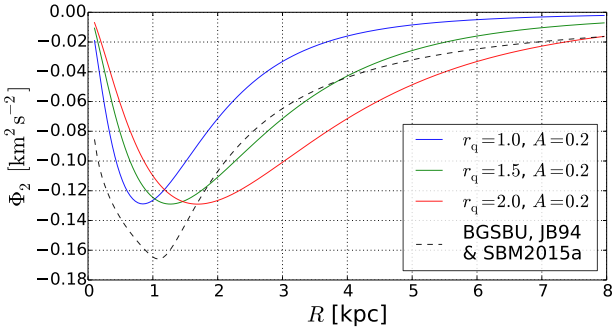


Figure 5.2: The quadrupole component Φ_2. In full lines, for the class of potentials used in this chapter, for the three different values of the bar length r_q and for a bar strength $A = 0.2$. The curves for different values of A are not shown but can be easily obtained by scaling linearly the curves shown. In dashed line the quadrupole component of the bar used in Chapter 3, Binney et al. (1991) and in Jenkins & Binney (1994) is shown for comparison.

2D potential in the plane can be obtained by a more realistic 3D density distribution. To be physical, a potential must come from a non-negative density distribution. We have checked that when our monopole is realised through a spherical symmetric distribution, the total density distribution is positive for all the values of the parameters used in this chapter.

5.2.2 Hydro Simulation Scheme

Our simulations methods are the same as in Chapters 3 and 4. We briefly review these here. We use a grid-based, Eulerian code based on the second-order flux-splitting scheme developed by van Albada et al. (1982). We assume that the gas is a two-dimensional inviscid isothermal fluid governed by the Euler equations (Eqs. 3.1). In this chapter we assume a sound speed $c_s = 10\,\mathrm{km\,s^{-1}}$. An additional term is introduced in the continuity equation to implement the recycling law of Athanassoula (1992). As before, the only effect of the recycling law is to prevent too much gas from accumulating in the very centre and to replace gas lost at the boundary due to the outflow boundary conditions. It does not affect the morphology of the results, so our results do not change if we disable the recycling law. We adopt the same recycling efficiency $\alpha = 0.3\,M_\odot\,\mathrm{pc^{-2}\,Gyr^{-1}}$ and initial density $\rho_0 = 1\,M_\odot\,\mathrm{pc^{-2}}$ as in previous chapters.

For the simulations in this chapter we used a grid $N \times N$ to simulate a square 20 kpc on a side, where N defines the resolution of the simulation. We start with gas in equilibrium on circular orbits in an axisymmetrized potential and, to avoid transients, turn on the non-axisymmetric part of the potential gradually during the first 615 Myr.

In most of our simulations the gas is non self-gravitating. In Appendix C we study the impact of the self-gravity, and for this study we disable the recycling law by setting $\alpha = 0$.

5.2.3 Projecting to the (l, v) plane

We adopt the same very simple projection procedure used in Chapter 3 to produce the predicted (l, v) distributions for each simulation snapshot $(\rho(\mathbf{x}), \mathbf{v}(\mathbf{x}))$. This procedure yields a brightness temperature that

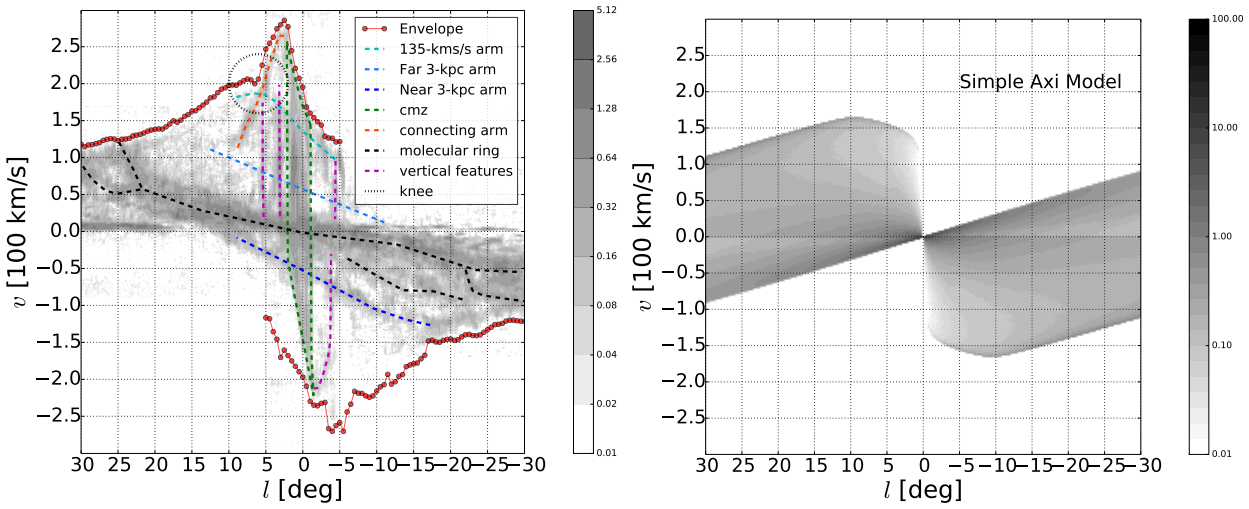


Figure 5.3: Left panel: CO observations with features superimposed. The red dots trace the envelope determined from HI data as explained in Chapter 2. Dashed lines trace internal features. The colorbar is in units of K. Right panel: appearance of the diagram in the absence of the quadrupole potential under the assumption that the gas density is proportional to $\exp(-R/2R_0)$ and that the circular velocity curve is given by the solid line in Fig. 5.1. The colorbar is in arbitrary units.

is linear in column density which is equivalent to the simplest radiative transfer calculation (optically thin case, see Chapter 3).

As before, we assume that the Sun is undergoing circular motion at a radius $R_0 = 8\,\mathrm{kpc}$ with speed $\Theta_0 = 220\,\mathrm{km\,s^{-1}}$. Calling ϕ the angle between the major axis and the Sun-Galactic Centre (GC) line, the Cartesian coordinates of the Sun are given by $x_\odot = R_0 \cos\phi$, $y_\odot = R_0 \sin\phi$. In our models, we only project material inside the solar circle. This means that our models cannot reproduce the non-forbidden emission in the two quadrants $l > 0$, $v < 0$ and $l < 0$, $v > 0$, except for some occasional emission due to non circular motions of material close to the Solar circle. However the focus of this chapter, as most of this thesis, is on the inner regions.

The resolution of our (l, v) diagrams is $\Delta l = 0.25°$ in longitude and $\Delta v = 2.5\,\mathrm{km\,s^{-1}}$ in velocity. The (l, v) intensities for a given line of sight are produced in the following way. Along the line of sight, we sample the xy density and the velocity by linearly interpolating the results of the simulations at points separated by $\delta s = 1\,\mathrm{pc}$. These points are then distributed in velocity bins of width $\Delta v = 2.5\,\mathrm{km\,s^{-1}}$. The final (l, v) intensity at the chosen longitude in each range of velocity is obtained by summing over all the relevant points along the line of sight weighted by their densities.

5.3 Signatures of the bar

In this section we briefly review the observational signatures of the Milky Way bar that we will use later in the chapter to compare our models with observations. A longer account was given in Chapter 2.

In the left panel of Fig. 5.3 the grey scale shows the intensity in the (l, v) plane of 2.6 mm radiation by CO at $|l| < 30°$. Coloured lines schematically trace features identifiable in the CO and 21-cm HI (l, v) diagrams. The right panel shows a simple model similar that shown in Fig. 2.1. The main difference between

the two simple models lies in the rotation curve: in the model in Fig. 2.1 it is completely flat, while here is the solid line in Fig. 5.1. Also the density distribution is different in the two models, being uniform in the first and exponentially decreasing in the second. The only noticeable consequence that this produces in the (l, v) plots is that the envelopes in the two quadrants $l > 0$, $v < 0$ and $l < 0$, $v > 0$ are sharper for the model with uniform density. However, as noted above, non-forbidden emission in these two quadrants is not the focus of this chapter.

The simple model in the right panel of Fig. 5.3 shows the emission we would expect in the absence of the quadrupole Φ_2. The latter has a big impact on (l, v) diagrams and we explore how the features marked in the upper panel can be used to constrain Φ_2.

The key features of the observed (l, v) plots are

1. *Emission at forbidden velocities.* If gas everywhere moved on circles, emission in the quadrants $(l > 0, v < 0)$ and $(l < 0, v > 0)$ could not be produced by material inside the Sun. Material that lies outside the Sun can produce emission in these quadrants, but at $|l| < 30°$, such emission is confined to small values of $|v|$. Velocities in the quadrants $(l > 0, v < 0)$ and $(l < 0, v > 0)$ that cannot be reached by gas outside the Sun are said to be "forbidden".

   For example, practically all the emission at $|l| < 5°$, $v > 100\,\mathrm{km\,s^{-1}}$ in Fig. 5.4 comes from inside the Solar circle (but the forbidden emission extends also up to $|l| \simeq 8°$). Forbidden emission is automatically generated by Φ_2.

2. *Velocity peaks.* At $|l| \sim 2°$ in Fig. 5.4 emission is seen at very high velocities, $|v| \simeq 270\,\mathrm{km\,s^{-1}}$. At one time these peaks were considered evidence of a very centrally concentrated monopole component of the Galactic potential, but since Binney et al. (1991) it has been widely accepted that they are generated by Φ_2.

3. *Internal features.* The right panel of Fig. 5.4 shows a complex pattern of ridges of enhanced emission. Fig. 5.3 gives the conventional names of the more important structures. These features would be absent if the gas distribution were axisymmetric (right panel of Fig. 5.3).

4. *The envelope.* The red dots in Fig. 5.3 trace the envelope of the observed emission, determined as explained in Sect. 2.3.1. The envelope carries the imprint of Φ_2 in three ways: through the velocity peaks, the emission at forbidden velocities, and the shoulders and bumps that arise where an internal feature touches the envelope.

5. *Bumps on the envelope.* A mechanism by which spiral arms generate bumps on the envelope is described in §9.1.2 of Binney & Merrifield (1998). We identify as particularly interesting the *knee* marked with a circle in Fig. 5.3. This knee occurs where the envelope shifts from touching the arm at *135-km/s* to touching the *connecting arm* (green dashed and light red dashed curves in Fig. 5.3). As noted in Chapter 2, the connecting arm, which contributes to the envelope on the low-l side of the knee, appears only at $b < 0$, while the arm at $135\,\mathrm{km\,s^{-1}}$, which forms the *envelope* on the other side, appears only at

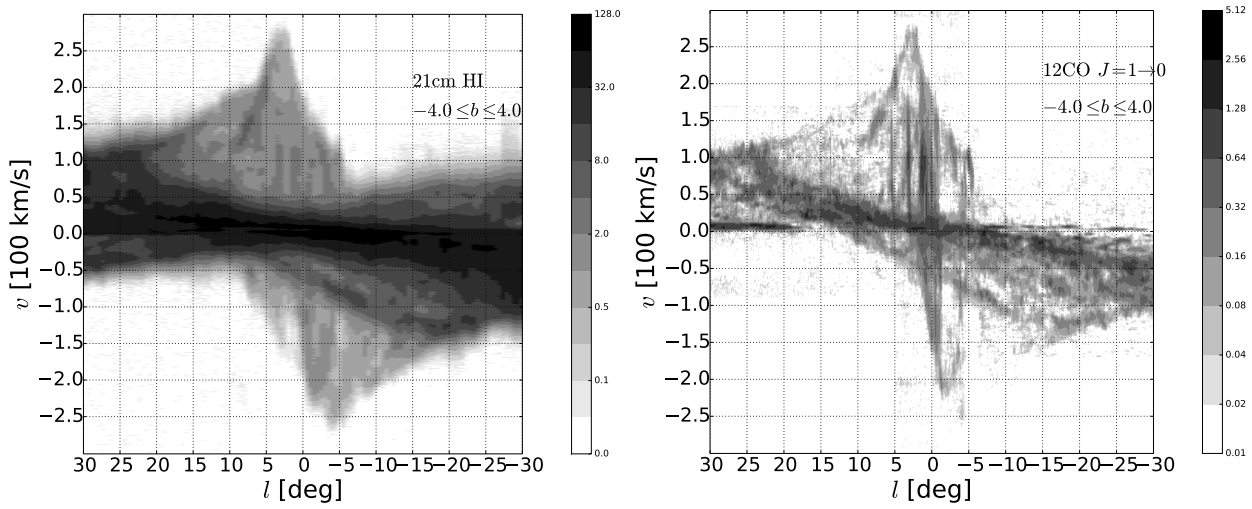


Figure 5.4: Left panel: HI observations integrated over $|b| < 4°$. Right panel: CO observations integrated over $|b| < 4°$. Colorbars are in units of K.

$b \geqslant 0$. This relative offset in b suggests that the gas disc is moving to larger b as one moves outwards along the black dotted line in the top-right panel of Fig. 5.5. Since the *connecting arm* and the arm at $135\,\mathrm{km\,s^{-1}}$ appear at distinct latitudes, they should represent distinct dynamical features. Thus the *knee*, which arises from the transition between these structures, cannot reflect merely a sudden change in the circular speed. Other interesting bumps occur when the near and far 3 kpc arms touch the envelope.

Towards the end of the Chapter we will return to this list to review how items constrain the Galactic bar.

5.4 A reference model

In this Section, we describe in detail a "reference" model. Table 5.1 gives the values of its defining parameters. Note that the model's pattern speed, $\Omega_\mathrm{p} = 40\,\mathrm{km\,s^{-1}\,kpc^{-1}}$, is lower than that, $63\,\mathrm{km\,s^{-1}\,kpc^{-1}}$, used in Chapter 3. The reference model has not been chosen because its parameters give the best fit to the observations; rather, it is just a representative model that displays many of the characteristic of interest for this chapter. In later sections, we will see how these characteristics change as a function of the model's parameters.

The top-left panel of Fig. 5.5 shows the gas density reached after a long time. The bottom-left panel of Fig. 5.5 shows the associated (l, v) plot for a bar angle $\phi = 20°$. The top-right panel marks with coloured lines prominent features in the top-left panel. The lower-right panel shows the location of these same features in the (l, v) plane.

The general characteristics of the gas flow can be understood as described in Chapters 3 and 4 and we now briefly summarise their picture. In the outer regions, where spiral arms are present, the velocity field of the gas is well approximated by elongated, closed ballistic x_1 orbits. However, the agreement is not perfect because the streamlines execute small librations around the underlying x_1 orbits. These librations generate

spiral arms as kinematic density waves. While flowing approximately on x_1 orbits, the gas also slowly drifts inwards. At a transition point, the gas stops following x_1 orbits and two straight offset shocks form. After passing the shocks, the gas plunges towards the central x_2 disc, where it settles and flows to a very good approximation on x_2 orbits. The x_2 disc is shown as the central grey structure in the top-right panel of Fig. 5.5, and the two straight shocks are the red and blue features that just touch the grey structure.

In contrast to the (l, v) plot generated by circular motion (right panel of Fig. 5.3), the (l, v) plot of the reference model (bottom-left panel of Fig. 5.5) bears a strong resemblance to the observed (l, v) plots. Because we are interested in the region inside the solar circle, we ignore emission in the quadrants $(l > 0, v < 0)$, $(l < 0, v > 0)$ other than emission at small $|l|$ and forbidden velocities. The position and strength of the velocity peaks of the reference model ($|l| \simeq 3°$, $v \simeq 280\,\mathrm{km\,s^{-1}}$) match well the peaks seen in the observational plots. The region covered by forbidden velocities in the reference model matches reasonably well that found in observations. It is a significant improvement over the model presented in Chapter 3, although in the reference model the envelope profile in the forbidden emission region is somewhat too steep.

The envelope of the reference model displays bumps wherever the envelope is touched by one of the lines that fan out from the centre. Each line is the projection of a bar-driven spiral arm, as can be seen comparing the upper and lower right-hand panels in Fig. 5.5. The stronger a spiral arm is, the larger the bump it creates on the envelope.

The black dots in the upper-right panel of Fig. 5.5 show the points in the Galactic plane that provide the emission that forms the envelope in the (l, v) plane. Each time a spiral arm in the (l, v) plane becomes tangent to the envelope, the black dots make a discontinuous jump in the Galactic plane. Particularly strong bumps in the (l, v) plane and large jumps in the Galactic plane occur at $l \simeq 18°$,$v \simeq 150\,\mathrm{km\,s^{-1}}$, $l \simeq -15°$,$v \simeq -160\,\mathrm{km\,s^{-1}}$ and $l \simeq -20°$,$v \simeq -140\,\mathrm{km\,s^{-1}}$. In Section 5.3, item (5), we have noted the presence of similar bumps in the observed envelope, including the knee circled in Fig. 5.3. A faint bump in the reference model at $l \simeq 9°$,$v \simeq 210\,\mathrm{km\,s^{-1}}$ is close to the right position, but is much less strong than the observed knee. We will see below that it can be made stronger by varying the quadrupole. Another bump is present in the observations where the 3 kpc arm touches the envelope at $|l| \simeq -18°$, $v = -150\,\mathrm{km\,s^{-1}}$. This is quite similar to the one present in the reference model at $l \simeq -20°$,$v \simeq -140\,\mathrm{km\,s^{-1}}$.

In the model of Chapter 3, bumps were present but were weaker than observed because the spiral arms were barely discernible in the (l, v) plane. The reference model is a significant improvement, but its knee needs to be strengthened and the positions of its bumps should be tweaked.

Some of the "arms" in the observed (l, v) plots can be identified with features in the reference (l, v) plot (Fig. 5.5). For example, the 3 kpc arm and its far-side counterparts are similar to the outermost red arm and its counterpart on the other side, the outermost blue arm. The connecting arm is well traced by part of the innermost green arm, and the arm at $135\,\mathrm{km\,s^{-1}}$ is similar to its outer neighbour. Not all arms of the reference model have a counterpart in the observations, but it appears as if all the principal features that contain *arm* in their name have a counterpart in the reference model.

The shocks are traced by the red and blue lines that just touch the x_2 disc in the top-right panel of

Fig. 5.5. According to the picture described in Chapter 3, the shocks should form two sides of the CO parallelogram, and the projection of the x_2 disc should match the region of CS emission. The reference model meets this expectation in that both the size of the x_2 disc and the position of the shocks in the (l, v) plane are right. The shocks are not as prominent in the reference (l, v) plane as in the observed one, probably because the model does not include conversion in shocks of gas to molecular form (see Chapter 3).

We note that the shocks are very narrow in the Galactic plane, but when projected to the (l, v) plane they extend over some range in longitude. This suggests that the vertical features are different portions of the two shocks. However, there appear to be two main problems with this interpretation: i) the vertical features are more distinct than Fig. 5.5 would suggest; ii) some vertical features lie in the forbidden quadrant $(l < 0, v > 0)$, while the red and blue shock lanes in the lower-right panel of Fig. 5.5 do not extend to the forbidden velocity regions. We conjecture that these shortcomings may be overcome once we take into account the efficient conversion of material from atomic to molecular form at the shocks and the instability discussed in Chapter 3. In general the distribution of gas in the Galaxy is not as smooth as in our models, and gas can be more concentrated along regions that when projected produce different vertical features. By including a law for the conversion of gas to molecular form, molecular clouds will be stochastically produced at shocks, and fluctuations are likely. This might solve problem (i). When the instability discussed in Chapter 3 is present, gas running down the shocks does not necessarily stop when it meets the edge of the x_2 disc, but it can have enough momentum to go beyond it, ending in a region that produces forbidden emission (see for example the panel $c_{\rm s} = 20$, d$x = 10$ in Fig. 3.2 and the middle panels in Fig. 3.9). This might solve problem (ii) (note also that even for stable simulations the shocks can partially extend to the forbidden velocity region, see for example Fig. 3.13). Thus, our conjecture on the origin of the vertical features should be tested alongside the conjectures made in Chapter 3 on the origin of the asymmetry and the prominence of the two sides of the CO parallelogram.

The molecular ring, schematically shown in Fig. 5.3, is well reproduced by our reference model, which becomes darker along a similar diagonal band in the (l, v) plane. The darkness of the molecular ring in the (l, v) plane is largely a consequence of velocity crowding: extended areas of material outside the outermost spiral arms in the top-left panel of Fig. 5.5 project to similar values of (l, v) and hence produce bright intensities in the (l, v) plane. This is an improvement over the model of Chapter 3, which generates an (l, v) diagram in which the dark band runs too steeply because the gas that generates it is too centrally concentrated, so it projects to low longitudes and high velocities.

We have seen that the reference model, qualitatively, manifests most of the observational signatures of the bar. The next step is to make the correspondence between the model and observed (l, v) plots quantitatively satisfying. This proves to be a challenging task because changing the model's physical parameters often improves the fit of one feature to the detriment of another. In the next section we study the dependencies of features on the parameters that define the bar (A, $r_{\rm q}$, $\Omega_{\rm p}$).

Table 5.1: Parameters of the reference model described in detail in Sec. 5.4.

model	A	r_q [kpc]	Ω_p [$\mathrm{km\,s^{-1}\,kpc^{-1}}$]	c_s [$\mathrm{km\,s^{-1}}$]	dx [kpc]
Reference	0.6	1.5	40	10	20

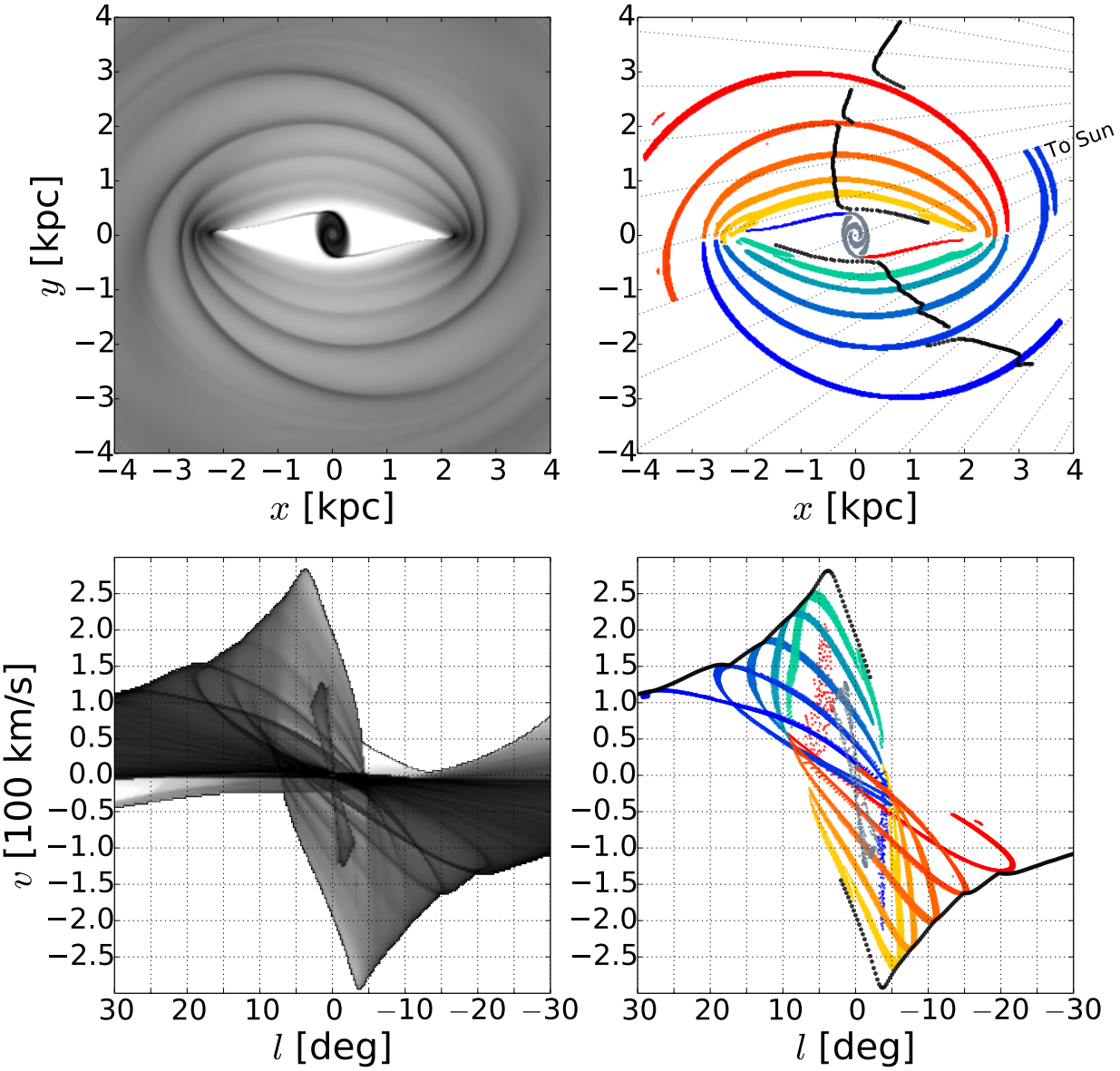


Figure 5.5: The reference model. Top-left: the gas density in the xy plane. Bottom-left: the projection of the reference model to the (l, v) plane assuming an angle $\phi = 20°$. In the two right panels the same features are shown in xy and (l, v) plane using the same colour coding. The black dots trace the points corresponding to the envelope in the (l, v) plane.

5.5 Effects of varying the quadrupole

First we investigate changes to the potential's quadrupole, which is characterised by three parameters: the strength A, the scale length $r_{\rm q}$ and the pattern speed $\Omega_{\rm p}$. The values of these parameters that we have explored are shown in Table 5.2.

Figs. 5.6 and 5.8 show the gas density for different values of quadrupole strength A and length $r_{\rm q}$ and for values of the pattern speed $\Omega_{\rm p} = 40\,{\rm km\,s^{-1}\,kpc^{-1}}$ and $\Omega_{\rm p} = 60\,{\rm km\,s^{-1}\,kpc^{-1}}$, respectively. Figs. 5.7 and 5.9 show the projections of the same snapshots onto the (l, v) plane, assuming bar angle $\phi = 20^\circ$. All snapshots are taken at $t = 1.25$Gyr, when all simulations have reached an approximate steady state (apart from any unsteadiness caused by the *wiggle instability*, see Chapter 3). The reference model has $\Omega_{\rm p} = 40\,{\rm km\,s^{-1}\,kpc^{-1}}$ and in Figs. 5.6 and 5.7 lies in the panel marked with $\{A = 0.6, r_{\rm q} = 1.5\}$. In order to contain the length of the main text of the chapter, we have relegated the results obtained for other values of the pattern speed to Appendix B.

The pattern speed controls the position of the resonances: increasing the pattern speed pushes all the resonances inwards. This is reflected in the morphology of gas flow; many characteristics of the gas flow roughly scale with the locations of the resonances. For example, comparing Fig. 5.6 with Fig. 5.8 demonstrates that increasing the pattern speed pushes spiral arms and the shocks inwards. Table 5.3 lists the position of the resonances for the explored values of the pattern speed.

From Figs. 5.6 and 5.8 we see that the morphology of the arms, the size of the x_2 disc, and the transition point where shocks are formed also depend on both A and $r_{\rm q}$. Consider the model in the top-left panel of Fig. 5.6, $\{A = 0.2, r_{\rm q} = 1\}$. This model is rather flat and featureless, and this is also reflected in the (l, v) projection (top-left panel of Fig. 5.7), which does not show anything resembling the internal features discussed in Section 5.3. The model lacks high-velocity peaks, and its x_2 disc is too big to coincide with the observed CS emission. It is, however, likely that higher velocity peaks and a smaller x_2 disc would be produced by an increase in sound speed or in resolution (Chapter 3).

By moving down the left column of Fig. 5.6, we increase the bar strength A at fixed bar length $r_{\rm q} = 1\,{\rm kpc}$. The shock region changes causing the x_2 disc to shrink, but the spiral arms remain weak. In the (l, v) plane (left column of Fig. 5.7), the envelope of the emission changes significantly in a way similar to what would be obtained by an increase in sound speed or resolution, but features due to spiral arms remain weak and the resulting diagrams are rather featureless compared to that of the reference model (Fig. 5.5). This happens because spiral arms appear outside the shock region and are weak when the quadrupole is weak there, as it is when $\Omega_{\rm p} = 40\,{\rm km\,s^{-1}\,kpc^{-1}}$ and $r_{\rm q} = 1\,{\rm kpc}$ (Fig. 5.2). To strengthen the spirals we need a substantial amount of quadrupole in the region where they are present, hence we must either extend the quadrupole by increasing $r_{\rm q}$, or bring the spirals in by increasing $\Omega_{\rm p}$. The efficacy of the first strategy is illustrated by moving horizontally in either Fig. 5.6 or Fig. 5.7, while the efficacy of the second strategy can be seen by comparing equivalent panels in Figs. 5.6 and 5.8.

The envelope, by contrast, depends on the flow in the vicinity of the shocks, where the quadrupole

Table 5.2: Values of the parameters explored in our study.

A	$r_{\rm q}$ [kpc]	$\Omega_{\rm p}$ [$\rm km\,s^{-1}\,kpc^{-1}$]
$\{0.2, 0.4, 0.6, 0.8\}$	$\{1.0, 1.5, 2.0\}$	$\{20, 30, 40, 50, 60, 70\}$

strength peaks, and is strongly affected by change in A. When the quadrupole is strong ($A = \{0.6, 0.8\}$) the gas reaches high velocities and a bigger portion of the forbidden velocity region is covered. Hence the extent of emission at forbidden velocities strongly constrains A, but barely constrains $r_{\rm q}$.

Stellar dynamics teaches us that bars cannot extend extend beyond the corotation radius (e.g. Sellwood & Wilkinson, 1993), and for any given value of $\Omega_{\rm p}$ this fact sets an upper limit on $r_{\rm q}$. The models in the right-hand column of Fig. 5.8 strongly violate this constraint and differ significantly from what it would be obtainable using a self-consistent stellar dynamical model.[1] These are the models in which the spiral arms are so strong that they have become wiggle-unstable shocks.

The bumps on the envelope are strictly connected with the strength of the spiral arms, and their positions are regulated by the parameters that characterise the bar (see, for example, the models with $A = \{0.6, 0.8\}$ in Fig. 5.7).

One of the motivation of the present study was to improve the model in Chapter 3, which has $\Omega_{\rm p} = 63\,\rm km\,s^{-1}\,kpc^{-1}$ and a quadrupole very similar to that obtained with $\{A = 0.2, r_{\rm q} = 1.5\}$ (Fig. 5.2). Therefore, let us compare the Chapter 3 model, computed with the same spatial resolution and sound speed, with the panels for $\{A = 0.2, r_{\rm q} = 1.5\}$ in Figs. 5.8 and 5.9. The two models are very similar in all aspects: the spiral arms, the transition point, and the size of the x_2 disc. One of the problems of the Chapter 3 model was the lack of spiral arms and consequently of internal features in the (l, v) plane. Fig. 5.8 shows that increasing the strength of the quadrupole produces internal features. Another problem of the Chapter 3 model was that it provided insufficient coverage of forbidden velocities. Fig. 5.9 shows that changing the quadrupole strength and length do not cure this problem. A decrease in the pattern speed is needed. We will discuss in more detail the relation to observations in the next section.

5.6 Implications of the observations

In this section we discuss what our simulations can say about the observations, and in particular about items listed in Section 5.3.

1. *Emission at forbidden velocities.* The extent to which emission extends into the forbidden zone of the (l, v) diagram depends strongly on the pattern speed, and more weakly on the strength and length of the quadrupole. Lower pattern speeds produce more emission in the forbidden zone. When

[1] Nevertheless, it is interesting to see what the gas flow becomes if we push some of the parameters to such unrealistic values. Indeed, one of the goals of this chapter is to understand what the gas response is when we change the underlying potential, and in order to understand this not necessarily we must restrict ourselves to realistic potentials.

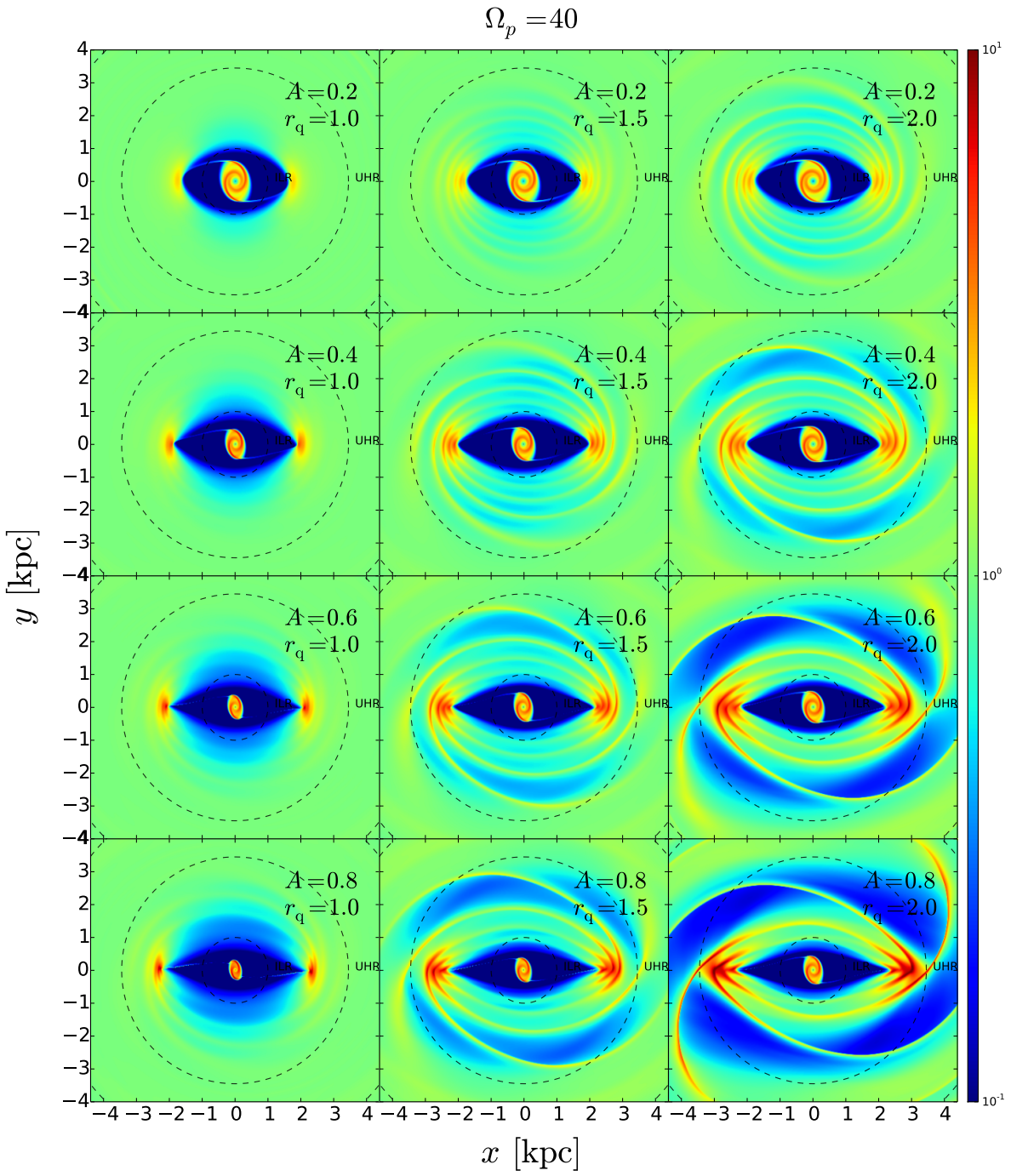


Figure 5.6: The gas surface density in hydro simulations for different values of the quadrupole strength A and quadrupole length r_q. r_q is increasing left to right taking values 1.0, 1.5 and 2.0 kpc. A is decreasing from top to bottom taking values 0.2, 0.4, 0.6 and 0.8. A and r_q are defined in Eq. (5.3). All snapshots are for a value of the pattern speed $\Omega_p = 40$ km s^{-1} kpc^{-1}. The major axis of the bar is aligned horizontally, and gas has reached an approximately steady state in the rotating frame and circulates clockwise. All snapshots are taken at $t = 1.25$ Gyr. The dotted circles mark the positions of the resonances, calculated from the monopole. Since the monopole and Ω_p are the same in all panels, the positions of the resonances are identical for all of them. The colorbar is in units of $M_\odot$ pc^{-2}.

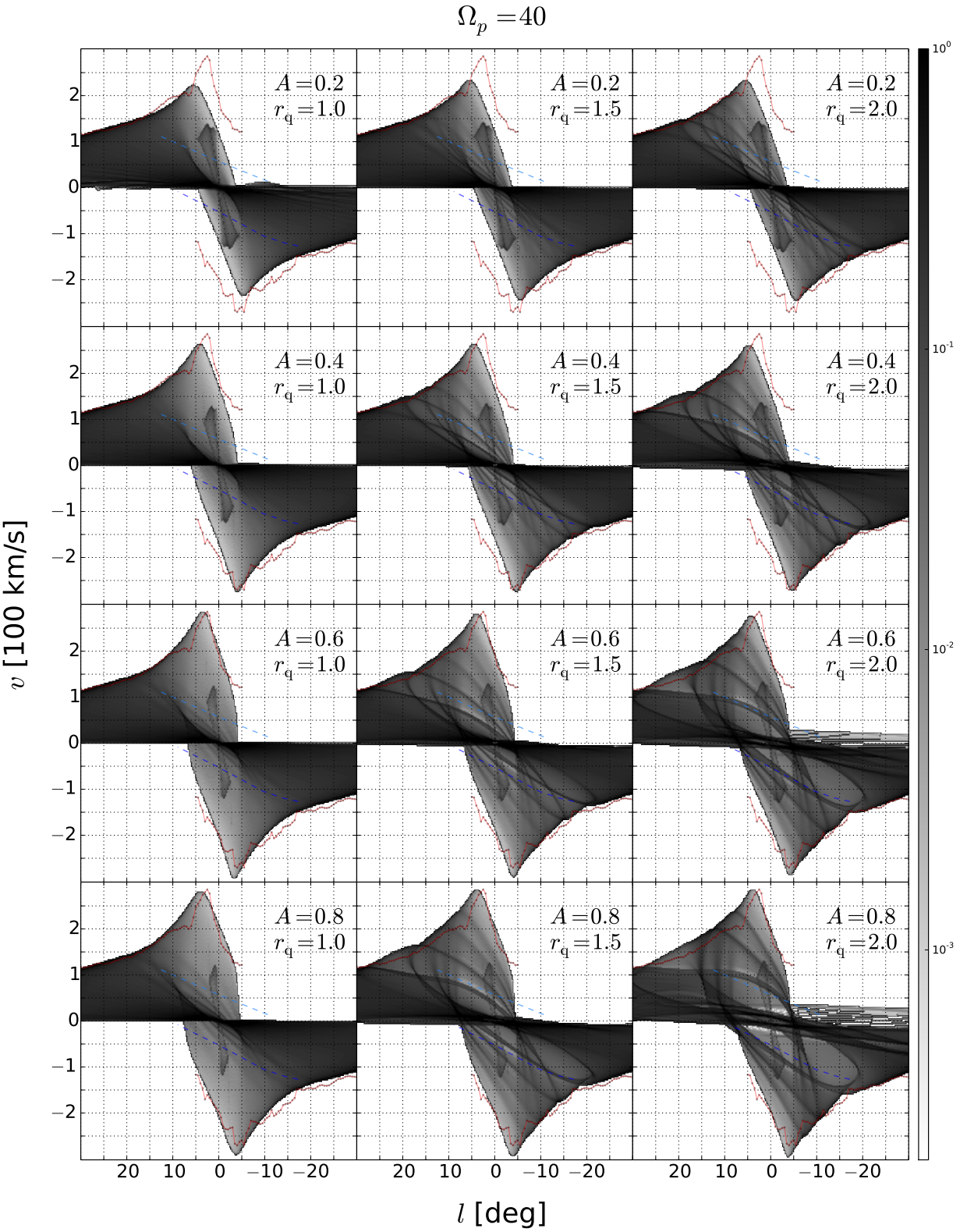


Figure 5.7: The simulations of Fig. 5.6 projected onto the (l, v) plane. The Sun is assumed to be in a circular orbit with $v = 220\,\mathrm{km\,s^{-1}}$ at $R_0 = 8\,\mathrm{kpc}$, and the bar major axis makes an angle $\phi = 20°$ with the Sun-Galactic Centre line. The red dots trace the envelope of the observations, while the blue dashed lines indicate the positions of the near and far 3 kpc arms. The colorbar is in arbitrary units.

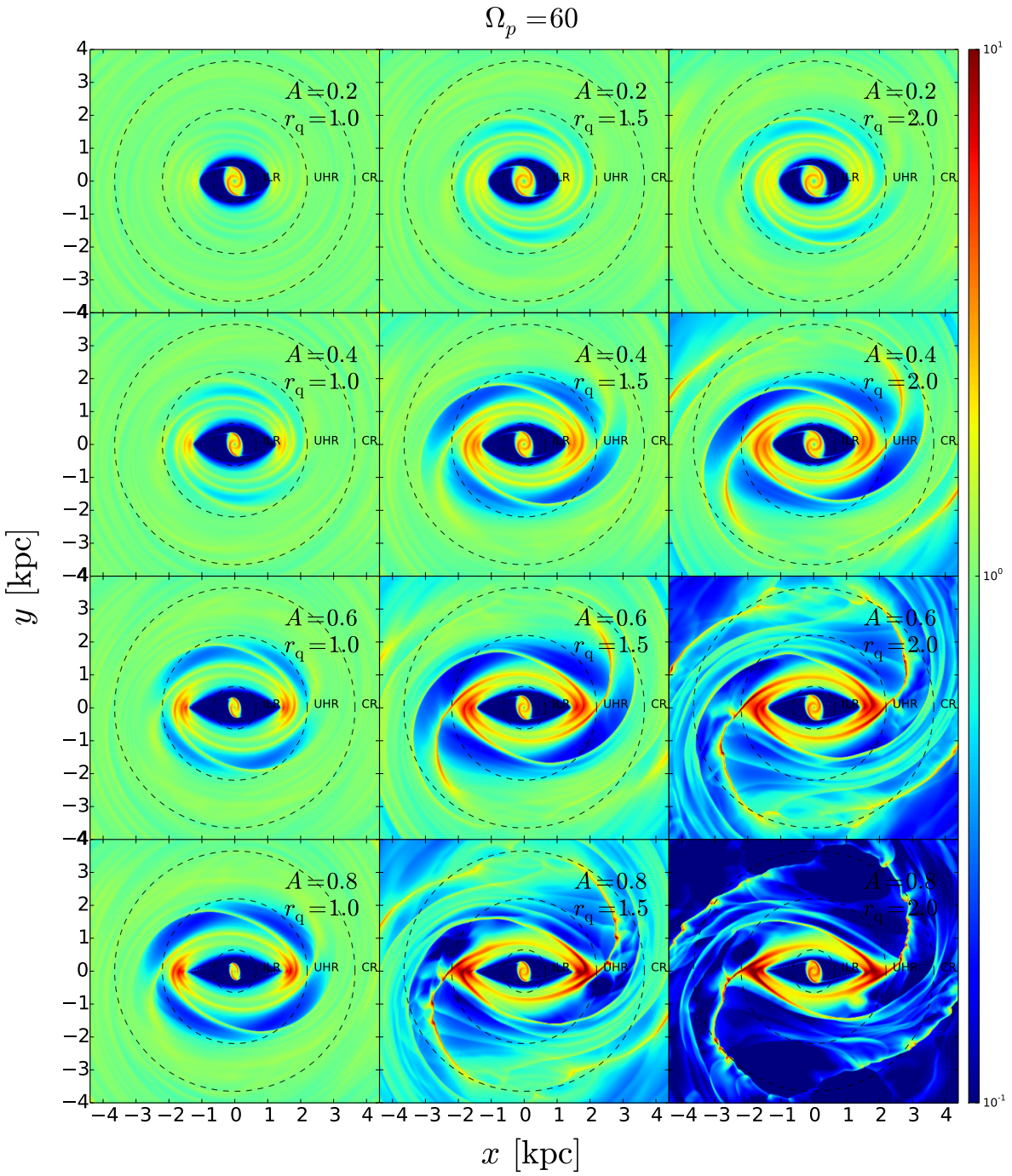


Figure 5.8: Same as Fig. 5.6 but for a pattern speed $\Omega_\mathrm{p} = 60\ \mathrm{km\,s^{-1}\,kpc^{-1}}$.

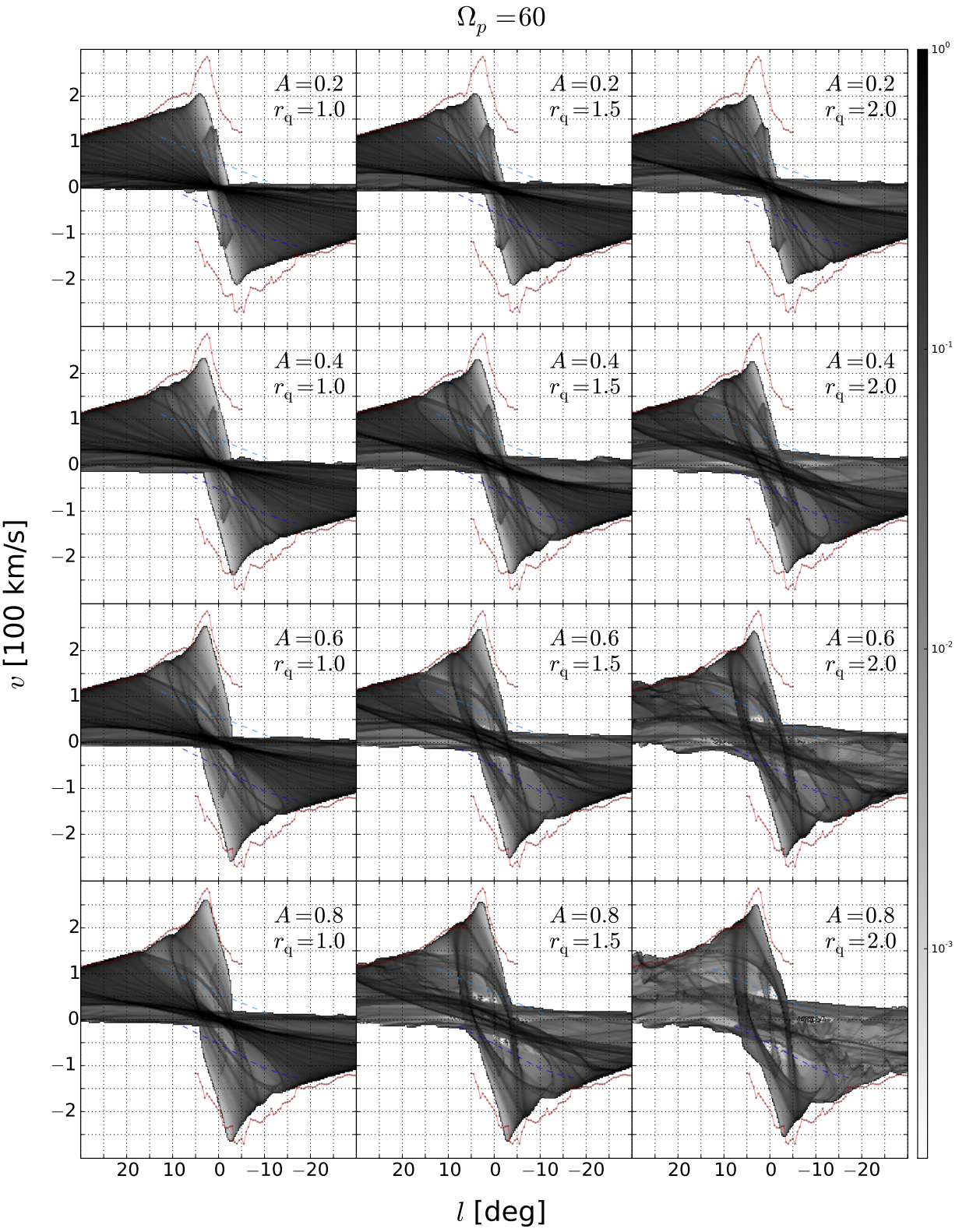


Figure 5.9: Same as Fig. 5.7 but referring to Fig. 5.8.

Table 5.3: Positions of resonances for different values of the pattern speed. ILR = inner Lindblad resonance, UHR = ultra harmonic resonance (i.e., 4/1), CR = corotation, OLR = outer Lindblad resonance. These are calculated from the monopole (black curve in Fig. 5.1).

Ω_p [$\mathrm{km\,s^{-1}\,kpc^{-1}}$]	ILR [kpc]	UHR [kpc]	CR [kpc]	OLR [kpc]
20	2.85	7.35	10.40	14.25
30	1.85	4.75	7.35	10.90
40	1.00	3.45	5.60	8.85
50	0.80	2.70	4.45	7.40
60	0.65	2.20	3.65	6.30
70	0.55	1.85	3.10	5.45

$\Omega_p \gtrsim 60\,\mathrm{km\,s^{-1}\,kpc^{-1}}$, the forbidden emission is insufficiently extended, regardless of the quadrupole's strength and length (Figs. 5.9 and B.8). This statement holds even if we allow the bar angle ϕ to vary. A good match with observations is obtained at pattern speeds of 30-40 $\mathrm{km\,s^{-1}\,kpc^{-1}}$ (Figs B.6 and 5.7). When the pattern speed is lower, $\Omega_p \lesssim 20\,\mathrm{km\,s^{-1}\,kpc^{-1}}$ (Fig. B.5), the region is clearly too big for bar angles $\phi \gtrsim 10^\circ$ that are compatible with photometric data (e.g. Gerhard & Wegg, 2014). In light of this finding, it is not surprising that Weiner & Sellwood (1999), who confined their fitting to the envelope, including the forbidden-velocity region, found $\Omega_p \simeq 42\,\mathrm{km\,s^{-1}\,kpc^{-1}}$. Thus, from the forbidden velocities alone, our simulations would suggest a pattern speed in the range 30-40 $\mathrm{km\,s^{-1}\,kpc^{-1}}$.

2. *Velocity peaks.* Discussion of these features is delicate because their structure is sensitive to the spatial resolution of simulations (Chapter 3). For example, in models with weak bars (e.g. top row in Fig. 5.9) the peaks are not reproduced, but the results of Chapter 3 show that they would be reproduced at higher resolution. The velocity peaks depend very weakly on the bar length (Figs. 5.7 and 5.9), and only marginally more strongly on the bar strength A, especially once we consider the effects of resolution. Increasing the pattern speed shifts the peaks to lower longitudes. For the chosen angle $\phi = 20^\circ$, the peak at positive velocity is at the right longitude when the pattern speed is high, $\Omega_p \simeq 60\,\mathrm{km\,s^{-1}\,kpc^{-1}}$, while it is at slightly too high longitudes when the pattern speed is lower. The peak at negative velocity instead is at the right longitude when the pattern speed is low, $\Omega_p \simeq 40\,\mathrm{km\,s^{-1}\,kpc^{-1}}$, and appears at slightly too low longitudes when the pattern speed is higher.

3. *Internal features.* Internal features provide some of the tightest constraints on the quadrupole's length and strength. As discussed in Section 5.5, a significant quadrupole is needed at $\sim 3\,\mathrm{kpc}$ to produce the observed arms (e.g. left column in Fig. 5.7). No model with a short bar, $r_q = 1$, produces the requisite quadrupole. Shifting the pattern speed significantly from $\Omega_p = 40\,\mathrm{km\,s^{-1}\,kpc^{-1}}$, in either direction, shifts the region in which spirals can form away from the radial range specified by the data. The 3 kpc arm extends to $|l| \simeq 20^\circ$, so a substantial quadrupole is required at the arm's tangent point, $R_0 \sin(20^\circ) \simeq 3\,\mathrm{kpc}$. If the bar is this long, stellar dynamics excludes pattern speeds as high as $\Omega_p = 60\,\mathrm{km\,s^{-1}\,kpc^{-1}}$. Thus, internal features strongly suggest that the bar pattern speed is around

$\Omega_p = 40\,\mathrm{km\,s^{-1}\,kpc^{-1}}$. This is perhaps why Rodriguez-Fernandez & Combes (2008), who used the 3 kpc arm as a fitting criterion, favoured a pattern speed in the range $\Omega_p = 30\text{-}40\,\mathrm{km\,s^{-1}\,kpc^{-1}}$. Other models with higher pattern speeds, (Englmaier & Gerhard, 1999; Bissantz et al., 2003) reproduce the internal features less well. The molecular ring is also better reproduced when the pattern speed is $\Omega_p = 40\,\mathrm{km\,s^{-1}\,kpc^{-1}}$. These models becomes darker along the right diagonal band in the (l, v) plane, similarly to the reference model in Sect. 5.4. When the pattern speed is higher or lower, this band becomes too much or not enough centrally concentrated respectively.

4. *The envelope.* In the positive-longitude permitted quadrant, $(l > 0, v > 0)$, the envelope is well matched by simulations with high pattern speed, $\Omega_p = 50\text{-}60\,\mathrm{km\,s^{-1}\,kpc^{-1}}$ (Figs. B.7 and 5.9). When the pattern speed is lower, (e.g., Fig. 5.7), the predicted envelope is generally higher than the observed one and the descent immediately after the peak towards larger longitudes is too shallow; in other words, the peaks are not sharp enough. In the other permitted quadrant however, $(l < 0, v < 0)$, the opposite is true and the envelope is well matched by simulations with low pattern speed, $\Omega_p = 40\,\mathrm{km\,s^{-1}\,kpc^{-1}}$, while at higher pattern speed the models do not reach adequately high velocities and are too steep. Regarding the envelope at forbidden velocities, a better match is obtained at low pattern speeds, as noted in item (1) above. Thus, the envelope is sending mixed message as regards the pattern speed. Perhaps tweaking the monopole and the bar angle ϕ one would obtain a good fit to the envelope in both the permitted quadrants for a low pattern speed. Although the mere presence of the quadrupole dramatically changes the envelope from its form in the axisymmetric case, the envelope's form is surprisingly insensitive to the quadrupole's exact length and strength.

5. *Bumps on the envelope.* The strength and position of the bumps in the envelope where it is touched by a spiral arm are sensitive to the quadrupole parameters A and r_q. The bump associated with the near 3 kpc arm is very well reproduced at low pattern speed (Fig. 5.7) when the bar is long and strong ($A \gtrsim 0.6$,$r_q \gtrsim 1.5$). The knee is also close to the right position in some of the same models - for example, in the model $\{A = 0.8, r_q = 2.0\}$. When the pattern speed is higher (Fig. 5.9) the knee is also well reproduced in some models but these models don't fit the 3 kpc arm and its bump.

Overall the observations are best fitted when the pattern speed is $\Omega_p \simeq 40\,\mathrm{km\,s^{-1}\,kpc^{-1}}$. Only one aspect is better explained with a higher pattern speed of $\Omega_p = 60\,\mathrm{km\,s^{-1}\,kpc^{-1}}$: the sharpness of the positive longitude velocity peak and the associated portion of the envelope in the positive-velocity permitted quadrant. It seems likely that these shortcomings can be resolved by tweaking the monopole component of the potential, as well as the other parameters that define the bar and the gas flow. Moreover, the internal features strongly suggest that the bar is longer than it could be with a high pattern speed. The presence of a long bar has also been confirmed by photometric evidence by the recent work of Wegg & Gerhard (2013) and Wegg et al. (2015). Hence, we favour a pattern speed $\Omega_p \simeq 40\,\mathrm{km\,s^{-1}\,kpc^{-1}}$, in agreement with the determinations of Fux (1999), Weiner & Sellwood (1999) and Rodriguez-Fernandez & Combes (2008).

We should also mention that many other parameters influence fits to the observations. The bar angle ϕ and the sound speed are particularly relevant. Also, simulations are known to be affected by resolution effects (Chapter 3). We have neglected these parameters in this chapter in part because they have been studied in Chapter 3 as well as in other papers (e.g. Englmaier & Gerhard, 1997; Patsis & Athanassoula, 2000). Had we included them in our discussion, the parameter space to explore would have been intractably large. As an addition, we report that we have repeated our simulations with a higher sound speed of $c_{\rm s} = 20\,{\rm km\,s^{-1}}$ and we have found increasing the sound speed can reduce the number of spiral arms.

Another question of interest regards the importance of the gas self-gravity. To test this, we have carried a small number of simulations in the potential of Chapter 3 taking into account the gravitational potential generated by the gas, in addition to that of the stars that is assumed to be externally imposed as before. We have found that self-gravity is generally negligible and should be taken into account as a refinement only after the signatures of the bar described above have been better constrained. The results of an experiment that includes self-gravity is described in Appendix C.

5.7 Conclusion

The model presented in Chapter 3 had two major shortcomings. One was the lack of internal features, that could reproduce, for example, the 3 kpc arm, and the other was the lack of emission at forbidden velocities. We have run many simulations, varying systematically the quadrupole component of the bar potential, and we have found that both these shortcomings can be cured by adjusting the pattern speed and the quadrupole length and strength. However, we have failed to find a set of parameters that reproduces all the important observational features simultaneously. Good fits to individual features can be often obtained to the detriment of other features.

Our exploration of the quadrupole parameter space suggests that the pattern speed of the bar is around $\Omega_{\rm p} = 40\,{\rm km\,s^{-1}\,kpc^{-1}}$ and that the bar exponential scale length as defined in equation (5.3) must be at least $r_{\rm q} = 1.5\,{\rm kpc}$, while $r_{\rm q} = 1\,{\rm kpc}$ is too short. The bar strength, defined in the same equation, must be at least $A = 0.4$. In our study we haven't explored all the parameters that are important in fitting the Milky Way: in particular, we have kept the angle between the major axis of the bar and the Sun-Galactic Centre line constant at the value $\phi = 20^\circ$. Other parameters characterising the model, such as the potential's monopole component, and the sound speed, play an important role in determining the gas flow. The resulting parameter space is too big to be tractable with by-eye fitting methods. To obtain a model that can reproduce all the important features simultaneously, automatic quantitative methods to search in parameters space should be used. This will be the subject of the next chapter.

References

Athanassoula E., 1992, MNRAS, 259, 345
Binney J., Gerhard O. E., Stark A. A., Bally J., Uchida K. I., 1991, MNRAS, 252, 210

Binney J., Merrifield M., 1998, Galactic Astronomy. Princeton University Press
Bissantz N., Englmaier P., Gerhard O., 2003, MNRAS, 340, 949
Dame T. M., Thaddeus P., 2008, ApJ, 683, L143
Englmaier P., Gerhard O., 1997, MNRAS, 287, 57
Englmaier P., Gerhard O., 1999, MNRAS, 304, 512
Fux R., 1999, A & A, 345, 787
Gerhard O., Wegg C., 2014, in Lessons from the Local Group, eds. Freeman, K. C., Elmegreen, B. G., Block, D. L., & Woolway, M. (Springer; New York), in press (arXiv:1408.0219)
Jenkins A., Binney J., 1994, MNRAS, 270, 703
Patsis P. A., Athanassoula E., 2000, A & A, 358, 45
Rodriguez-Fernandez N. J., Combes F., 2008, A & A, 489, 115
Sellwood J. A., Wilkinson A., 1993, Reports on Progress in Physics, 56, 173
van Albada G. D., van Leer B., Roberts, Jr. W. W., 1982, A & A, 108, 76
Wegg C., Gerhard O., 2013, MNRAS, 435, 1874
Wegg C., Gerhard O., Portail M., 2015, MNRAS, 450, 4050
Weiner B. J., Sellwood J. A., 1999, ApJ, 524, 112

CHAPTER 6

Fitting Dynamical Models to the Milky Way

6.1 Introduction

Fitting the Milky Way kinematic data to dynamical models is not a trivial problem. This chapter is an attempt to address this issue. The difficulties were briefly mentioned in Sect. 1.3.3, but we also met them concretely in Chapter 5, where we found difficult to find a good model due to the high number of parameters involved in the problem.

As we have discussed in Chapter 2, (l, v) plots are the main tool to compare models with data. We have reviewed the data for the two most studied species, HI and ^{12}CO. However, many other species have also been studied, although less comprehensively, probing different temperature and density regimes of the gas, for example ^{13}CO, CS or more exotic species (Bally et al., 1987; Dahmen et al., 1997; Nakagawa et al., 1998; Sawada et al., 2001; Riquelme et al., 2010, and references therein).

We can produce synthetic (l, v) plots through gas dynamical models of the Milky Way, such as those mentioned in Sect. 1.2 or those described in Chapters 3 and 5. The problem now is: once we are given a synthetic (l, v) plot, how do we compare it to the available (l, v) plots to produce a number that says how "good" it is? If we can answer this question we may be able to unravel a yet unexploited amount of information from the spectral line data that can be used to constrain the Galactic potential.

An important limitation of all the works to date (e.g. Peters, 1975; Liszt & Burton, 1980; Mulder & Liem, 1986; Binney et al., 1991; Jenkins & Binney, 1994; Weiner & Sellwood, 1999; Englmaier & Gerhard, 1999; Fux, 1999; Bissantz et al., 2003; Rodriguez-Fernandez & Combes, 2008; Pettitt et al., 2014) is the lack of a fully satisfactory way of comparing synthetic models' predicted (l, v) distributions against the observed ones. The ballistic models tried only to superimpose the projected traces of orbits to ridges and edges identified by eye in the observed (l, v) distribution, without any consideration about the intensity produced by such traces. Authors who studied hydrodynamical simulations have mostly tried to take snapshots of

the gas density projected to the (l, v) plane and then qualitatively tried to identify by visual inspection counterparts of features found in observations. Some works (Fux, 1999) have given a very detailed and coherent interpretation of many features. Others have carried out a quantitative comparison based only the envelope of the (l, v) distribution (Weiner & Sellwood, 1999; Englmaier & Gerhard, 1999), ignoring the extra information contained in the internal structure of the data. More recently, Pettitt et al. (2014) have proposed a fit statistic akin to χ^2 that makes use of the full (l, v) distribution. Although this is in principle a correct approach which makes full use of the available data, as we are going to explain we believe that it is not well suited for the task at hand.

Suppose that we want to model the (l, v) CO data. Suppose we have a model in which all the gas dynamics is right, all the spiral arms are at the right positions, but we have a bad emissivity model (that relates density to how much radiation gets out) and a bad radiative transfer model (that relates how the intensity of radiation is modified during its travel toward us, for example by processes such as absorption, stimulated emission and scattering). Then, despite the bright ridges being at the right positions in the (l, v) plots, the exact (l, v) intensities that we produce will not be similar to CO observations. Thus a χ^2 fit will tell us that our model is poor, even though we have the right dynamics. As an extreme example of this situation, suppose that our emissivity and radiative transfer models were well suited for HI but not for CO. Our final synthetic (l, v) plot may be identical to the HI data. From the point of view of understanding dynamics, this is a success, because we have all the right spiral arms, and so on. But the χ^2 will tell us that this is a bad model for CO data.

As a somewhat different example, suppose now that our emissivity and radiative transfer models are good while the dynamics is *almost* right, but not quite. For example, the spiral arms projected to the (l, v) plane are in slightly different positions than in the data. To the eye, the fit might look quite good: you just need to move those features by a small amount, but they are all there. But again the χ^2 might tell us a different story: by comparing only contents of the same bins in data and model, it fails to see that a spiral arms is just slightly displaced. Thus, so many different ingredients concur in producing the final detailed (l, v) intensity at a single point, that is difficult to get all of them right at the same time, and failing in modelling accurately only one (just the dynamics or just the emissivity model, for example) of them can completely undermine the χ^2 fitting procedure.

Thus, the observational (l, v) plots of different species can be considered very similar or very different depending on the viewpoint. On the one hand, the χ^2 insists in telling us that their detailed intensities at each point can vary very much from species to species, and that a good model for HI is not a good model for CO. On the other hand, HI and CO data share many common characteristics, for example, the internal features of Sect. 2.3.2. We would like to produce a quantitative fitting that compares only the common part, which as we argue in this chapter is mostly dependent on the underlying dynamics rather than on the details of the density distribution and of the local physics on small scales.

That this must be possible can be explained using the following story, which also shows the key idea of our approach to the problem. Suppose that some advance in technology allows a new (l, v) plots to

be measured for our Galaxy in a hitherto unmeasured species, for example molecular hydrogen H_2. We show the new (l, v) plots to an experienced astronomer. It is very likely that the astronomer will recognise immediately that the data are an (l, v) plot for material in our Galaxy. How does he know? We can do an analogy between our problem and the way humans recognise human faces. Suppose that a friend sends us a picture portraying him, tanned by the Sun and lying on the beach. How do we know he is the same guy that we saw in his dark office the week before? He also has some beard now, that he did not use to have! Clearly our brain does not use χ^2 fitting to identify him among all the people we know: the different lighting conditions, the beard, the different environment and the different viewing angle would have fooled the χ^2 fitting. What we have recognised in our friend are some of his characteristic features: his long nose, his typical expression, the shape of his eyes, etc. Indeed, computer face-recognition algorithms do not use χ^2; instead, they work by matching features.

The idea of this chapter is to borrow methods from face-recognition and fingerprinting recognition algorithms to identify the "face" of our Galaxy. We present a first attempt to develop this idea, but it is probably susceptible of improvement and further development; however, it gives an indication of the direction that we think should be taken.

We argue that the problem of interpreting (l, v) plots is best split into two steps: one should first constrain the gross distribution of gas and the overall Galactic potential by fitting simple dynamical models to "features" in the observed (l, v) distribution; then, once this gross structure has been found, the model can be refined by including more detailed treatments of gas chemistry, physics on small scales, emissivity models, radiative transfer models and so on, for example along the lines proposed by Pettitt et al. (2014). We focus on the first step: the problem of fitting features. We present a new method that can be seen as an automated way of performing the task that has been previously done by visual inspection: comparing broad scale features in synthetic (l, v) plots against those in the observed (l, v) plots. In our method, features in the models are identified automatically by a computer, while features in observations are identified by the astronomer, but only once. The method returns a single number that measures the dissimilarity between a synthetic (l, v) plot and the observations. The approach takes inspiration from human face- and fingerprint-recognition algorithms.

The chapter is organised as follows. In Section 6.2 we explain how we use the observational data in our fitting method. In Section 6.3 we describe our simple hydrodynamical models and show how they reproduce the same kinds of features. Section 6.4 is the core of the chapter, where we describe in detail our method for defining a "distance" between observed and model features. Section 6.5 presents a range of tests on mock data to assess the performance of the method, including how it compares to other methods, such as envolope fitting or χ^2. In Section 6.6 we show an example of application to real data, before summing up in Sections 6.7 and 6.8.

6.2 How to use the observations

As we have seen in Chapter 2, the (l, v) distributions of HI and CO spectral lines contain an incredibly rich and diverse amount of information. They exhibit clumpiness and complicated structure on small scales, but also coherent, broad features on large scales.

In our fitting method the first step is to identify features in the observational data. We want features that depend the least on small scales physics (that controls how much light exactly is produced at each point) and on the details of radiative transfer (that controls how the light intensity is modified while travelling toward us) while depending the most on the large scale dynamics of the gas. The most obvious such feature is the envelope of the emission. We have discussed in detail how to identify this in the observational data in Sect. 2.3.1. The other type of feature are internal features of the type discussed in Sect. 2.3.2. These consist mostly of bright ridges, some of which are thought to correspond to spiral arms in the Galaxy.

Our fitting method requires that features in the observations (but not in the models) are identified by a human being and not by a computer. The advantage of this subjective identification is that it allows to include our own insight when identifying features; this can be tested by selectively omitting features or adding new ones. However, there clearly are also dangers in this subjective identification, i.e. biases. Unfortunately, the task of determining automatically features in the real data for the moment remains too hard (see also Section 6.7.2.1).

Following the analysis of Chapter 2, all the important features in their final format are represented in Fig. 6.1. This represents the observational input for our fitting method in the case of real data. It is clear that it represents a compromise between using the full information available in the data (because we have thrown away much information in going from the data to Fig. 6.1) but it uses more information than using only the envelope.

6.3 Methods

In this section we explain how we obtain synthetic (l, v) distributions given a model for the Galactic potential. We split the process into two steps:

1. Construct snapshots of the density and velocity distribution of the gas (i.e., obtain $\rho(\mathbf{x})$ and $\mathbf{v}(\mathbf{x})$ for each point in the Galaxy).

2. Project the density and velocity distribution onto (l, v) space

Sect 6.3.1 and Sect. 6.3.3 deal respectively with these two steps.

6.3.1 Hydro simulation scheme

Our methods are very similar to those described in Sect. 3.2. The main difference is that here the recycling law is disabled.

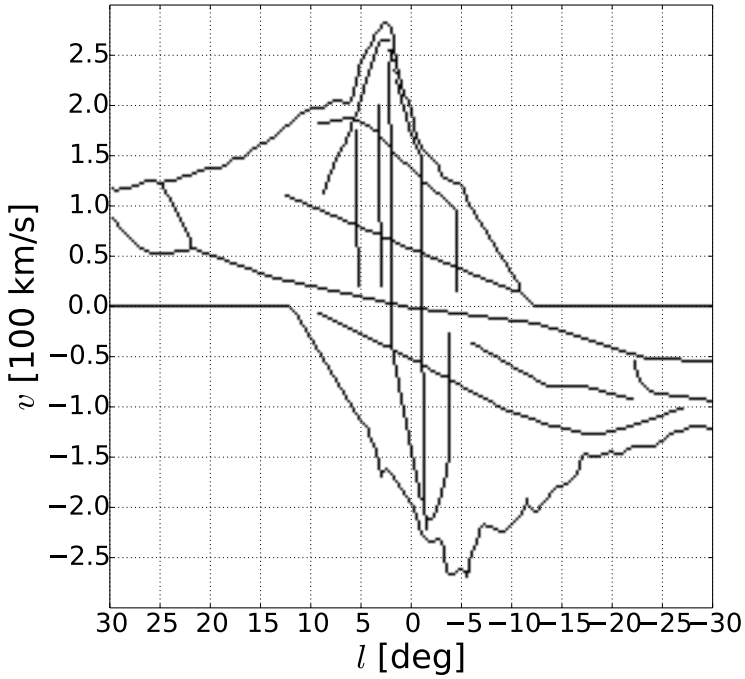


Figure 6.1: All the features used in comparison with models in the final format. This is the data input for the SMHD (defined in Section 6.4.2).

As before, we assume that the gas is a two-dimensional inviscid isothermal fluid and we use the same grid-based, Eulerian code based on the second-order flux-splitting scheme described in Sect. 3.2.

In this chapter, we use a sound speed of $c_{\rm s} = 10\ {\rm km\,s^{-1}}$. We used a 400×400 grid, with cells of side 50 pc. Thus the total simulated area is a square of 20 kpc side.

In each run the initial conditions are as follows. We start with an axisymmetrized bar and, to avoid transients, turn on the non-axisymmetric part of the potential gradually during the first 150 Myr, in such a way that the total mass of the Galaxy is conserved in the process. The gas starts with a uniform density profile and moves on circular orbits at the local circular velocity. For the models presented here, we verified that the gas reaches an approximate steady-state, although this is not a requirement for the application of our comparison method (section 6.4 below).

We have disabled the recycling law in this chapter. As a consequence, in the central region our simulations reach very high densities that should be taken as upper bounds on real densities.[1] We use outflow boundary conditions such that the gas can freely escape the simulated region, after which it is lost forever. The potential well is sufficiently deep that very little gas escapes from our simulation box.

6.3.2 Models

As in Chapters 3, 4 and 5 the gas is assumed to flow in an externally imposed barred potential. For much of this chapter we will use our reconstruction of the "standard" potential of Englmaier & Gerhard (1999) from the multipole moments they plot in their Figure 3; our reconstruction reproduces their rotation curve and

[1] To test the importance of this we have run some models that do include a simple gas recycling law. The central spike vanishes in these models, but the rest of the features are unchanged.

effective potential (their Figures 4 and 5, respectively) correctly. To allow comparison with their results, in Fig. 6.2, panel (a), we show the gas density obtained by our simulations in this reconstructed potential for the bar pattern speed of $\Omega_{\rm p} = 55$ km s^{-1} kpc^{-1} that they assumed in their Fig. 9: there is a very good match between the density produced by our Eulerian simulations and that of their SPH scheme. Panels (b) and (c) of Fig. 6.2 show our reconstruction of two other models from the literature, Bissantz et al. (2003) and Rodriguez-Fernandez & Combes (2008). We discuss these models in Section 6.6.

6.3.3 Calculating model (l, v) distributions

Producing accurate synthetic (l, v) maps for a given spectral line would require us to proceed as follows. First, we should find the emissivity and opacity at each point in our simulated galaxy. These would be determined by variables such as the temperature $T(\mathbf{x})$, the density $\rho(\mathbf{x})$ and the chemical composition at each point.[2] Second, we should perform radiative transfer ray-tracing, following the intensity of rays from point to point until they reach our location in the Galaxy, where the intensity of the rays can be converted to brightness temperature to give us our synthetic (l, v) map.

In the simple models used in this thesis, we do not include ISM heating and cooling processes such as heating from cosmic rays, heating due to stellar winds, photoelectric heating, or cooling from atomic lines. Hence, we do not keep track of the temperature of the ISM and of its chemical composition. We also do not have a real radiative transfer model and we use instead a much simpler projection law, similar to what has been used in Chapter 3 and 5, which assumes that the gas is optically thin. However, we do include an extra parameter α which allows for a crude exploration of varying the emissivity law. This is described below.

The projections are performed as follows. We adopt a very simple binning procedure to produce the predicted (l, v) distributions corresponding to each simulation snapshot $(\rho(\mathbf{x}), \mathbf{v}(\mathbf{x}))$. We use an (l, v) grid with spacing $\Delta l = 0.25°$, $\Delta v = 2.5\,$km s^{-1}, whereas the simulations use a 50 pc $\times$ 50 pc grid.

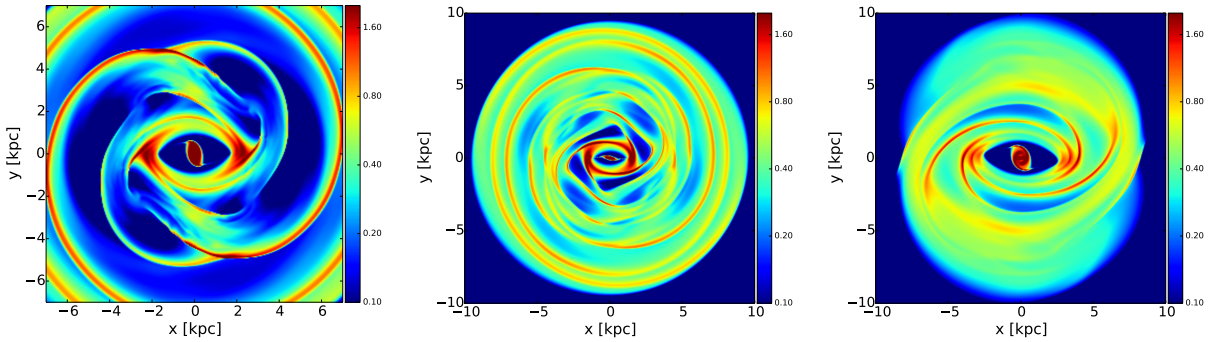
First, we use linear interpolation to resample the simulation's $\rho(\mathbf{x})$ and $\mathbf{v}(\mathbf{x})$ grids down to finer 6.25 pc grids. Then, given an assumed position and velocity of the Sun, we calculate the Galactic longitude l_i and line-of-sight velocity v_i corresponding to each resampled grid point and bin the resulting (l_i, v_i) onto the (l, v) grid with weight

$$w_i = \frac{\rho_i^\alpha}{s_i^2 + (2\text{ kpc})^2}, \tag{6.1}$$

where s_i is the distance of the grid point from the Sun. The 2 kpc softening term is used to avoid sampling artefacts from grid points that lie close to the Sun's position.

We include the exponent α in Eq. (6.1) as a very crude proxy for effects of changing the emissivity law, which allows us to test how sensitive features in our synthetic (l, v) maps are to details of the local physics. Fig. 6.3 shows projections made for different values of α. We usually adopt $\alpha = 1$, in which case

[2] Under some simplifying assumptions such as local thermodynamic equilibrium, the chemical composition can be calculated given T and ρ.



(a) (b) (c)

Figure 6.2: Gas densities produced by our simulation scheme for some models from the literature. In all panels the x-axis is coincident with the major axis of the bar. Density is in arbitrary units, and the initial density is uniform with value 1. (a) Englmaier & Gerhard's (1999) standard potential with $\Omega_\mathrm{p} = 55\ \mathrm{km\,s^{-1}\,kpc^{-1}}$ taken at evolutionary time $t = 367\,\mathrm{Myr}$. Compare with Fig. 9 of their paper. (b) Bissantz et al.'s (2003) standard potential at evolutionary time $t = 310\,\mathrm{Myr}$. The bar pattern speed is $\Omega_\mathrm{p} = 58.6\ \mathrm{km\,s^{-1}}$. A spiral component with pattern speed $\Omega_\mathrm{spiral} = 19.6\ \mathrm{km\,s^{-1}\,kpc^{-1}}$ is also included in the potential. Compare with their Fig. 12. (c) The potential of Rodriguez-Fernandez & Combes (2008) with $\Omega_\mathrm{p} = 30\ \mathrm{km\,s^{-1}\,kpc^{-1}}$ at evolutionary time $t = 367\,\mathrm{Myr}$. Compare with their Fig. 8, second row.

our projection is linear in the density and gives results equivalent to some simple cases such as emission by optically thin atomic hydrogen at constant temperature (see Chapter 3).

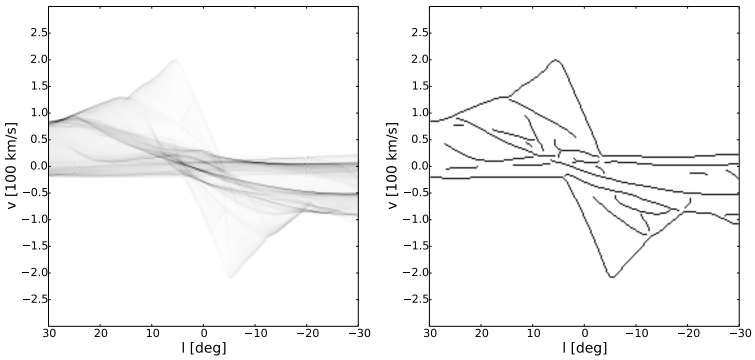
Throughout this chapter, we assume that the Sun is undergoing circular motion at a radius $R_0 = 8$ kpc with speed $v_\odot = 220\ \mathrm{km\,s^{-1}}$. The major axis of the bar always lies along the x axis in our models. Calling ϕ the angle between the major axis and the Sun–GC line, the cartesian coordinates of the Sun are given by $x_\odot = R_0 \cos\phi$, $y_\odot = R_0 \sin\phi$.

A number of further minor comments are in order regarding this projection. As we do not have a recycling law that lowers the gas in high-density regions, gas accumulates at the very center reaching implausibly high densities. So, before projecting the density, we clip all gas with density more than four times higher than the initial density to this maximum value. This has no major effect on the identification of features because, as we will see below (Section 6.4), the velocity structure is much more important than the density structure in determining the strength of the projected (l, v) features. We also project only the gas inside $R < 8$ kpc: this is justified by the fact that we focus only on the central disk ($|l| < 30°$) and exclude low-velocity emission from our comparisons. Material outside the Solar circle would produce emission only at high longitudes or low line-of-sight velocity, which we do not include in our fits.

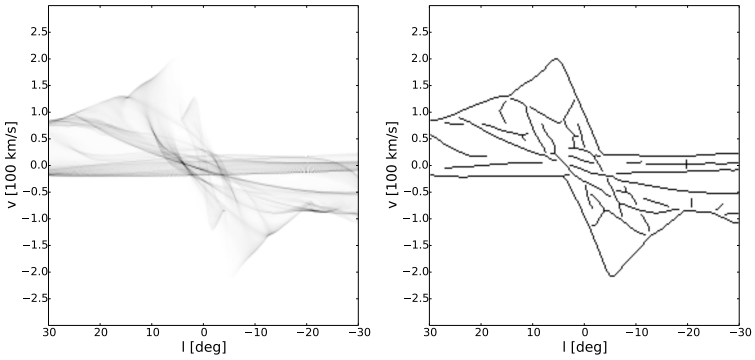
6.4 Comparing models and observations

Given a set of synthetic (l, v) distributions, we would like to judge which one is "closest" to the observations. This is actually a very challenging task, and the main focus of this chapter.

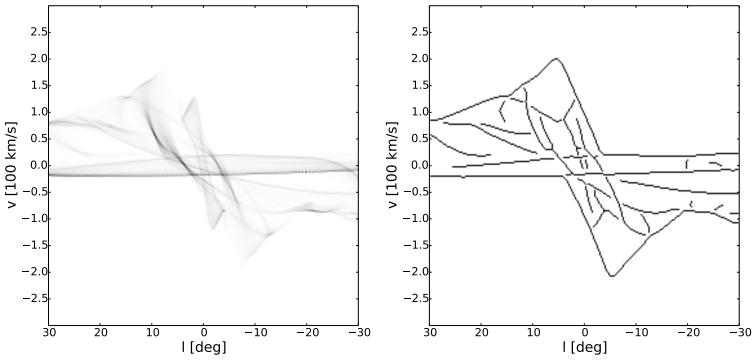
On small scales, the detailed intensities at each point are the result of complicated micro-physics, and therefore depend on local density, temperature, and chemical composition of the gas. Also the local clumpiness of the gas plays a major role in determining the observed longitude-velocity intensities on small scales. On



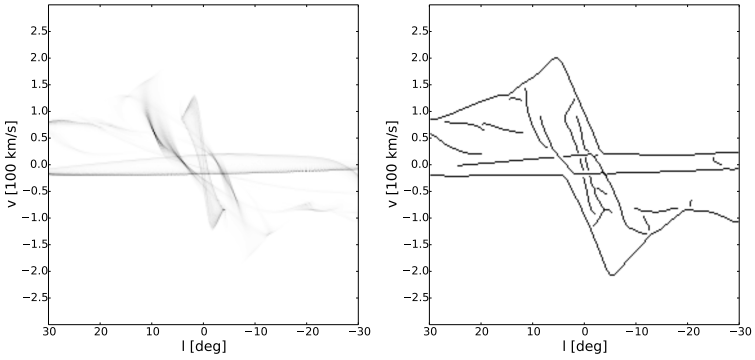
(a) $\alpha = 0$



(b) $\alpha = 1$



(c) $\alpha = 2$



(d) $\alpha = 3$

Figure 6.3: The effect of the exponent α (the emissivity law) adopted in Eq. (6.1). The panels on the left show how the model's predicted $T_{\rm B}(l, v)$ varies with α. The underlying model is our reconstruction of Englmaier & Gerhard's (1999) potential with bar pattern speed $\Omega_{\rm p} = 55\ {\rm km\,s^{-1}\,kpc^{-1}}$ viewed at time $t = 367\,{\rm Myr}$ with angle $\phi = 20°$. The panels on the right show the corresponding features found using the algorithm of Section 6.4.1. For $\alpha = 0$, the gas density in the (x, y) plane is effectively uniform, and all information about density variations is washed away. Yet, many features in the (l, v) plot remain visible, particularly those corresponding to spiral arms. This suggests that the velocity field is more important than the density distribution in producing (l, v) features (see also Mulder & Liem 1986).

the other hand, the presence of large-scale features, such as bright ridges tracing spiral arms, is more robust to changes in the micro-physics, and more sensitive to changes in the large-scale dynamics of the gas. The evidence for this can be seen both in data and models:

- The main features believed to trace spiral arms are present both in CO and HI. The CO and HI features are coincident with one another on the scale of the (l, v) maps of Figure 2.3. The same cannot be said for the detailed intensities, as features have different relative intensities in the two species. This is to be expected, since the two species probe different density-temperature regimes; therefore, spatial densities are not completely correlated.

- The data also show a clumpy sub-structure that can be modeled as the result of heating and cooling processes, such as stellar feedback and supernova feedback; interestingly, if a smooth simulation is run turning off these processes, the main large-scale features stay the same (Baba et al., 2010).

- In Fig. 6.3 we see the result of projecting the gas density for four different values of α. Clearly, the same bright ridges can be identified at the same locations for a wide range of values of α, while the relative intensity of these ridges varies. Only at extreme values ($\alpha = 3$) does the difference start to be noticeable.[3] Particularly striking is the case $\alpha = 0$, where any information about density variations (for example along spiral arms) has been washed out: the features stay more or less the same as long as *some* gas is present at all points, suggesting that the large-scale velocity field is what mostly determines the features (see also Mulder & Liem, 1986).

- Pettitt et al. (2014) produced a longitude-velocity diagram from a full radiative transfer calculation for CO, and compared it to the diagram obtained by a simpler approach akin to ours. Their results show that while detailed intensities are different, the overall morphology and large-scale features are identical in the two versions.

Therefore we argue that, loosely speaking, detailed intensities at each point probe the local physics and the details of the emissivity and radiative transfer models, while broad features trace the large-scale dynamics of the gas.

Proper modelling of local physics, radiative transfer *and* gas dynamics at the same time is certainly possible (e.g., Shetty & Ostriker, 2008; Dobbs et al., 2011; Tasker, 2011; Pettitt et al., 2014), but its computational expense renders it unattractive when the parameter space to explore is large (see, e.g., Section 2.3.2 of Pettitt et al., 2014). Nevertheless, the above considerations suggest that it is possible to break the problem of understanding the Galaxy's ISM into two steps: first constrain the large-scale gravitational potential and dynamics by matching only the broad features in the (l, v) distribution; then, once the potential has been constrained, go back and use more sophisticated models to understand the internal structure and chemistry of the ISM in detail.

[3] Nevertheless, our fitting scheme method still finds the correct parameters in this case; see Sect. 6.5.

Here we present a quantitative comparison scheme to use in the first step, focusing only on features that are believed to depend the most on gas dynamical physics and the least on local small-scale physics, the emissivity model and the radiative transfer model. This "feature comparison" problem has much in common with the problem of matching fingerprints, in which, given a smudged, incomplete, contaminated print taken from a crime scene, the goal is to find which members in a database of prints look most like it. Here the Milky Way is the crime scene and we use a set of simple hydrodynamical models as the database of potential culprits.

6.4.1 Identifying features in model (l, v) distributions

As discussed above, features that mostly probe the large-scale dynamics of the gas, besides the envelope, are generally bright ridges in the longitude-velocity plane. Therefore our first step is to identify such ridges: given a binned model (l, v) distribution (Section 6.3.3 above), the task is to return a corresponding binary image in which each bin is 1 if the emission at that bin is part of a feature, 0 otherwise. Features are lines 1-pixel wide and at least 5 pixels long. Our procedure relies heavily on the widely used edge-detection algorithm of Canny (1986), modified to detect ridges instead of edges. The steps of our procedure are the following, each illustrated in Fig. 6.4.

- **Envelope enhancement** In a typical model, some parts of the envelope are brighter than others. Since we want our algorithm to pick up the whole envelope, not only its brightest parts, we increase the intensity of pixels lying on the envelope. Thus we put the value of all pixels on the envelope equal to the value of the brightest pixel in the image. A pixel is defined to belong to the envelope if it has at least one side in common with an empty pixel which is at higher velocity than the highest velocity pixel with positive emission or at lower velocity than the lowest velocity pixel with positive emission.

- **Smoothing** We convolve the image (including the enhanced envelope) with a Gaussian to remove noise. We used a standard deviation $\sigma = 2$ pixels. (Recall that a pixel size is $\Delta l = 0.25^\circ$ in longitude and $\Delta v = 2.5\ \mathrm{km\,s^{-1}}$ in velocity.)

- **Ridge filter** This step filters the image in order to highlight ridges. Let $F(x, y)$ denote a two-dimensional function and $H(x, y)$ the matrix of its second-order partial derivatives (the Hessian matrix):

$$H(x, y) = \begin{bmatrix} F_{xx} & F_{xy} \\ F_{xy} & F_{yy} \end{bmatrix}.$$

  A measure of the presence of a ridge at a point is the value of the main negative eigenvalue of H (Lindeberg, 1996). The direction of the ridge is orthogonal to the direction of the eigenvector associated with this main negative eigenvalue. We therefore compute the lowest eigenvalue λ_{low} of the Hessian matrix and its associated direction p. The direction of p is orthogonal to the ridge. The greater the value of $R \equiv -\lambda_{\mathrm{low}}$, the stronger the ridge. To compute the derivatives F_{xx}, F_{xy} and F_{yy} we use Sobel

operators (Sobel & Feldman, 1968) with a kernel size of 3 pixels.

- **Non-maximum suppression** A search is carried out to determine whether the ridge strength R assumes a local maximum in the direction p orthogonal to the ridge. At every pixel, we round the p direction to the nearest 45° (that is, the rounded direction points to one of the 8 nearest neighbours). Then we compare the ridge strength at the current pixel with the ridge strength of the pixels in the positive and negative p direction. If R at the current pixel is greater than or equal to both, we mark the point as a possible ridge, otherwise we suppress it, i.e., we declare that no ridge goes through this point.

- **Hysteresis thresholding** We need to decide which points that are left unsuppressed by the previous step are actual ridges. Large values of R are more likely to correspond to ridges. It is in many cases difficult to specify a unique threshold at which points switch from corresponding to ridges to not doing so. For this reason, following Canny (1986), we use thresholding with hysteresis. Thresholding with hysteresis requires two thresholds, high and low. All points above the high threshold are marked as certainly ridges. All points below the lower threshold are marked as certainly non-ridges. The points between the two are marked as ridges only if they are connected to a point above the high threshold. This allows to follow a fainter section of a strong ridge, that is likely to be a genuine ridge. We use the mean value of R over the whole image as the high threshold and one half of this value for the low threshold. We have experimented with varying these by a factor of two (but keeping the same high-to-low ratio) and find little change in the resulting feature maps.

- **Thinning** We make each detected line thinner, reducing its width to one-pixel. This is done according to the algorithm of Zhang & Suen (1984), which preserves end points and pixel connectivity.

- **Remove small components** As a final polishing, we remove features that are too small and likely to be noise. We therefore remove from the image all the connected components that are less or equal than 4 pixels to obtain our final result.

6.4.2 Comparing the features

Having two binary images A and B representing the features of model and data respectively, we need to produce a single number that quantifies their dissimilarity. For this purpose, we have tried different options. The one we found intuitively most appealing involved applying the "Earth-mover distance" (EMD, see Appendix E). Surprisingly, in practical tests this did not perform as well as a much simpler alternative, the Modified Hausdorff Distance[4] (Dubuisson & Jain, 1994).

Call $a = \{a_1, ..., a_N\}$ the set of pixels containing 1 in the first image and $b = \{b_1, ..., b_M\}$ the same for

[4]However, as we discuss in more detail in Appendix E, we believe that EMD used in a qualitatively different way could prove to be a useful alternative to χ^2.

the second image. Then the Modified Hausdorff Distance (MHD) is defined as

$$\mathrm{MHD}(a,b) \equiv \sum_i \min_j \left(d(a_i,b_j)\right) \,, \tag{6.2}$$

where $d(a_i,b_j)$ is a suitable metric between pixels. In words, for each positive pixel in the first image, find the distance from the closest positive pixel in the second image, according to the chosen metric d. Then the MHD is the sum of these distances over all positive pixels in the first image. An unattractive feature of the MHD is that it is not symmetric in its arguments. We define a symmetrized version as

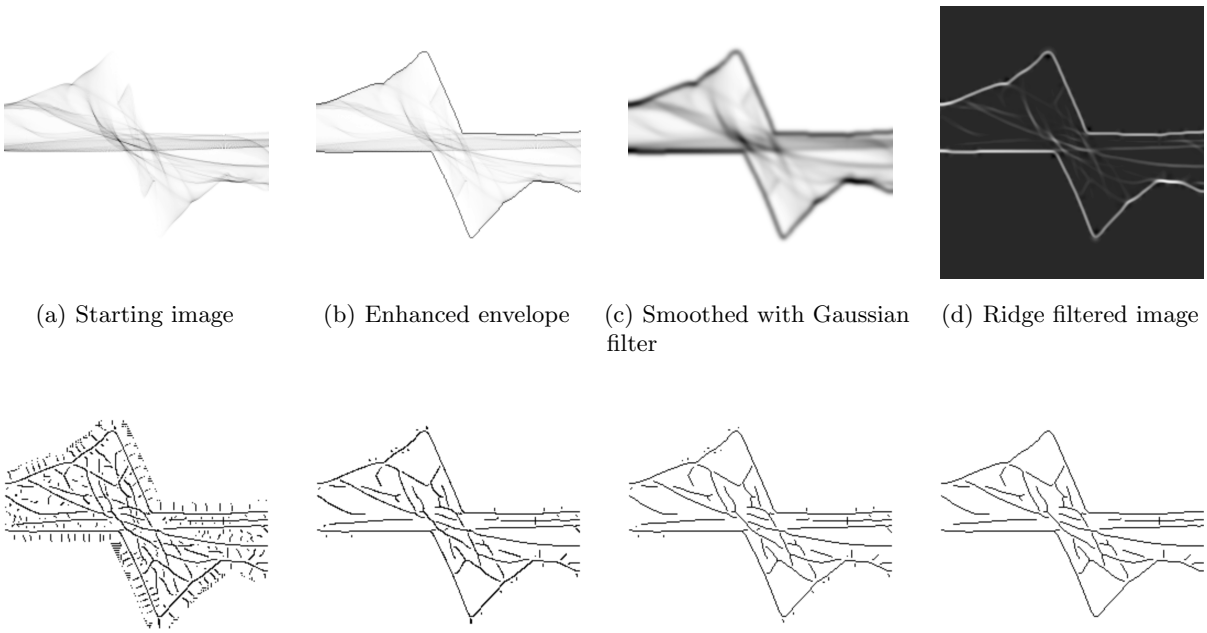
$$\mathrm{SMHD}(a,b) \equiv \frac{1}{2N}\mathrm{MHD}(a,b) + \frac{1}{2M}\mathrm{MHD}(b,a)\,. \tag{6.3}$$

Our recipe is that the dissimilarity between two binary images is their SMHD. The only thing left to decide is $d(a,b)$, the metric in the pixelated (l,v) plane. After some experimentation, we decided to use the city-block distance,

$$d(a,b) = \frac{|l_a - l_b|}{\Delta l} + \frac{|v_a - v_b|}{\Delta v}, \tag{6.4}$$

where (l_a, v_a) are the coordinates of pixel a in the first image, and, similarly, (l_b, v_b) are the coordinates of pixel b in the second. The city-block distance between two points is the sum of the absolute differences of their Cartesian coordinates. We found very little difference between this and the "Euclidean" distance obtained by squaring each of the terms in (6.4). The results of the fitting method depend also on the choice of the ratio $\zeta = \Delta v/\Delta l$, which should be adapted to the nature of features under consideration. A natural choice is to take ζ equal to the ratio of the typical velocity extension and the typical longitude extension of a feature, which in our case is approximately $\zeta = 10$. This value could in principle be adjusted to better suit other situations.

One might ask whether we need to symmetrize the MHD: we extract features from data and models in different ways, so why should we impose symmetry on the function we use to compare the resulting sets of features? In fact, variants of our method can be defined that do not symmetrize the MHD. Let D be the features in data and M the features in models. If we measure the dissimilarity of data and observations using $\mathrm{MHD}(M,D)$, then models are penalized if a model feature is not present in the data, but not viceversa. Thus, $\mathrm{MHD}(M,D)$ is insensitive to contaminants in the data that cannot be reproduced by the model. On the other hand if we use $\mathrm{MHD}(D,M)$, the opposite would be true: we are not penalized at all if our model predicts extra features in addition to those present in the data. An example of such a situation occurs in Figure 6.7, panel (b), below. We decided to use the symmetrized version as giving a good compromise, but one could choose to calculate both and process independently the separate bits of information acquired.



(a) Starting image (b) Enhanced envelope (c) Smoothed with Gaussian filter (d) Ridge filtered image

(e) Points after non-maximal suppression (f) Points after hysteresis thresholding (g) Points after thinning (h) Final output

Figure 6.4: The steps of our feature-finding algorithm (section 6.4.1), ordered from top left to bottom right. The final output represents the features detected in the model, constituted by the bright ridges and the envelope. Features are 1-pixel wide lines at least 5 pixels long. The final output is then used as input for the SMHD (defined in Section 6.4.2).

6.5 Tests with mock data

In this Section we use mock data generated from a variety of simulated galaxies to asses the performance of our method. By fitting the mock data with a family of parametrized simulated galaxies, we test how well we can recover the correct parameters describing the potential underlying the mock data. We also compare the performance of our SMHD distance against two other measures of goodness-of-fit that have been used previously, namely χ^2 and envelope distance.

Unlike our SMHD, which measures distance between model and observed features, χ^2 is a direct measure of the difference between model and observed brightness temperatures. It is defined as

$$\chi^2 \equiv \sum_n \left[\frac{T_{\rm B}^D(l_n, v_n) - T_{\rm B}^M(l_n, v_n)}{\Delta T_{\rm B}^D(l_n, v_n)} \right]^2 , \tag{6.5}$$

where $T_{\rm B}^M(l_n, v_n)$ is the model's prediction for the brightness temperature at the point (l_n, v_n) and $T_{\rm B}^D(l_n, v_n)$ is the corresponding "measurement" from the simulated dataset, with measurement uncertainty $\Delta T_{\rm B}^D(l_n, v_n)$. For the tests here we take $T_{\rm B}$ to be directly proportional to the binned (l, v) distribution constructed in section 6.3.3, with $\Delta T_{\rm B} =$ constant.

The envelope distance (ED) is defined as

$$D_{\mathrm{e}}^2 = \frac{1}{N}\left[\sum_{n=1}^{N}\left[v_+^D(l_n) - v_+^M(l_n)\right]^2 + \sum_{n=1}^{N}\left[v_-^D(l_n) - v_-^M(l_n)\right]^2\right] \tag{6.6}$$

where $v_\pm^D(l)$ and $v_\pm^M(l)$ are the positive- and negative-velocity envelopes of the "data" and "model" respectively. We consider the range $-6 \leqslant l \leqslant 30\,^\circ$ for the positive-velocity envelope and $-30 \leqslant l \leqslant 6\,^\circ$ for the negative; we omit portions of the envelope that in the real observations are seen to be heavily influenced by material outside the Solar circle. The envelope distance is closer in spirit to our SMHD than χ^2, as it involves measuring the distance between the terminal velocity features of model and data.

For both χ^2 and SMHD we exclude all data/features at low velocities, $|v| < 40\,\mathrm{km\,s^{-1}}$. This is done to simulate the modelling of real data, in which the low velocity features are dominated by foreground emission.

The models we fit in this section are all based on our reconstruction of the Englmaier & Gerhard (1999) potential. They have only three free parameters: the pattern speed Ω_{p}, the angle ϕ between the major axis of the bar and the Sun-Galactic Centre line and the evolutionary time t. We shall see in Section 6.5.1 below that, as these particular models settle into an approximate steady state, we can eliminate the time t, reducing the number of interesting parameters to be fit to just the pair $(\phi, \Omega_{\mathrm{p}})$. For each model, we evolve an initial axisymmetric state, as described in Section 6.3.1 and project onto the (l, v) plane assuming an exponent $\alpha = 1$ in equation (6.1). At this point we can calculate χ^2 and the ED (6.6). For the SMHD (6.3), we use the algorithm of Section 6.4.1 to find the features.

6.5.1 On stationarity

The Englmaier & Gerhard (1999) models we consider here quickly reach a steady state. Fig. 6.5 shows two snapshots of the model having pattern speed $\Omega_{\mathrm{p}} = 55\ \mathrm{km\,s^{-1}\,kpc^{-1}}$ viewed with $\phi = 20^\circ$ taken at two different evolutionary times. The features are almost identical, particularly outside the $|v| < 40\,\mathrm{km\,s^{-1}}$ band we exclude from our fitting procedure.

To quantify this, we calculate the SMHD and ED between a synthetic (l, v) plot at a fiducial evolutionary time, $t = 367\,\mathrm{Myr}$, and synthetic (l, v) plots at *different* evolutionary times in the *same* simulation. Thus all synthetic (l, v) plots considered in this section come from exactly the same potential, pattern speed Ω_{p} and viewing angle ϕ; only the time t is allowed to vary. Fig. 6.6 shows the SMHD and ED between the fiducial snapshot at $t = 367\,\mathrm{Myr}$ and all other snapshots in the same simulation. At $t = 0$ the gas moves on purely circular orbits and the SMHD and ED are both very large. As the bar is gradually turned on over the first 150 Myr of the run, the gas settles into a distribution that approaches that of the fiducial snapshot at $t = 367\,\mathrm{Myr}$. While this is happening, the SMHD and ED both decrease from their high initial values to much lower values, reaching an approximate plateau starting from 300 Myr. Of course, at the special point $t = 367\,\mathrm{Myr}$ both distances go to zero, but the values of the SMHD and ED in the plateau region away from this point provide lower bounds on the level of noise expected when we fit only relaxed, evolved versions of our models to mock data. Indeed, we will see below that varying one of the other parameters

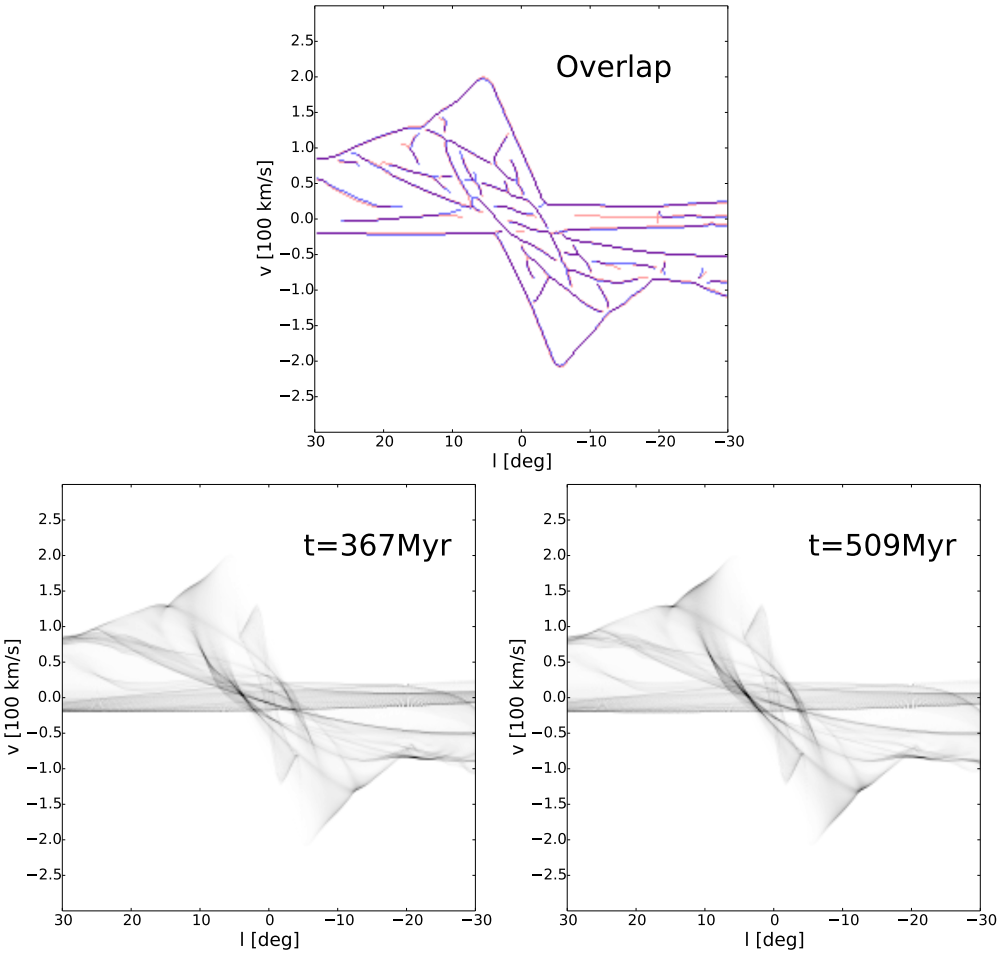


Figure 6.5: (l, v) plots of the simulation with $\Omega_\mathrm{p} = 55\,\mathrm{km\,s^{-1}\,kpc^{-1}}$ viewed from $\phi = 20°$ at different evolutionary times. The top panel overlays the features extracted from the two snapshots.

$(\phi, \Omega_\mathrm{p})$ produces differences in SMHD and ED much greater than the value of the plateau. An analogous test shows that the χ^2 value also reaches a plateau, but the curve leading to the final asymptotic value is not gently decreasing from $t = 0$ as for SMHD and ED, but instead it decreases more suddenly when the gas reaches a configuration close to the final steady state.

We emphasize that, although this shows that the particular models we consider here do reach an effective steady state, the feature-fitting algorithm and SMHD distance of Section 6.4 do not require that such a steady state be reached. In the following subsections we make use of this effective steady state and compare mock data only to model snapshots taken at some time $t > 300$ Myr long enough to allow the model to settle into a steady state; when fitting more general models with more than one pattern speed (e.g., Fux, 1999) one would have to include t as a parameter to be fit.

6.5.2 Recovering the correct model parameters

Having eliminated time t, we now turn to the more interesting question of how reliably one can identify the model parameters Ω_p and ϕ. We construct three different mock datasets, shown in Figure 6.7. All are generated from a fiducial model having pattern speed $\Omega_\mathrm{p} = 55\,\mathrm{km\,s^{-1}Myr^{-1}}$ viewed at $t = 310$ Myr and angle $\phi = 20°$, but differ in how the mock data are generated from that model. The first is constructed by

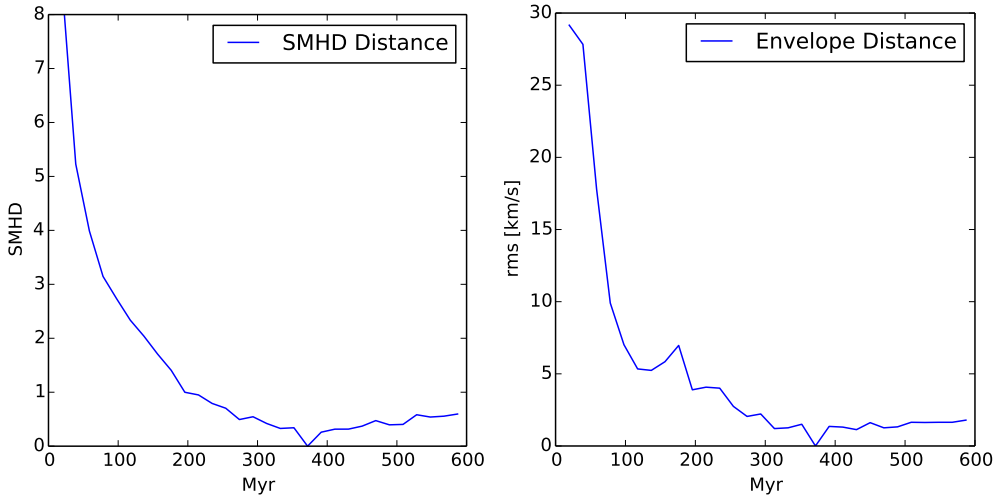


Figure 6.6: Variation of SMHD distance (left) and envelope distance (right) versus time for models with $\Omega_{\rm p} = 55$ km s^{-1} kpc^{-1} and $\phi = 20°$. The mock "data" used are the snapshots taken at time $t = 367$ Myr. SMHD is calculated using all features, including the envelope.

projecting the fiducial model assuming $\alpha = 1$ and using the algorithm of Section 6.4.1 to extract features. The second is identical to the first, but projected assuming $\alpha = 3$. The third is a modification of the first to which additional contaminating features have been added by hand.

We construct a set of models having $\Omega_{\rm p}$ in the range 20-70 km s^{-1} kpc^{-1} in steps of 2 km s^{-1} kpc^{-1} and ϕ in the range 0-60° in steps of 2°. For each model we project a single snapshot taken at $t = 370$ Myr assuming $\alpha = 1$. This set of models is then compared in turn to each of the three mock data sets. Notice that none of our models uses the "correct" values of $t = 310$ Myr and $\Omega_{\rm p} = 55$ km s^{-1}Myr^{-1} from which the mock data are generated.

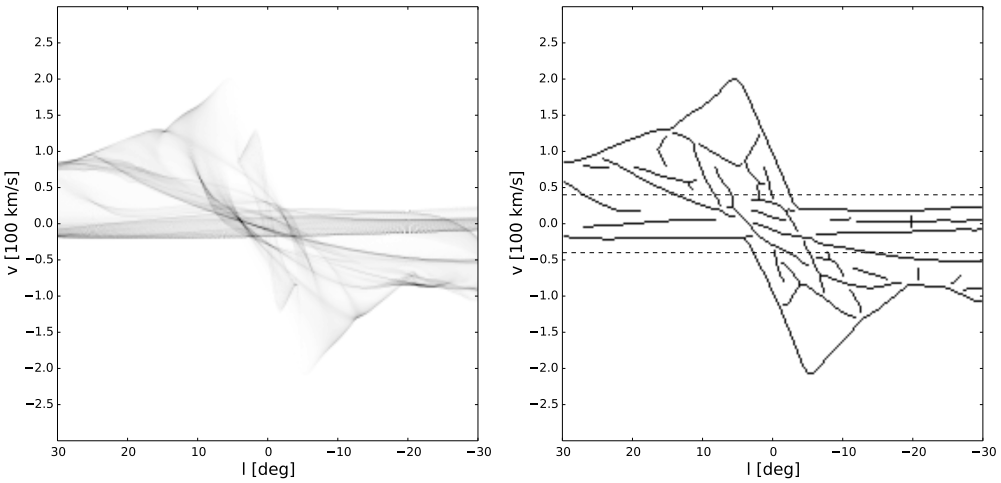
6.5.2.1 The basic test

The most basic test is given by checking how reliably the model parameters $(\phi, \Omega_{\rm p})$ are recovered when the models adopt the correct emissivity law and the data are uncontaminated by misidentified features. The mock dataset used for this test is shown in Fig. 6.7(a). Figure 6.8 shows by how much models with assumed parameters $(\phi, \Omega_{\rm p})$ differ from this mock dataset, as measured by SMHD, ED, and χ^2.

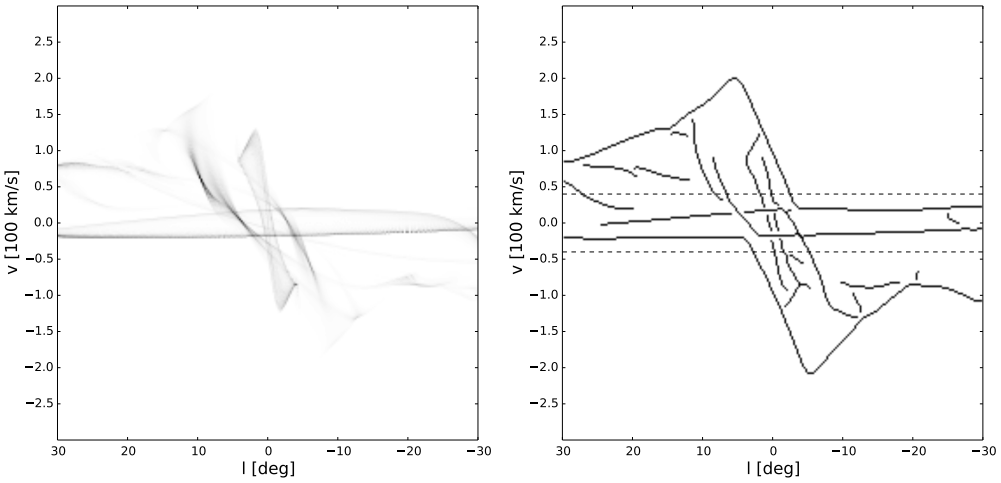
The plots demonstrate that when the emissivity law is correct (which is our simple proxy for how correct our treatment of local physics is), then all three distances can be used to locate the correct model; apart from some noisiness at low values of $\Omega_{\rm p}$, they all descend smoothly to a minimum at (or very close to) the correct values of $(\phi, \Omega_{\rm p})$. Therefore if we start away from the correct model, e.g., with a pattern speed wrong by 20 km s^{-1} kpc^{-1}, then all three distances indicate in which direction we should move to get to the right model.

One point to note from these plots is that knowledge of the envelope alone suffices to identify the correct model among the restricted family of models we consider here. This is not true in general, a point to which we return in Section 6.5.3.

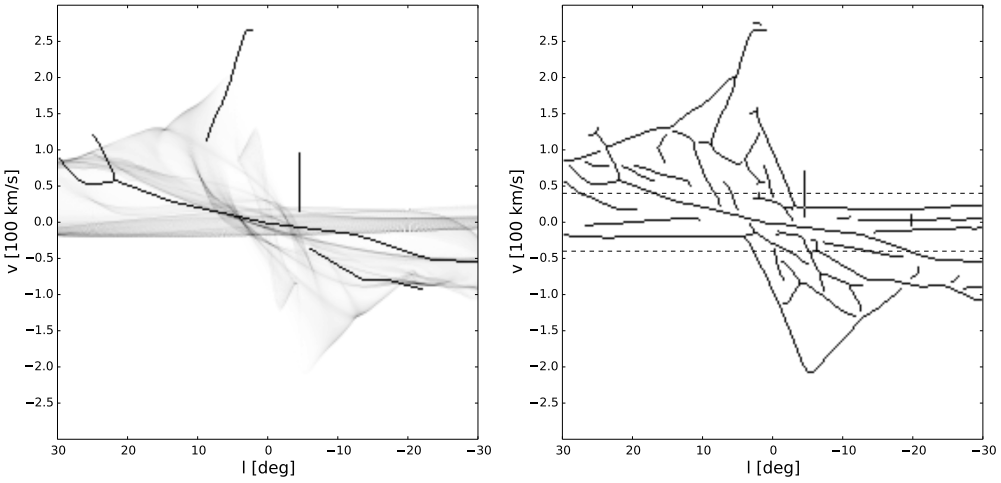
Another point is that the minimum in χ^2 is much sharper than the minimum in either the SMHD or ED.



(a) Mock data for the **basic test**, generated using $\alpha = 1$.



(b) Mock data for the **alpha test**, generated using $\alpha = 3$.



(c) Mock data for the **contamination test**, using $\alpha = 1$ and adding contaminating features.

Figure 6.7: The mock data used in the tests of Section 6.5.2. All data are based on our reconstruction of Englmaier & Gerhard's (1999) potential with $\Omega_\mathrm{p} = 55\ \mathrm{km\,s^{-1}\,kpc^{-1}}$ and $\phi = 20^\circ$ taken at $t \simeq 310\,\mathrm{Myr}$. The top row shows the results of projecting this model assuming $\alpha = 1$ in Eq. (6.1). The panel on the left shows the resulting (l, v) distribution, while the panel on the right shows the full set of features extracted using the algorithm of Section 6.4.1. The middle row shows the second dataset, projected assuming $\alpha = 3$. The bottom row shows the third dataset, projected assuming $\alpha = 1$, but to which the additional contaminating features indicated in the panel on the left have been added.

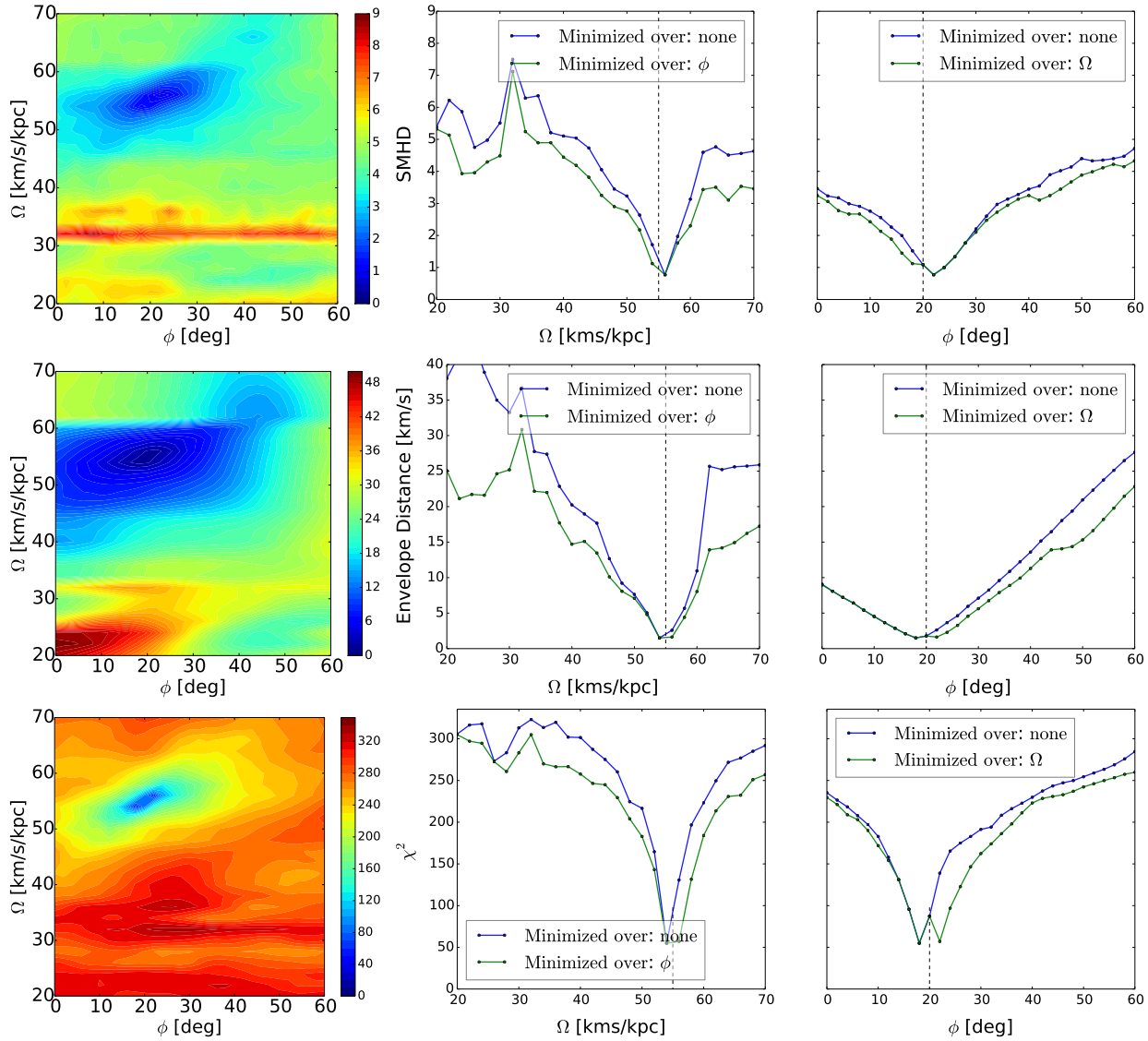


Figure 6.8: Variation of SMHD (top), ED (middle) and χ^2 (bottom) for the mock dataset constructed using $\alpha = 1$ in Section 6.5.2.1. The first column of each row shows how the corresponding distance varies as a function of the assumed (ϕ, Ω_p). The other two columns plot one-dimensional slices that pass through the location of minimum value of the distance. For the blue curves in the middle (right) column the slices are vertical (horizontal). The green curves are constructed by taking the distance of the best-fitting ϕ as a function of Ω_p (middle column) and the distance of the the best-fitting Ω_p as a function of ϕ (right column). The vertical lines plot the values of the parameters used to construct the mock data.

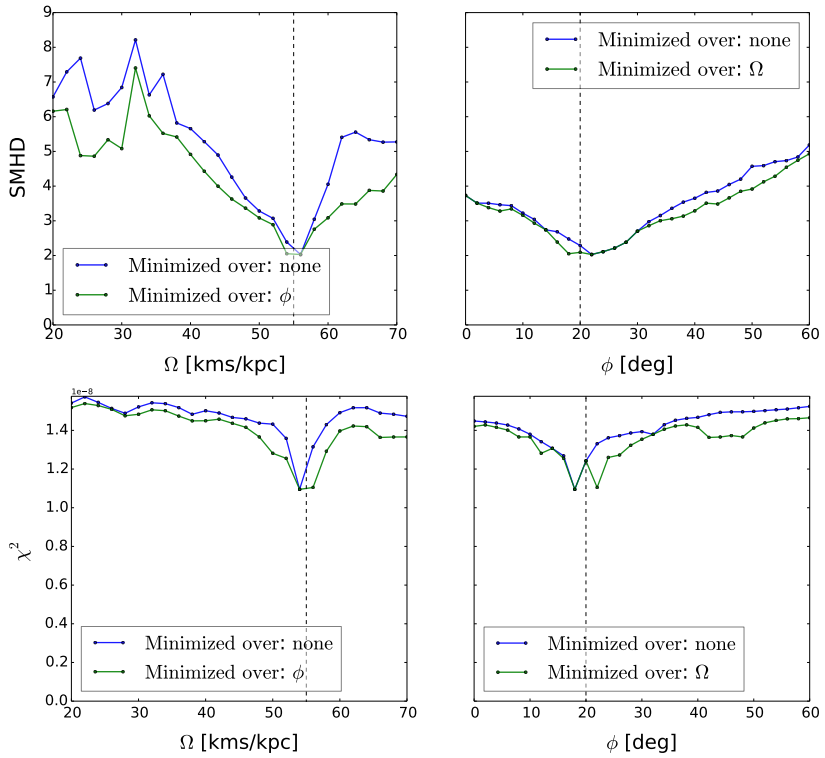


Figure 6.9: Variation of SMHD (top) and χ^2 (bottom) for the test of the emissivity law described in Section 6.5.2.2. The curves have the same meaning as in Fig. 6.8.

χ^2 measures the overlap between the data and model densities. Adjusting (ϕ, Ω_p) moves density around the (l, v) plane. If the model density is similar to the observed one, but shifted in the (l, v) plane, then the overlap between the two is low and χ^2 tells us that the model is bad, despite the fact that a small adjustment to the model can lead to a significant improvement. This happens because χ^2 compares only the contents of the same bin, and does not take into account any cross-bin information. So, the width of the minimum in the χ^2 is essentially measuring the width of the features. We discuss this further in Section 6.5.3.

None of the distances reaches the value of zero because, as we noted above, the model used to produce the mock data set is not included, and also the evolutionary time between the two is different. For the SMHD the value of the minimum is above the value of the plateau in Fig. 6.6, as are variations produced in the SMHD by all but the smallest variations of the parameters. Thus for this particular set of models we are justified in asserting that the time is a parameter that can be neglected if we are interested in recovering the value of the pattern speed and of the angle.

6.5.2.2 Effects of the emissivity law

The second test we run is the **alpha test**. In this test, we change the value of the exponent in Eq. (6.1) from $\alpha = 1$ to $\alpha = 3$ when building our mock dataset. All the other parameters retain the values they had in the basic test. Since a synthetic (l, v) plot depends on the emissivity law α, so do the features, even though, as shown in Fig. 6.3, they do so only weakly when the emissivity law is varied within physically plausible limits. The purpose of this test is to test whether we can recover the correct parameters if the models are built using the wrong emissivity law, which we can think of as a crude test of how sensitive the fit is to

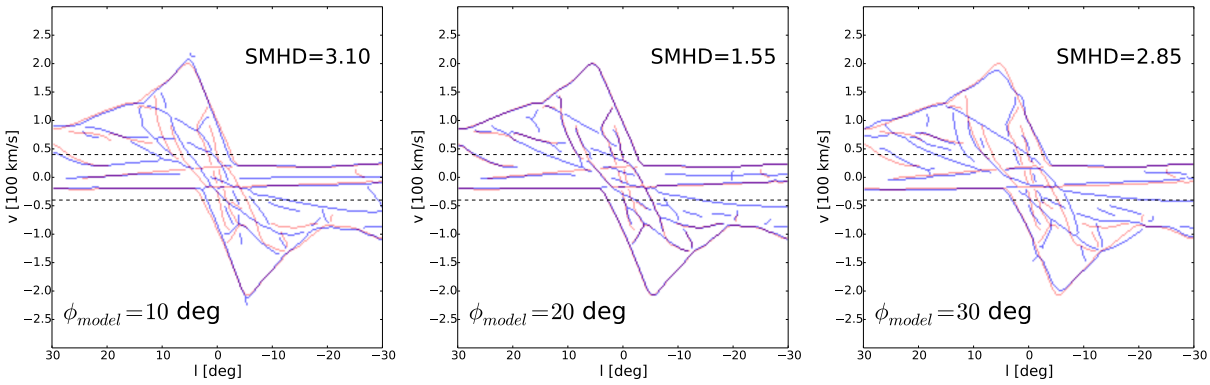


Figure 6.10: Mock data for the alpha test (Figure 6.7, panel (b)) overlaid on three models. The red curves show the mock data features, the blue the models. The purpose is to show how the features and the SMHD change when we move away from the correct model. The central picture is the model with $\Omega_{\rm p} = 54$ km s^{-1} kpc^{-1} and the correct value for the angle $\phi = 20°$. We see that the features of the mock data are almost a subset of the features of the model (but not quite). The other two pictures show models with the same $\Omega_{\rm p}$ but $\phi = 10°$ and $\phi = 30°$. In these, all the features are moved slightly in the (l, v) plane with respect to their positions in the middle model. This movement is one of the reasons that makes χ^2 unsuited for matching the longitude-velocity diagrams. On each figure the values of SMHD between the blue and red models are shown.

assumptions about ISM chemistry and local physics.

When $\alpha = 3$, the brightness temperature of the mock dataset does not depend linearly on the total column density, but on its third power instead. This is quite an extreme choice, probably well over the edge of the physically reasonable values. The mock dataset for this test is shown in Fig. 6.7(b). We see that the features change significantly from the basic test. In particular, some features disappear and are not visible anymore, while some features are enhanced. Features that disappear tend to be associated with velocity crowding, while features that survive tend to be associated with real overdensities in the gas distribution. The features for the $\alpha = 3$ are almost, but not exactly, a subset of the features for $\alpha = 1$. Therefore in this test we are trying to match a mock dataset with much fewer features than our models have.

Fig. 6.9 shows the result of the fitting. Again the model is clearly identified by the minimum SMHD, though the minimum is not as deep as found in the basic test when the correct value of α was adopted. This happens because the best model now contains extra features that are not contained in the mock dataset. Nevertheless, the SMHD still does quite a good job in identifying the correct model. This test illustrates the robustness of the method: it shows that if we use the wrong emissivity law we should still be able to retrieve the correct model.

In this test, χ^2 is clearly outperformed by SMHD (compare top and bottom rows in Fig. 6.9). χ^2 does exhibit a minimum around the parameters of the correct model, but it is weaker than the minimum of the SMHD. χ^2 is much flatter than SMHD when we are far from the correct model, with minima appearing in χ^2 at $\phi \simeq 45°$ in regions, while in the same regions SMHD points to the correct model. The explanation for this is that χ^2, by being heavily dependent on the intensities at each point, can fail to recognise that the overall morphology is similar.

As a by product of the fact that for $\alpha = 3$ the features diminish, this test shows that if for some other

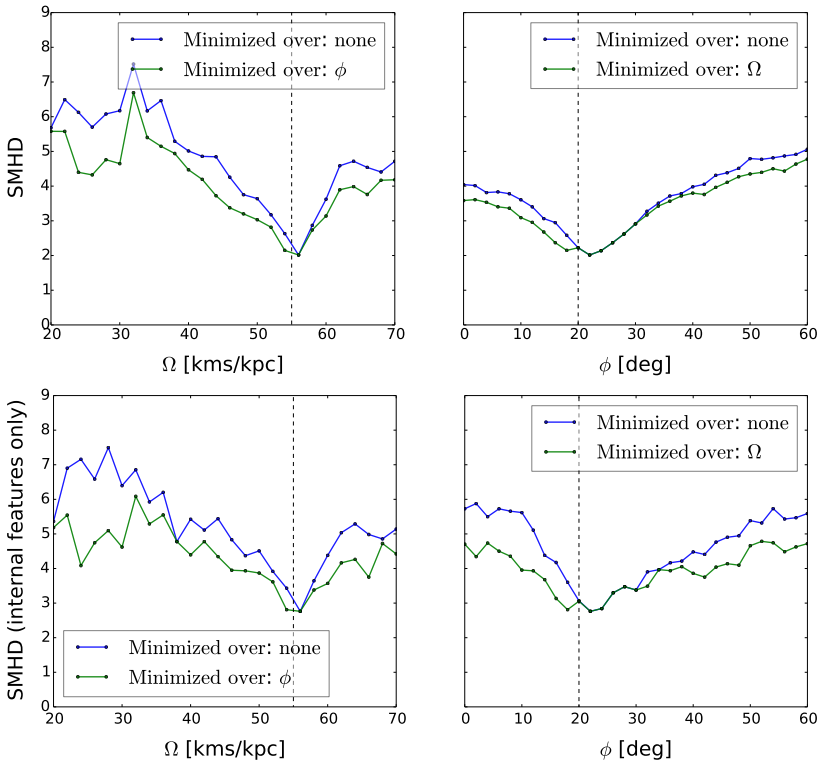


Figure 6.11: Variation of SMHD in the contamination test (Section 6.5.2.3). In the top row the SMHD is calculated on all features, while in bottom row SMHD is calculated only on internal features (i.e., excluding the envelope). Colours mean the same as in Fig. 6.8.

reason we fail to identify some features in the data, then we might still be able to find the correct model. In Fig. 6.10 we overlay the mock data of the alpha test with the model that has the correct parameters and with two models that have one parameter, the angle, different from the correct value.

We do not plot the results for ED; they are unchanged from Fig. 6.8, as one might expect.

6.5.2.3 Effects of contamination

The last we run is the **contamination test**. The mock dataset for this is built by manually adding some extra features on top of the mock dataset for the basic test and it is shown in Fig.6.7(c). The purpose is to test the robustness of the method against the presence of spurious features in the data. These could represent some features that have been included in the data but are not really wanted, for example because they are caused by effects not taken into account by the models (if the connecting arm were due to magnetic fields and we do not include these in our simulations, this would count as contamination in this case). Fig. 6.11, top row, shows the results of this test. These are similar to the results for the alpha test, thus showing robustness of the method against the presence of contamination. This test can be viewed as adding features as opposed to the alpha test were we are removing features. Thus the method is robust both against adding extra features and removing good ones.

6.5.2.4 Variations on SMHD

In Fig. 6.11, bottom row, we show the results of using the SMHD *without* the envelope using the contaminated dataset. It shows that even if one considers *only* the internal features, the correct model can still be identified, although the SMHD becomes more noisy. So, for this particular class of models, the internal features alone contain enough information to identify the correct model, albeit not as well as the envelope alone.

Finally, we have tested how the SMHD performs compared with unsymmetrized versions of the MHD. Given a choice between the two possibilities of using the MHD, the unsymmetrized version performs better than the symmetrized version either in the alpha test or in the contamination test, but not in both. At worst, the unsymmetrized MHD displays a shallower minimum, which is more difficult to identify, and weak secondary minima can appear. This agrees with our considerations in Sect. 6.4.2. The symmetrized version provides a compromise able to handle a wider range of situations.

6.5.3 Behaviour for families of models that are far from the fiducial model

Finally, as a more realistic test, we consider what happens when the models we search over are very different from the model from which the data are generated. We use mock data generated from the model given in Appendix D, with pattern speed $\Omega_\mathrm{p} = 48\ \mathrm{km\,s^{-1}\,kpc^{-1}}$ and bar angle $\phi = 30°$, projected with $\alpha = 1$. As the form of the underlying potential of this galaxy model is very different from the Englmaier & Gerhard (1999) potentials that we try to fit to it, we do not expect these parameters to be retrieved correctly. The aim is instead to compare the quality of the best fits to the mock data according to the SMHD, ED and χ^2 distances.

Fig. 6.12 shows the mock data used in this case, together with the models that minimize SMHD, ED and χ^2. It is clear that the model that reproduces the features best is, unsurprisingly, the one that minimizes the SMHD. This shows that the SMHD is the method that finds the model whose visual appearance is closer to that of the mock dataset. In other words, it is the method that best mimics the results of using visual inspection, the main method of comparison used to date in the literature. The model that minimises the envelope distance matches the envelope very well, but fails to match the internal features well. It is evident that the model that minimises χ^2 is entirely unsatisfactory.

In Fig. 6.13 we show how the distances vary with model parameters. The ED exhibits multiple minima, indicating degeneracy. In fact, it has a secondary minimum at the location of the best SMHD model, that is weaker than the main minimum. The χ^2 is more flat, with weak minima here and there. If we consider that in this test the same emissivity law is used for the data and the models, we argue that the situation would be even more hopeless than the plot indicates if one were trying to use χ^2 without knowledge of the (unknown) correct models for the ISM chemistry, small-scale physics, emissivity and radiative transfer. This indicates that χ^2 is not an appropriate goodness-of-fit measure, at least not until one understands in detail all the different parts of physics involved. In Sect. 6.7 we come back to this topic and discuss the reasons for this behavior.

We note that in the example shown in Fig. 6.13 the ED is the only method that recovers parameters that are close to the right ones. This suggests that the ED might in fact be more robust in recovering the parameters ϕ and $\Omega_{\rm p}$, which are the two parameters that most influence the envelope. However the ED can be highly insensitive to other parameters that influence more the internal features: for example it is extremely insensitive to the quadrupole's exact length and strength, see item 4 in Section 5.6. This means the ED cannot recover these characteristics of the bar, while the SMHD can.

6.6 Application to real data

Having applied the SMHD to mock data, we now test how well it works when applied to the features in the real data identified in Section 6.2. We do not produce any new models here, but instead compare the fits provided by three of the best models from the literature: the standard model of Englmaier & Gerhard (1999), that of Bissantz et al. (2003) (their Table 1) and the best overall model of Rodriguez-Fernandez & Combes (2008), the rotation curve of which is shown in their Fig. 6.

The model of Englmaier & Gerhard (1999) is stationary in a frame that corotates with the bar pattern speed of $\Omega_{\rm p} = 55\ {\rm km\,s^{-1}\ kpc^{-1}}$; there is no spiral arm component in their potential. The model of Bissantz et al. (2003) is nonstationary, as the potential includes a bar and a spiral arm component rotating respectively at $\Omega_{\rm p} = 58.6\ {\rm km\,s^{-1}\ kpc^{-1}}$ and $\Omega_{\rm p} = 19.6\ {\rm km\,s^{-1}\ kpc^{-1}}$. Rodriguez-Fernandez & Combes (2008) includes two bars, a big bar and a smaller nuclear bar making a constant angle of 55° with the first, both rotating at the same pattern speed $\Omega_{\rm p} = 30\ {\rm km\,s^{-1}\ kpc^{-1}}$. The potential has no spiral arm component.[5]

Our reconstructions of the surface density distributions of these three models are shown in Fig. 6.2. The corresponding (l, v) distributions for a viewing angle of $\phi = 20°$ are shown in Fig. 6.14. We have compared these density distributions and (l, v) plots with the appropriate figures from the original papers and we found that the location of the features agree very well. This agreement is remarkable, given that our models are based on a Eulerian grid simulation, whereas Englmaier & Gerhard (1999) and Bissantz et al. (2003) used a SPH code while Rodriguez-Fernandez & Combes (2008) use a sticky-particle code. It is particularly interesting that we reproduce the latter so well: sticky particle simulations in principle solve different fluid equations than our FS2-based method. However, this may not come as totally unexpected if we take into account the results of Chapter 3: we have found that the results of our Eulerian grid-based simulations are well understood in terms of closed orbits, but we also expect sticky particle to follow orbits except when they collide. The fact that the three methods give similar results on the scales we are interested in corroborates our claim that we should first try to match the observational data with simple models, so as to reproduce the overall structure, and only later move to more refined models to reproduce the details. When the gross structure is not known, it is in general a pointless exercise to add effects if they turn out to be of secondary

[5] In order to reproduce the rotation curve plotted in Fig. 6 of Rodriguez-Fernandez & Combes (2008) we found that we had to replace the exponent of 1/4 in their expression (9) for r_s with an exponent of 1/2. We assume that this is a typographical error in the paper, even though it means that their boxy Gaussian bulge is actually a boxy exponential bulge.

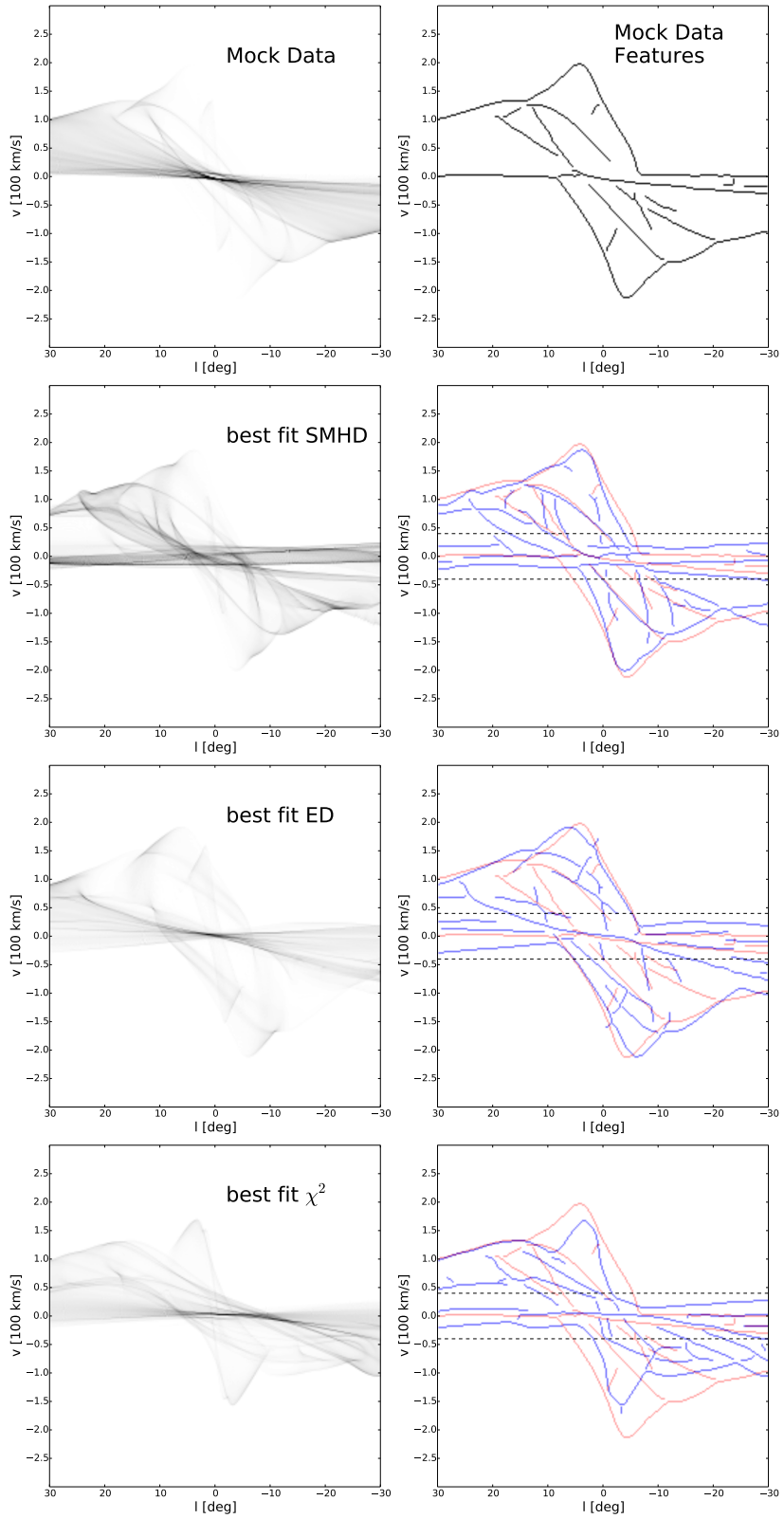


Figure 6.12: A comparison of the best fits obtained by minimising the SMHD, χ^2 and ED for the mock data of Section 6.5.3. The top panel shows the mock data and its features. Subsequent panels show the models that minimise SMHD, ED and χ^2, respectively. In blue the models and in red the data. The best SMHD model matches the mock data features remarkably well, while the best χ^2 model looks very different from the mock data. The best ED model displays the best-matching envelope, but the internal features are not matched well. The mock data are drawn from a very different potential than the models, and so the purpose of this is not to retrieve the correct parameters, but to show that the best SMHD model has features better matching the mock data than the best ED model and the best χ^2 model.

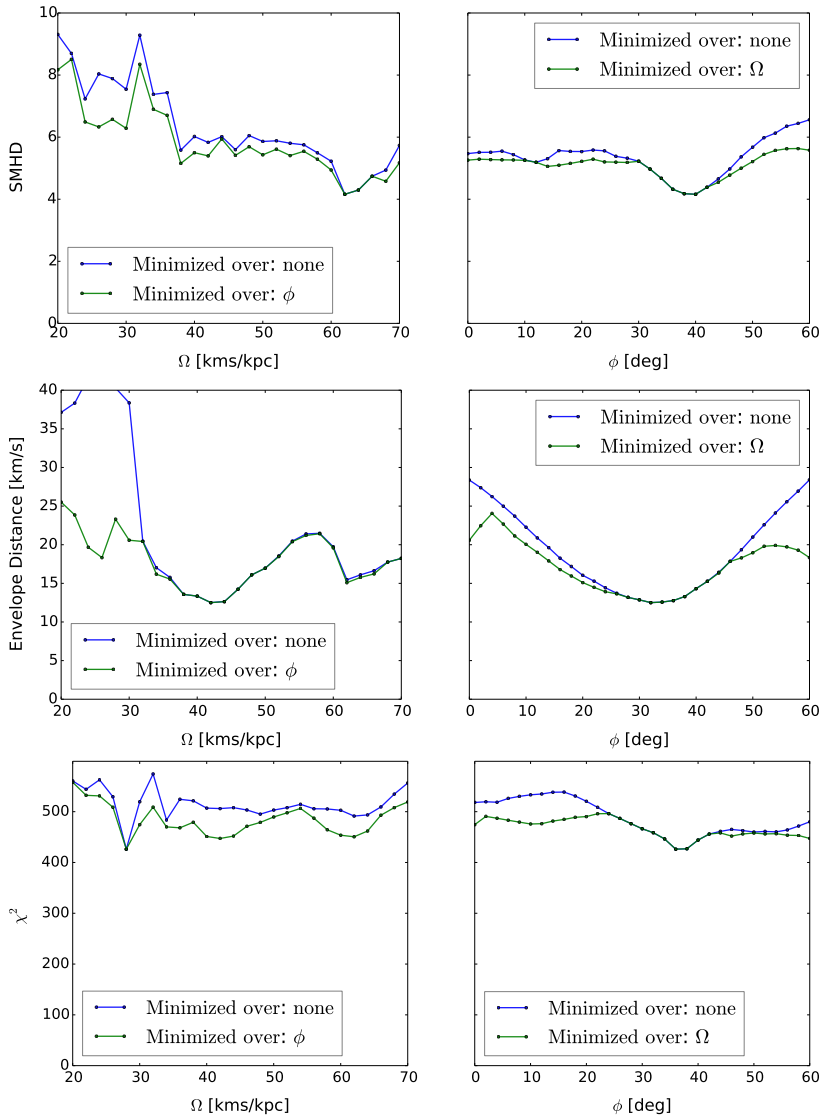
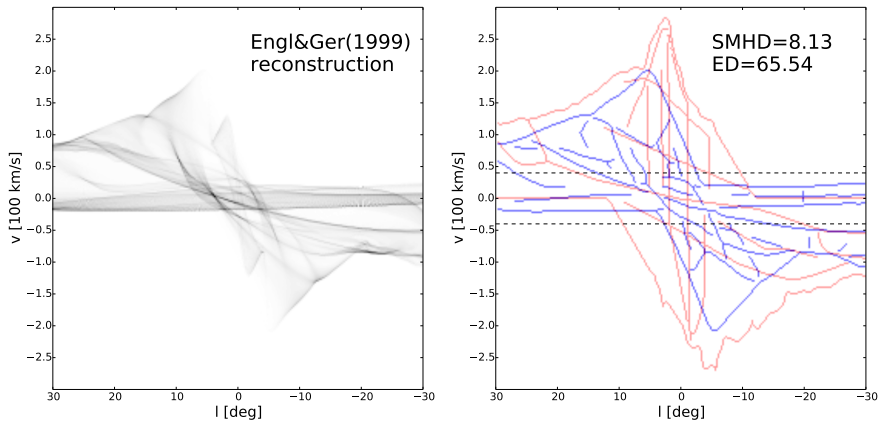
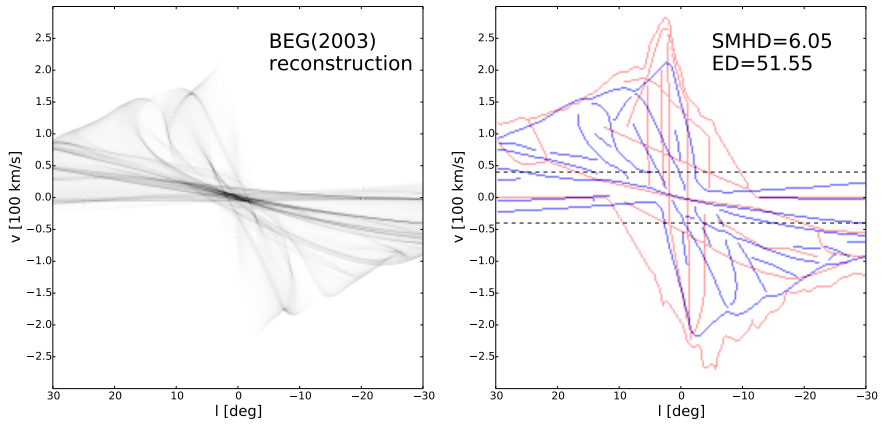


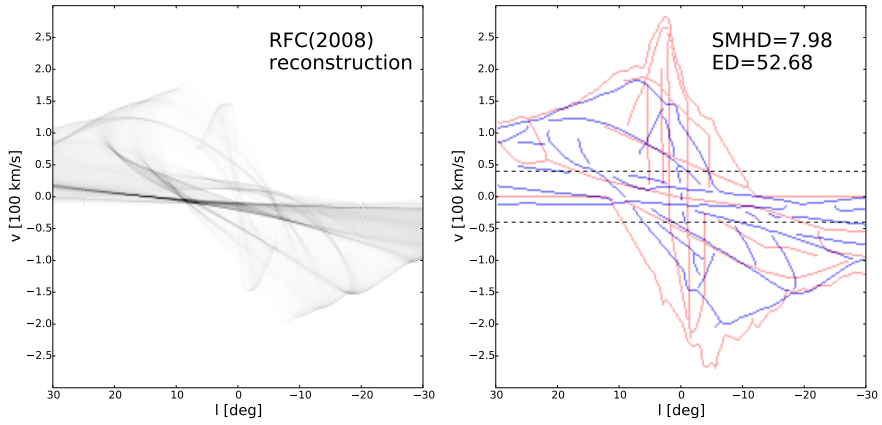
Figure 6.13: Variation of SMHD (top row), ED (middle) and χ^2 (bottom row) with model parameters for the mock data of Section 6.5.3. Colours indicate the same thing as in Fig. 6.7. We see that the ED is degenerate, displaying two minima. The χ^2 is flat and does not indicate clearly the correct direction to the nicer model. The SMHD shows a more definite minimum. Moreover, as shown in Fig. 6.12, the best SMHD is superior to the best ED and the best χ^2 models.



(a)



(b)



(c)

Figure 6.14: (l, v) plots corresponding to our reconstructions of some of the best models in the literature compared to observations. The corresponding face-on densities are shown in Fig. 6.2. From top to bottom, our reconstruction of Englmaier & Gerhard (1999) standard model, Bissantz et al. (2003) standard model and Rodriguez-Fernandez & Combes (2008) best overall fitting model. On the right we overlay the features of the models with the observational features discussed in Sect. 6.2. In blue the models and in red the data. All models have a viewing angle (angle between the Sun-Galactic Centre line and the major axis of the bar) of $\phi = 20°$.

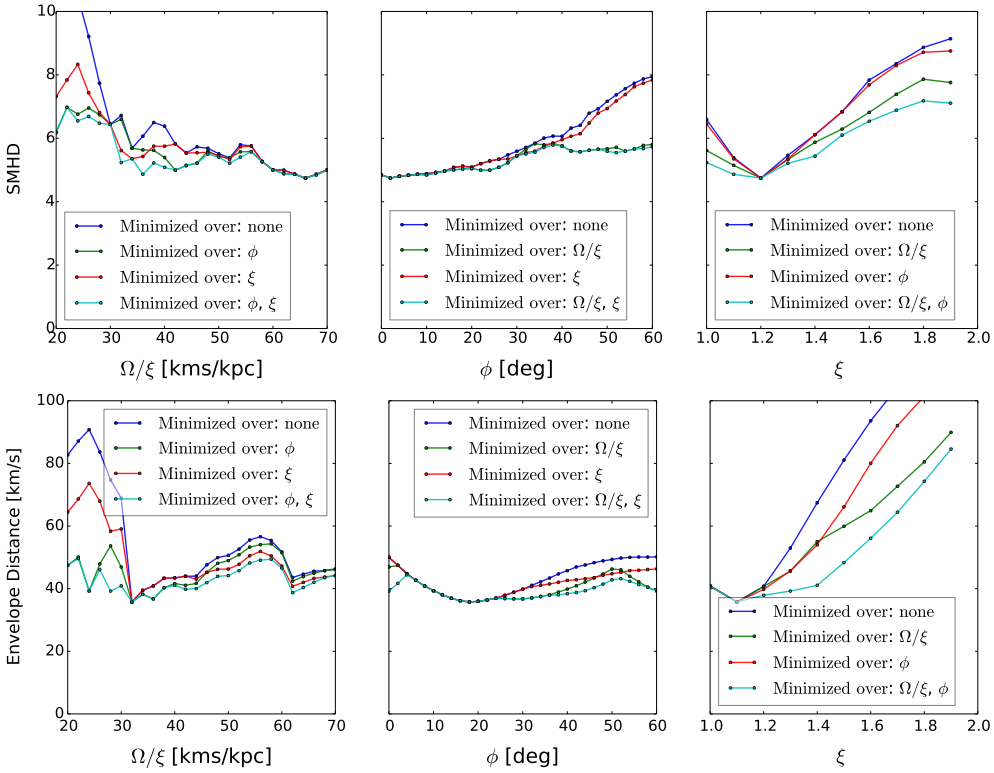


Figure 6.15: SMHD (top row) and ED (bottom) for our reconstructions of the models of Englmaier & Gerhard (1999) scaled to fit features in real data (Section 6.6). The free parameters of the models are the viewing angle ϕ, the pattern speed Ω_p and the velocity scaling factor ξ. As in figure 6.8, the blue curves show how the distances change along straight lines aligned with the $(\phi, \Omega_\mathrm{p}, \xi)$ coordinate system that pass through the location $(\phi, \Omega_\mathrm{p}, \xi)$ of the best fit. The other curves plot the minimum distances when one or both of the other parameters is allowed to vary freely.

importance.

For each of our three reconstructions, we use the algorithm of Section 6.4.1 to extract the features. These are overlaid on the right column of Fig. 6.14 with the features extracted by eye from the observations (Section 6.2). Unfortunately, these plots highlight the shortcomings of the models we employed. The main problems are:

1. No model is able to reproduce the high velocity peaks at $l \simeq \pm 3^\circ$ and $l \simeq -4^\circ$. In particular the Rodriguez-Fernandez & Combes (2008) model has a very low peak velocity of less than 200 $\mathrm{km\,s^{-1}}$. However, results of Chap. 3 strongly suggest that this is due to insufficient resolution.

2. No model reproduces the huge forbidden velocities at $(l > 0, v < 0)$ and $(l < 0, v > 0)$. Rodriguez-Fernandez & Combes (2008) do better than the others in this, but large uncovered portions remain.

3. Broad features, such as the 3-kpc arm, are not reproduced well by the Englmaier & Gerhard (1999) or Bissantz et al. (2003) models. The model of Rodriguez-Fernandez & Combes (2008) does a better job here, providing a good fit to both the near- and far-side 3kpc arms.

4. The very complicated central structure, for example the vertical features, is not reproduced in any model. According to our interpretation of Chapter 5, this requires the inclusion of chemistry.

As a very limited test of whether one could easily improve on this situation, we fit the same set of models used in Sect. 6.5, based on our reconstruction of the Englmaier & Gerhard (1999) potential, to the real Galaxy features. As before the models have Ω_p in the range 20-70 $\mathrm{km\,s^{-1}\,kpc^{-1}}$, ϕ in the range 0-60°, all viewed at evolutionary time $t \simeq 370\,\mathrm{Myr}$ and projected assuming $\alpha = 1$. To allow the models some extra freedom we allow an extra parameter ξ that scales all velocities of the gas. When velocities are scaled such that $v(r) \to \xi v(r)$, the other quantities scale in the following way: $\Phi \to \xi^2\Phi$, $M \to \xi^2 M$, $c_\mathrm{s} \to \xi c_\mathrm{s}$, $\Omega_\mathrm{p} \to \xi\Omega_\mathrm{p}$.

In Fig. 6.15 we show the results of minimising the SMHD (top row) and ED (bottom row). We note that the graphs are smooth and not dominated by noise. The best fits are different according to the two methods. In Fig. 6.16, panel (a) and (b), we show the best fits according to SMHD and ED respectively. The parameter values for these are the overall minima in the graphs of Fig. 6.15. These are $\Omega_\mathrm{p}/\xi = 64$ $\mathrm{km\,s^{-1}\,kpc^{-1}}$, $\phi = 2^\circ$ and $\xi = 1.2$ for the best SMHD, and $\Omega_\mathrm{p}/\xi = 32$ $\mathrm{km\,s^{-1}\,kpc^{-1}}$, $\phi = 18^\circ$ and $\xi = 1.1$ for the best ED model. Unfortunately, we believe that both should be considered unsatisfactory, as in each important ingredients are missing. No model is able to reproduce high forbidden velocity and, at the same time, the high velocity peaks, and many features are also not reproduced well. Interestingly, in the best ED model a weak vertical feature appears at negative velocities, approximately at $l \simeq -7^\circ$, $-200 < v < -100$ $\mathrm{km\,s^{-1}}$. In the face-on view of the Galaxy, this feature corresponds to an offset shock lane, in agreement with our interpretation in Chapter 5. Inclusion of the chemistry would make these features more prominent in the (l, v) plot.

In panel (c) of Fig. 6.16 we show a further model, labelled "GE", that we found in our reconstruction of the potential of Bissantz et al. (2003) for $\Omega_\mathrm{p}/\xi = 30$ km s^{-1} kpc^{-1}, $\phi = 34°$ and $\xi = 1.1$. This model has an envelope that matches the observed one amazingly well, filling the right forbidden velocities region. If one were to judge this model only from the envelope, akin to what Weiner & Sellwood (1999) did, this would be considered a very good model. We do not consider this to be a particularly good model, however. The internal features are completely wrong, except close to the molecular ring region. The very central region is almost featureless, and it exhibits nothing similar to the 3-kpc arm, the connecting arm or the CMZ. This model illustrates for a real case that the envelope is not enough to constrain the Galaxy potential, and in this chapter we argue that the next piece of information that should be taken into account is given by the internal features.

We note however that the ED does identify values for the parameters ϕ and Ω_p for our Galaxy that are closer to the suspected truth (see Chapter 5). As described in the previous Section, it is possible that the ED is more robust with respect to these two parameters. However, the fact that we found a model (GE) with an almost perfect envelope means that the information the ED could have provided has already been exploited. To gain further information about our Galaxy it is necessary to turn to other methods.

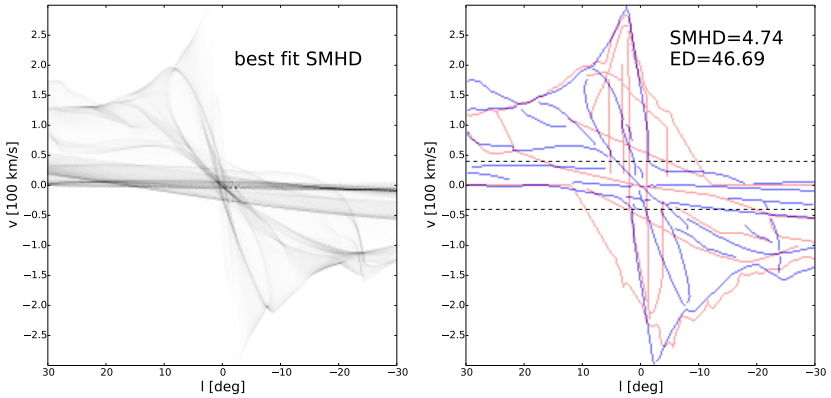
6.7 Discussion

6.7.1 Comparison of SMHD with ED and χ^2, and their limitations

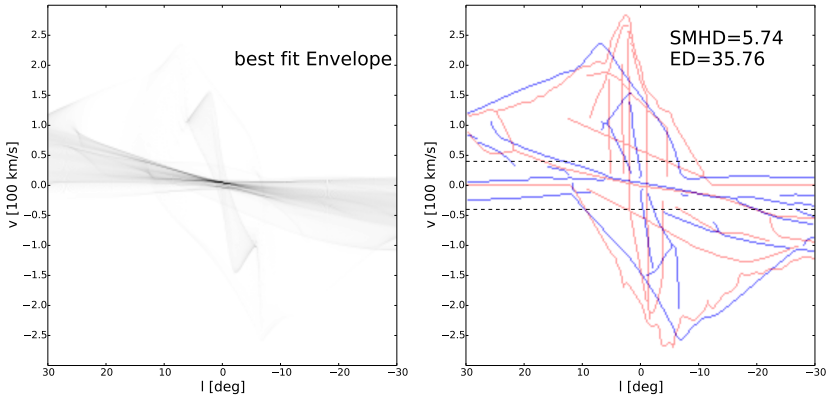
Apart from "by-eye" comparisons, the two most widely used ways of fitting models to observed (l, v) distributions have been some variations of the ED (eq. 6.6, e.g., Weiner & Sellwood, 1999; Englmaier & Gerhard, 1999) and χ^2 (eq. 6.5, e.g., Pettitt et al., 2014).

While the ED is very robust with respect to changes in the emissivity law, it has the obvious disadvantage of neglecting all the information coming from internal features; thus, the degeneracy of the problem is increased. We know, for example, that it is possible to reproduce *any* terminal velocity curve in the $(l > 0, v > 0)$ and $(l < 0, v < 0)$ quadrants by means of gas moving on purely circular orbits. In Sect. 6.5.3 we have shown that the ED can be degenerate when the SMHD is not, so the latter can provide a better-fitting model than the former. In Sec. 6.6 we have shown a model that fits well the envelope of the real data, but is overall unsatisfactory as it fails to fit the internal features. The internal features must certainly contain additional information and should be taken into account when comparing the data with the models. In Fig. 6.11 we give an example of a case in which internal features alone are sufficient to identify the correct galaxy model, albeit not quite so securely as when the envelope is known as well.

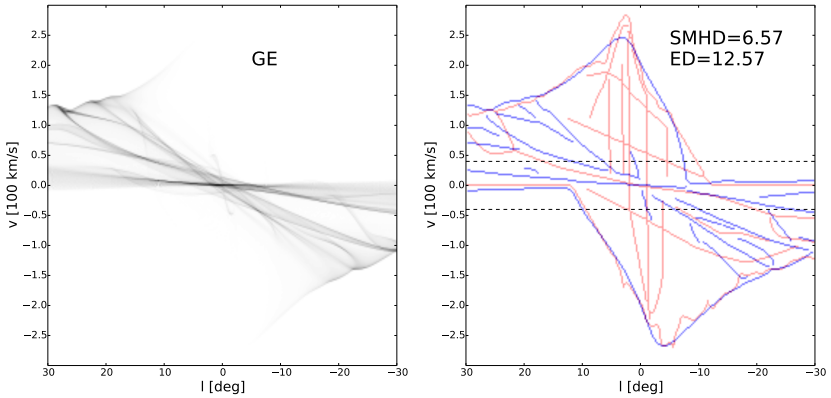
In constrast, χ^2 makes full use of all of the available data, but it suffers from serious drawbacks when one tries to use it to contrain the geometry and dynamics of the Galaxy. One of the main concerns is that χ^2 only compares model versus observed intensities in the same (l, v) bin and does not take into account any cross-bin information. If, for example, a model displays features that are very similar to the observational features,



(a)



(b)



(c)

Figure 6.16: Panels (a) and (b) show the best fit models in our reconstruction of the potential of Englmaier & Gerhard (1999), according respectively to SMHD and ED. These two models have values of the parameters corresponding to the values of the minima in Fig. 6.15. Panel (c) shows a model with a very well matching envelope. The simulation underlying this model runs in our reconstruction of the potential of Bissantz et al. (2003), with $\Omega_\mathrm{p} = 30\ \mathrm{km\,s^{-1}\ kpc^{-1}}$, $\phi = 34°$ and $\xi = 1.1$. All models are taken at evolutionary time $t = 367\,\mathrm{Myr}$.

but are slightly misplaced in the (l, v) plane, the χ^2 distance can suggest that the model is terrible; in the most extreme case, even a bland, featureless model might be a formally better fit, while visual inspection would suggest that the models are actually quite good. This means that raw χ^2 fails to capture the essence of what is important in comparing model and data in this case. This is what happens in the situation of Sect. 6.5.3, where we have shown that χ^2 provides unsatisfactory fits even when the same emissivity law is used for the data and the models.

χ^2 has other drawbacks. It is computationally expensive. As it is entirely dependent on local intensities, its use necessarily requires detailed modelling of local physics on small scales, chemistry, radiative transfer, and gas dynamics all at the same time, which is very time consuming. On the other hand, models need to be cheap to test because the space of possible models is large. We would like to know not only the pattern speed and orientation of the bar, but also its mass, length, axis ratio, and possibly more. Computational expense is the reason why Pettitt et al. (2014), despite having a full radiative transfer model, did not use it when fitting the data and relied on a very simplified projection model akin to ours. χ^2 is clearly not the best choice in situations such as the alpha test in Sect. 6.4, where everything in the model is correct except the emissivity law, which we used as a proxy to describe how accurately the local physics is modelled.

Current models of the gas flow in the Milky Way (Sect. 6.6) are unsatisfactory. We argue that the sensible way of addressing this is to note that, as we found in Sect. 6.4, the features in the (l, v) distribution give valuable constraints on the Galactic potential that are largely independent of the details of ISM chemistry and small scale physics. Therefore one should first build models that match well the broad morphology of the observations to narrow down the potential, and only later refine this to match the details. As we've shown in Sect. 6, the envelope is too degenerate for such a task, even in the case of real data: a model that reproduces only the envelope well can still be unsatisfactory.

Thus one should use the SMHD or similar scheme to locate the range of broadly acceptable potentials and pattern speed(s). Only when this large-scale structure has been constrained does it make sense to switch to more sophisticated models that include chemistry, proper small scale physics and proper radiative transfer modelling. We believe that χ^2 (or similar) should play an important role only in this last step. An alternative to χ^2 that might be worth considering in this last step is the "earth mover distance" (Appendix E).

We emphasise that, unlike χ^2, the SMHD is a purely qualitative measure that cannot sensibly be used to provide formal uncertainties on the parameters of models that fit the observations. Given the present ambiguity as to the overall form of the Galactic potential, we would argue that any such attempt would be misleading.

6.7.2 Identification of "features"

The SMHD returns a number that quantifies the dissimilarity of two sets of features in the (l, v) plane: the higher the number, the more dissimilar the features. The procedure for identifying features in data and

models requires some remarks. For models, we have a fully automatic algorithm (Sect 6.4.1) that, given a model (l, v) distribution returns the features as 1-pixel wide lines. The features we identified as being important are bright ridges and the envelope. Therefore, the algorithm simply detects ridges and envelope given a model (l, v) distribution.

6.7.2.1 Features in data

It is natural to ask whether the data could be analysed in the same way as the models to extract features automatically. Unfortunately, this turned out to be problematic: unlike our simple, smooth hydrodynamical models, the real data exhibit substructures due to clumpiness that are identified as spurious ridges by our ridge-detection algorithm. Moreover, the analysis of features in the data often involves looking at different latitude slices (e.g. in connection with the tilt and the envelope, see item (6) in Sect. 2.2 and item (5) in Sect. 5.3), and each feature may require a special analysis and considerable work as the example of the far-side 3-kpc arm (Dame & Thaddeus, 2008) shows. This is clearly beyond our algorithm's capabilities and requires the skills of experienced astronomers. For this reason, we rely on human wisdom for the identification of features in the real data.

6.7.2.2 Suitability of models

Any model for the gas flow comes with a series of implicit or explicit simplifying assumptions. In the present chapter our gas models are 2D; we neglect the vertical dimension, which could also play an important role. We do not include heating and cooling processes due to a variety of sources, such as supernova explosions and stellar winds. Modeling the gas as a smooth fluid means neglecting all the grainy structure, such as individual clouds with peculiar velocities. Indeed, the crude modelling with the Euler equation is applicable only in a coarse-grained sense, and does not take into account explicitly local turbulence, temperature variations and the multiphase nature of the ISM. Lastly, we neglect the self-gravity of the gas that could be important especially near shocks or other structures where gas accumulates. There is nothing to prevent us from applying the SMHD method to more (or less) sophisticated models, as long as they produce smooth (l, v) distributions; as discussed above, we argue that for finding the large-scale structure of the gas distribution, which is currently poorly understood, the simpler the model the better.

6.7.2.3 Possible extensions

The method can easily be adjusted to incorporate our beliefs about how features are generated. For example, one could argue that the envelope is of a different nature than the internal features, and therefore we should calculate two separate SMHDs, one that matches only the envelopes and one matching only the internal features. Another example is provided by the fact that, because of absorption, for many features we know whether they are caused by material in front of the Galactic Center (for example the 3-kpc arm) or behind it (for example, the $135\,\mathrm{km\,s^{-1}}$ arm, see Cohen, 1975). It would therefore be natural to match features that we know are in front of the GC with features that are in front also in the models. A further example is

provided by the tilt of the Inner Galaxy (Burton et al., 1992): if we believe that a part of the Galaxy – for example the inner nuclear disk – is tilted, then we would like to match observational tilted features only with features that are produced within the corresponding region in the models. This is straightforward to do by fitting multiple SMHDs.

As an illustrative example, to incorporate the information on whether features are in front or behind the GC, we can proceed as follows. We divide data features in three sets: $\mathrm{D_f}$ are features we know lie in front of the GC, $\mathrm{D_b}$ those that lie behind, $\mathrm{D_u}$ those whose position is unknown. Model features are divided only in two sets, $\mathrm{M_f}$ and $\mathrm{M_b}$, as for each model feature we always have information on its position with respect to the GC. Then a suitable definition of SMHD that takes into account the new information (and reduces to the previous definition in absence of new information) is:

$$\mathrm{SMHD_{new}}(a,b) = \frac{A}{2N} + \frac{B}{2M}, \tag{6.7}$$

where

$$A = \mathrm{MHD}(\mathrm{D_f}, \mathrm{M_f}) + \mathrm{MHD}(\mathrm{D_b}, \mathrm{M_b}) + \mathrm{MHD}(\mathrm{D_u}, \mathrm{M_f} + \mathrm{M_b}) \tag{6.8}$$

and

$$B = \mathrm{MHD}(\mathrm{M_f}, \mathrm{D_f} + \mathrm{D_u}) + \mathrm{MHD}(\mathrm{M_b}, \mathrm{D_b} + \mathrm{D_u}). \tag{6.9}$$

Other prior information could be taken into account in a similar way, dividing the features in different sets and defining the rules by which these sets should be matched.

6.8 Conclusion

We have proposed a new way of fitting model (l, v) distributions to observations. We have argued that one can separate the effects of the large-scale dynamics and structure of the Galaxy from those due to details of chemistry and other small-scale physics. Based on this, our SMHD provides a way of measuring distances between features in models versus corresponding features in the observations.

We have tested the ability of our method to fit models to mock data generated under a variety of conditions, and have compared it to alternative methods. To the best of our knowledge, this is the first time that the ability of such methods to retrieve model parameters by fitting to (l, v) distributions has been investigated systematically. We have also explicitly demonstrated the importance of internal features in the (l, v) plots, that have been the basis for comparison of (l, v) plots to observations by many authors who have run simulations.

We find that our feature-based SMHD method works well and is much more robust than other methods. It works in cases in which the assumed (crude) emissivity law is wrong or when the data are contaminated.

The ED, as expected, is often degenerate in cases where the SMHD is not, as the ED exploits only a small part of the information available in the data. We found this to be true both for mock and real data. On the other hand, we found χ^2 to be unsuited to the task of matching longitude-velocity diagrams when this requires exploration of a huge parameter space. It works well only in the very vicinity of the correct solution, and at the price that *all* pieces of physics are taken into account in producing synthetic (l, v) plots, including those that can be disentangled from the dynamics of the gas. The main reasons for this behaviour are (i) that χ^2 fails to take into account cross-bin information, which means that it tends to favour models that have little structure; (ii) its use is computationally expensive as it requires accurate modelling of (l, v) detailed intensities, which in turn requires accurate modelling of the chemistry of the ISM, small scale physics, emissivity, radiative transfer and so on. We argued that such calculations are unnecessary if all one wants to do is to constrain the Galaxy's gravitational potential and the large-scale distribution of its gas. Given that current dynamical models of the Galaxy appear to be far from the truth and that fitting the envelope alone is not enough, one should first constrain the gross morphology found in data using, for example, SMHD, and only later, once one has almost nailed down the potential, turn on more details of physics and finally use χ^2 to take advantage of its statistical interpretation.

Our method is computationally inexpensive because it relies on simple hydro simulations and avoids the need to model chemistry or to carry out sophisticated radiative-transfer modelling. This makes it suitable for carrying out large, systematic scans of model parameter space. It is easily applied to time-dependent simulations. It can be used to test the reality of observed features (by comparing fits with and without the feature present) and can naturally be extended to test hypotheses, such as "this feature belongs to a foreground spiral arm" or "that feature is the trace of the x_1 orbit" with a little extra analysis of the internal dynamics of the models used for comparison. It does, however, require some work in that it relies on features being identified "by hand" in the data. It also requires that the models used for comparison are smooth enough to allow the feature extraction algorithm to work; sophisticated models that produce clumpy structures would probably not be suitable.

We have reconstructed, and reanalysed by applying our method, some of the best models from the literature that were constructed to fit the Dame et al. (2001) and Kalberla et al. (2005) data on the Milky Way, but find that they produce surprisingly poor fits. We have made an initial attempt to fit the data, but our family of models was too limited. A better fit is likely to be obtained if we include in the model parameters that control the characteristic of the bar (such as its mass, length, strength and axis ratio) and the circular velocity curve (for example, the density profile in the inner region). These parameters can all be fit simultaneously using the SMHD method, and the bottleneck is the computational time required to run such a high number of models.

We were able to find a model that reproduces the envelope of the emission accurately, but this fails in explaining the internal structure of the data; nevertheless, this demonstrated only that the envelope alone is insufficient to constrain the Galactic potential. As an interesting by product, we found that the large-scale morphology is not very sensitive on the simulation method. The sticky-particles code used by

Rodriguez-Fernandez & Combes (2008) gave a large-scale morphology very similar to our grid-based code, despite the fact that in principle it solves a different set of hydrodynamical equations. This does not come as completely unexpected given the results of Chapter 3 and corroborates our claim that to find the gross morphology of the Galaxy one should focus on simple hydrodynamical models.

The problem is now that of finding a sufficiently general class of model potentials to use in the comparison. One possibility is to use a generalisation of the models presented in Chapter 5 and express the first few multipole moments $\rho_l(r)$ of the Galaxy's mass density distribution in terms of splines and to develop an automatic scheme for adjusting the spline weights to minimize the SMHD distance. A good model should also take into account constraints coming from different sources, for example infrared data from the 2MASS survey (Skrutskie et al., 2006) and the correlations expected between the three-dimensional distribution of gas and independent results on the three-dimensional distribution of dust (e.g., Marshall et al., 2006; Green et al., 2014; Sale & Magorrian, 2014). The models presented in Chapter 5 show that there is every prospect for a good model to be found.

References

Baba J., Saitoh T. R., Wada K., 2010, PASJ, 62, 1413
Bally J., Stark A. A., Wilson R. W., Henkel C., 1987, ApJS, 65, 13
Binney J., Gerhard O. E., Stark A. A., Bally J., Uchida K. I., 1991, MNRAS, 252, 210
Bissantz N., Englmaier P., Gerhard O., 2003, MNRAS, 340, 949
Burton W. B., Elmegreen B. G., Genzel R., eds., 1992, Saas-Fee Advanced Course 21: The Galactic Interstellar Medium
Canny J., 1986, IEEE Trans. Pattern Anal. Mach. Intell., 8, 679
Cohen R. J., 1975, MNRAS, 171, 659
Dahmen G. et al., 1997, A&AS, 126, 197
Dame T. M., Hartmann D., Thaddeus P., 2001, ApJ, 547, 792
Dame T. M., Thaddeus P., 2008, ApJ, 683, L143
Dobbs C. L., Burkert A., Pringle J. E., 2011, MNRAS, 417, 1318
Dubuisson M.-P., Jain A. K., 1994, ICPR94, A:566
Englmaier P., Gerhard O., 1999, MNRAS, 304, 512
Fux R., 1999, A & A, 345, 787
Green G. M. et al., 2014, ApJ, 783, 114
Jenkins A., Binney J., 1994, MNRAS, 270, 703
Kalberla P. M. W., Burton W. B., Hartmann D., Arnal E. M., Bajaja E., Morras R., Pöppel W. G. L., 2005, A & A, 440, 775
Lindeberg T., 1996, International Journal of Computer Vision, 30, 465
Liszt H. S., Burton W. B., 1980, ApJ, 236, 779
Marshall D. J., Robin A. C., Reylé C., Schultheis M., Picaud S., 2006, A & A, 453, 635
Mulder W. A., Liem B. T., 1986, A & A, 157, 148
Nakagawa T., Yui Y. Y., Doi Y., Okuda H., Shibai H., Mochizuki K., Nishimura T., Low F. J., 1998, ApJS, 115, 259
Peters, III W. L., 1975, ApJ, 195, 617
Pettitt A. R., Dobbs C. L., Acreman D. M., Price D. J., 2014, MNRAS, 444, 919
Riquelme D., Bronfman L., Mauersberger R., May J., Wilson T. L., 2010, A & A, 523, A45
Rodriguez-Fernandez N. J., Combes F., 2008, A & A, 489, 115
Sale S. E., Magorrian J., 2014, MNRAS, 445, 256
Sawada T. et al., 2001, ApJS, 136, 189
Shetty R., Ostriker E. C., 2008, ApJ, 684, 978
Skrutskie M. F. et al., 2006, AJ, 131, 1163

Sobel I., Feldman G., 1968, A 3x3 Isotropic Gradient Operator for Image Processing, never published but presented at a talk at the Stanford Artificial Project

Tasker E. J., 2011, ApJ, 730, 11

Weiner B. J., Sellwood J. A., 1999, ApJ, 524, 112

Zhang T. Y., Suen C. Y., 1984, Commun. ACM, 27, 236

CHAPTER 7

Conclusions

7.1 Overview

We started our study of the gas flow in Chapter 2 with a simple axisymmetric model in which material flows on circular orbits. To improve on such a simple model we needed to understand the general characteristics of the flow caused by a central bar. In Chapter 3 we did this by studying in detail one particular barred model. We found that the keys to understand the gas flow are non-circular ballistic closed orbits and shocks. In the outer and inner regions, gas follows the x_1 and x_2 family of closed orbits respectively, while at intermediate radii the gas transfers along the shocks from one family to the other. The assumed sound speed of the gas affects significantly the gas flow, and by increasing it we push the shocks inwards and decrease the size of the nuclear x_2 disc. Unexpectedly, also the spatial resolution of the simulation plays an important role, and simulations by previous authors often did not have the necessary resolution to resolve the central regions. Depending on the parameter on the gas flow, the shocks can become unstable as a consequence of the wiggle instability.

In Chapter 4 we refined in great detail our understanding of one particular aspect of the gas flow that was neglected in the previous chapter: spiral arms that emerge from the end of the bar. We found that while it is true that in the outer region streamlines are well approximated by ballistic x_1 orbits as claimed in Chapter 3, tiny librations of the streamlines around closed orbits generate the spiral arms as kinematic density waves. Crucially, these librations are not around circular orbits as assumed in the epicyclic approximation, but around elongated orbits. Up to this point, we based our discussion on the particular bar model introduced in Chapter 3. In Chapter 5 we dropped this limitation by exploring how the gas flow is affected by a change in the parameters that characterise the bar, such as its length, strength and pattern speed.

We have compared our models with observations in the Milky Way. The model presented in Chapter 3 was based on the Binney et al. (1991) model and gave it strong support, by showing that the velocity peaks

(Chapter 2) can be explained as these authors hypothesised. We refined their model and showed that the sides of the CO parallelogram must be produced by material at the shocks rather than on the cusped orbit. We also conjectured that the wiggle instability provides a promising explanation for the observed asymmetry of the molecular emission inside $|l| \leqslant 4^\circ$. However, the model in Chapter 3 had two main drawbacks: the lack of internal features like the 3-kpc arms, and insufficient coverage of the forbidden velocity emission. The models in Chapter 5 showed that these drawbacks can be corrected by modifying the parameters that define the bar while keeping alive the key ingredients that defined the model in Chapter 3. This allowed us to constrain the pattern speed, the length and the strength of the bar. However, we have failed in finding a model that accounts for all the important features simultaneously. Due to the high dimensionality of parameter space involved, automatic fitting methods are necessary. In Chapter 6, we addressed this problem by developing a method that is based on feature matching. Now the challenge is that of applying such methods to a sufficiently general class of models to produce a quantitative fit to the Milky Way.

7.2 Directions for future work

This thesis shows that while we understand the key ingredients of the large-scale gas flow in barred galaxies individually, a model that combines them to give a quantitative fit of the Milky Way is still missing. The most promising way to produce such a model is to apply a quantitative fitting method similar to that developed in Chapter 6. Without doubt, the method in Chapter 6 can be improved in many respects, for example in the choice of features, or in the choice of the distance between features. But the basic idea of quantitatively comparing features, widespread in the field of computer vision and behind many modern algorithms for face and fingerprinting recognition, is probably the correct approach.

In our models, many features in the (l, v) diagrams were not sufficiently prominent in our models, most notably the shocks that in our interpretation form the sides of the Binney et al. (1991) parallelogram and the vertical features. However, when gas is compressed at a shock, gas is converted from atomic to molecular form and this should make features corresponding to post-shocked gas more prominent in CO observations than our simulations suggest. To see whether this is really the case and overcome these shortcomings we need to follow the chemistry of the gas and to include proper modelling for the small-scale physics, the gas emissivity and the radiative transfer. A law for the conversion of atomic to molecular gas is also necessary to test whether the wiggle instability can explain the observed asymmetry of molecular gas within $|l| \leqslant 4^\circ$, as hypothesised in Chapter 3. Besides a law for the creation of molecular gas, a law for the inverse process that takes into account the finite lifetime of molecular clouds is probably needed. We expect molecular clouds created on one side of the bar at a shock not to survive enough to reach the other side, otherwise the mean density would be averaged out and the asymmetry would not be present. Thus, we expect molecular clouds created at a shock to survive for a fraction of the dynamical time, $t_{\rm clouds} \lesssim 20\,{\rm Myr}$, which is well in agreement with theoretical studies (e.g. Blitz & Shu, 1980).

A related problem is the physical nature of the wiggle instability. The analysis of Kim et al. (2014) needs

to be understood at a more fundamental level and complemented by a more physical understanding of the instability. Only when this is done can we confidently explain physical phenomena which are likely to be caused by the wiggle instability, such as the asymmetry of the molecular zone or the feathering of spiral arms.

The simulations presented in this thesis are very simple from a fundamental point of view. They are non-self gravitating, non-magnetised, non-viscous. We adopted a simple isothermal equation of state. They are only two-dimensional. We do not include any heating (e.g., due to stellar radiation) or cooling (e.g., due to collisions) in our simulations. The gravitational potential is externally imposed and time-independent, except for a solid body rotation. In retrospect, it is surprising to see the richness that this simple scheme has produced. Indeed, the first studies on the wiggle instability attributed it as caused to magnetic fields, because it was difficult to believe that a new type of instability could be hiding in too simple physics.

The real ISM is far more complex than our simple models assume. However, we believe that the complexity of the models should be increased only when required by a clear understanding of why simpler models are not enough, such as in our suggestion for inclusion of chemistry above. Without a sound physical understanding, the addition of complex physics will lead to obscure results that are difficult to interpret and do not constitute real progress.

Finally, we should mention that modelling the gas flow is only part of the bigger quest of modelling the whole Galaxy in all its aspects and of understanding how it works as a machine. Big efforts are underway to determine the current structure and history of this complex physical system. We are in the middle of an epoch of massive surveys of our Galaxy: for example, the recently launched *Gaia* satellite will provide us with extraordinarily accurate information on the position and motion of stars, as well as their ages and chemical compositions. Stellar and gas dynamics are complementary to each other: the equations of motion are different, and while gas observations are unaffected by dust, some additional information such as distances is available only for stars. Sophisticated dynamical models, necessary to make sense of the huge amount of data, are under development.

References

Binney J., Gerhard O. E., Stark A. A., Bally J., Uchida K. I., 1991, MNRAS, 252, 210
Blitz L., Shu F. H., 1980, ApJ, 238, 148
Kim W.-T., Kim Y., Kim J.-G., 2014, ApJ, 789, 68

APPENDIX A

Solving the Floquet Equation

In this Appendix we show how to solve Eq. (4.28). This equation arises during a Floquet analysis to find librations around closed orbits in a strongly barred potential.

A.1 Simplified case

Before attacking Eq. (4.28), let us first consider a simpler differential equation which already contains all the important ingredients:

$$\ddot{x} + \kappa(t)x + \lambda\dot{x} = F(t)\,, \tag{A.1}$$

where $\lambda > 0$ and $\kappa(t)$ and $F(t)$ are given periodic functions of time with period T:

$$\begin{aligned} \kappa(t) &= \kappa(t+T)\,, \\ F(t) &= F(t+T)\,. \end{aligned} \tag{A.2}$$

We want to solve this equation to find $x(t)$. If κ were constant and not time dependent, Eq. (A.1) would be the equation of a damped and driven harmonic oscillator. The general solution of this equation is a transient that decays exponentially plus a periodic term with the same periodicity as $F(t)$. Our problem is more general, and κ is a function of time. We are interested in the solution of Eq. (A.1) that is equivalent to the solution that discards transients. This solution is the one such that $x(t)$ is periodic with the same periodicity of F and k. This is also a requirement if we want to obtain a closed orbit, and we have to assume such periodicity if our phenomenological model is to make sense. Therefore let us assume that all functions are

periodic with period T and expand in Fourier series:

$$\begin{aligned} F(t) &= \sum_n F_n e^{in\omega t} \\ \kappa(t) &= \sum_n K_n e^{in\omega t} \\ x(t) &= \sum_n X_n e^{in\omega t}\,, \end{aligned} \tag{A.3}$$

where $\omega = 2\pi/T$ and n is an integer. Substituting Eqs. (A.3) into Eq. (A.1) we obtain

$$\begin{aligned} -\sum_n (n\omega)^2 X_n e^{in\omega t} + \left(\sum_m K_m e^{im\omega t}\right)\left(\sum_p X_p e^{ip\omega t}\right) + \lambda \sum_n X_n (in\omega) e^{in\omega t} \\ = \sum_n F_n e^{in\omega t}\,. \end{aligned} \tag{A.4}$$

Fortunately, the product of two Fourier series gives us another Fourier series of the same type:

$$\begin{aligned} \left(\sum_m K_m e^{im\omega t}\right)\left(\sum_p X_p e^{ip\omega t}\right) &= \sum_m \sum_p K_m X_p e^{i(m+p)\omega t} \\ &= \sum_m \sum_{n-m} K_m X_{n-m} e^{in\omega t} \\ &= \sum_m \sum_n K_m X_{n-m} e^{in\omega t} \\ &= \sum_n \left(\sum_m K_m X_{n-m}\right) e^{in\omega t} \\ &= \sum_n C_n e^{in\omega t}\,, \end{aligned} \tag{A.5}$$

where

$$\boxed{C_n = \sum_m K_m X_{n-m}}\,. \tag{A.6}$$

Substituting Eq. (A.5) into (A.4) we have

$$-\sum_n (n\omega)^2 X_n e^{in\omega t} + \sum_n C_n e^{in\omega t} + \lambda \sum_n X_n (in\omega) e^{in\omega t} = \sum_n F_n e^{in\omega t}\,. \tag{A.7}$$

In this last equation we can equate the coefficients term by term in the Fourier Series. We obtain

$$\boxed{-(n\omega)^2 X_n + \sum_m K_m X_{n-m} + \lambda X_n (in\omega) = F_n}\,. \tag{A.8}$$

In the boxed equation above, K_n and F_n are assumed to be known. The unknowns are X_n, and thus the boxed equation constitutes a set of linear algebraic equations that can be solved to find X_n. We want to

rewrite this system in the form

$$\mathbb{A}\mathbf{X} = \mathbf{F}\,, \tag{A.9}$$

where $\mathbb{A}$ is an infinite matrix and

$$\mathbf{X} = \begin{pmatrix} \vdots \\ X_{-1} \\ X_0 \\ X_1 \\ \vdots \end{pmatrix}\,. \tag{A.10}$$

The vector $\mathbf{F}$ is

$$\mathbf{F} = \begin{pmatrix} \vdots \\ F_{-1} \\ F_0 \\ F_1 \\ \vdots \end{pmatrix}\,. \tag{A.11}$$

The matrix $\mathbb{A}$ can be written as the sum of two matrices:

$$\mathbb{A} = \mathbb{D} + \mathbb{K}\,. \tag{A.12}$$

$\mathbb{D}$ is diagonal matrix whose elements are $-(n\omega)^2 + \lambda(in\omega)$

$$\mathbb{D} = \begin{pmatrix} \ddots & & & & & \\ & -(\omega)^2 - \lambda(i\omega) & & & & \\ & & 0 & & & \\ & & & -(\omega)^2 + \lambda(i\omega) & & \\ & & & & \ddots & \end{pmatrix}\,. \tag{A.13}$$

$\mathbb{K}$ is the following matrix

$$\mathbb{K} = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ \cdots & K_0 & K_1 & K_2 & K_3 & K_4 & \cdots \\ \cdots & K_{-1} & K_0 & K_1 & K_2 & K_3 & \cdots \\ \cdots & K_{-2} & K_{-1} & K_0 & K_1 & K_2 & \cdots \\ \cdots & K_{-3} & K_{-2} & K_{-1} & K_0 & K_1 & \cdots \\ \cdots & K_{-4} & K_{-3} & K_{-2} & K_{-1} & K_0 & \cdots \\ \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} . \tag{A.14}$$

On the diagonal of the matrix $\mathbb{K}$ there is K_0. Suppose we truncate the Fourier expansion of $\kappa(t)$ at K_0, so we recover the simple case of the normal damped and driven harmonic oscillator. Then the matrix $\mathbb{A}$ is diagonal and the system is very easy to solve line by line. Now suppose we truncate the Fourier expansion of $\kappa(t)$ at some higher order. In this case, the matrix $\mathbb{A}$ is not diagonal, but is a band matrix that has elements on the sides of the diagonal up to the order of truncation. For example, if we truncate at $K_{\pm 1}$, then we have a band of width 3. To find the solution we need to truncate the system at some finite order and in general we cannot solve the system exactly as in the simple case of the harmonic oscillator, where all equations were decoupled.

Thus there are two truncations involved. The first is where to truncate the Fourier expansions of κ and F. This truncation has to be done at some order where κ and F are well represented by their Fourier expansions. The second is the size of truncation of the system $\mathbb{A}$. This truncation must be done for a size sufficiently high that the solution is converged and does not change anymore by a further increase in the size of the system. We have found that this convergence takes place quite rapidly.

A.2 Full case

We have discussed how to solve a simplified equation. Let us now turn our attention to the real equation to be solved. This involves not a real function $x(t)$, but a vector $\mathbf{x}_1(t)$. The equation is

$$\ddot{\mathbf{x}}_1 = -\left[(\mathbf{x}_1 \cdot \nabla)\nabla\Phi_0 + \nabla\Phi_P\right]_{x_c(t)} + \Omega_\mathrm{p}^2 \mathbf{x}_1 - 2\Omega_\mathrm{p}\left(\hat{e}_z \times \dot{\mathbf{x}}_1\right) - 2\lambda\dot{\mathbf{x}} . \tag{A.15}$$

Expanding into components it is

$$\begin{aligned} \ddot{x}_1 &= -(x_1\partial_x + y_1\partial_y)\partial_x\Phi_0 - \partial_x\Phi_P + \Omega_\mathrm{p}^2 x_1 - 2\Omega_\mathrm{p}\dot{y}_1 - 2\lambda\dot{x}_1 , \\ \ddot{y}_1 &= -(x_1\partial_x + y_1\partial_y)\partial_y\Phi_0 - \partial_y\Phi_P + \Omega_\mathrm{p}^2 y_1 + 2\Omega_\mathrm{p}\dot{x}_1 - 2\lambda\dot{y}_1 . \end{aligned} \tag{A.16}$$

Rearranging

$$\begin{aligned}\ddot{x}_1 &= -x_1\partial_x^2\Phi_0 - y_1\partial_{xy}^2\Phi_0 - \partial_x\Phi_P + \Omega_\mathrm{p}^2 x_1 - 2\Omega_\mathrm{p}\dot{y}_1 - 2\lambda\dot{x}_1\,,\\ \ddot{y}_1 &= -x_1\partial_{xy}^2\Phi_0 - y_1\partial_y^2\Phi_0 - \partial_y\Phi_P + \Omega_\mathrm{p}^2 y_1 + 2\Omega_\mathrm{p}\dot{x}_1 - 2\lambda\dot{y}_1\,.\end{aligned} \tag{A.17}$$

Renaming

$$\begin{aligned}\ddot{x}_1 &= -x_1\kappa_{xx}(t) - y_1\kappa_{xy}(t) + F_x(t) + \Omega_\mathrm{p}^2 x_1 - 2\Omega_\mathrm{p}\dot{y}_1 - 2\lambda\dot{x}_1\,,\\ \ddot{y}_1 &= -x_1\kappa_{xy}(t) - y_1\kappa_{yy}(t) + F_y(t) + \Omega_\mathrm{p}^2 y_1 + 2\Omega_\mathrm{p}\dot{x}_1 - 2\lambda\dot{y}_1\,,\end{aligned} \tag{A.18}$$

where

$$\begin{aligned}\kappa_{xx}(t) &= \left[\partial_x^2\Phi_0\right]_{x_c(t)}\,,\\ \kappa_{xy}(t) &= \left[\partial_{xy}^2\Phi_0\right]_{x_c(t)}\,,\\ \kappa_{yy}(t) &= \left[\partial_y^2\Phi_0\right]_{x_c(t)}\,,\\ F_x(t) &= -\left[\partial_x\Phi_P\right]_{x_c(t)}\,,\\ F_y(t) &= -\left[\partial_y\Phi_P\right]_{x_c(t)}\,.\end{aligned} \tag{A.19}$$

Now expand all time dependent quantities as Fourier series, assuming they are periodic with period T

$$\begin{aligned}x_1(t) &= \sum_n X_n e^{in\omega t}\,,\\ y_1(t) &= \sum_n Y_n e^{in\omega t}\,,\\ \kappa_{xx}(t) &= \sum_n K_{xx;n} e^{in\omega t}\,,\\ \kappa_{xy}(t) &= \sum_n K_{xy;n} e^{in\omega t}\,,\\ \kappa_{yy}(t) &= \sum_n K_{yy;n} e^{in\omega t}\,,\\ F_x(t) &= \sum_n F_{x;n} e^{in\omega t}\,,\\ F_y(t) &= \sum_n F_{y;n} e^{in\omega t}\,.\end{aligned} \tag{A.20}$$

The unknowns are X_n, Y_n, while the other functions (κ's and F's) are assumed to be known. Substituting the Fourier expansions into Eqs. (A.18) and then equating coefficients term by term we arrive at

$$\begin{aligned}
&-\left[(n\omega)^2+\Omega_{\rm p}^2\right]X_n = \\
&-\sum_m K_{xx;m}X_{n-m}-\sum_m K_{xy;m}Y_{n-m}-(in\omega)\left[2\lambda X_n+2\Omega_{\rm p}Y_n\right]+F_{x;n}\,, \\
&-\left[(n\omega)^2+\Omega_{\rm p}^2\right]Y_n = \\
&-\sum_m K_{yy;m}Y_{n-m}-\sum_m K_{xy;m}X_{n-m}-(in\omega)\left[2\lambda Y_n-2\Omega_{\rm p}X_n\right]+F_{y;n}\,.
\end{aligned} \tag{A.21}$$

The last two equations are an algebraic system of equations in the unknown X_n and Y_n that we can solve. Again, we want to rewrite the system in the form

$$\mathbb{A}\mathbf{X}=\mathbf{F}\,. \tag{A.22}$$

We can choose how to order the unknowns inside $\mathbf{X}$. We choose

$$\mathbf{X}=\begin{pmatrix}\vdots\\ X_{-1}\\ Y_{-1}\\ X_0\\ Y_0\\ X_1\\ Y_1\\ \vdots\end{pmatrix}\,. \tag{A.23}$$

Then $\mathbf{F}$ is

$$\mathbf{F}=\begin{pmatrix}\vdots\\ F_{x;-1}\\ F_{y;-1}\\ F_{x;0}\\ F_{y;0}\\ F_{x;1}\\ F_{y;1}\\ \vdots\end{pmatrix}\,. \tag{A.24}$$

To write $\mathbb{A}$, let us divide it as before into a part $\mathbb{K}$ containing the Fourier coefficients of the $\kappa's$ and a part $\mathbb{D}$ that does not.

$$\mathbb{A} = \mathbb{D} + \mathbb{K}\,. \tag{A.25}$$

$\mathbb{D}$ is now a block-diagonal rather than a diagonal matrix: it has 2×2 blocks along the diagonal. Each block is

$$\mathbb{D}_n = \begin{pmatrix} -(n\omega)^2 + 2\lambda(in\omega) + \Omega_{\rm p} & (in\omega)2\Omega_{\rm p} \\ -(in\omega)2\Omega_{\rm p} & -(n\omega)^2 + 2\lambda(in\omega) + \Omega_{\rm p} \end{pmatrix}. \tag{A.26}$$

This is the block referring to the vector (X_n, Y_n). The entire matrix $\mathbb{D}$ will then be

$$\mathbb{D} = \begin{pmatrix} \ddots & & & & & \\ & \mathbb{D}_{-1} & & & & \\ & & \mathbb{D}_0 & & & \\ & & & \mathbb{D}_1 & & \\ & & & & \ddots \end{pmatrix}. \tag{A.27}$$

The matrix $\mathbb{K}$ will be

$$\mathbb{K} = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ \cdots & \mathbb{K}_0 & \mathbb{K}_1 & \mathbb{K}_2 & \mathbb{K}_3 & \mathbb{K}_4 & \cdots \\ \cdots & \mathbb{K}_{-1} & \mathbb{K}_0 & \mathbb{K}_1 & \mathbb{K}_2 & \mathbb{K}_3 & \cdots \\ \cdots & \mathbb{K}_{-2} & \mathbb{K}_{-1} & \mathbb{K}_0 & \mathbb{K}_1 & \mathbb{K}_2 & \cdots \\ \cdots & \mathbb{K}_{-3} & \mathbb{K}_{-2} & \mathbb{K}_{-1} & \mathbb{K}_0 & \mathbb{K}_1 & \cdots \\ \cdots & \mathbb{K}_{-4} & \mathbb{K}_{-3} & \mathbb{K}_{-2} & \mathbb{K}_{-1} & \mathbb{K}_0 & \cdots \\ \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{A.28}$$

where each $\mathbb{K}_n$ is a 2×2 block given by

$$\mathbb{K}_n = \begin{pmatrix} K_{xx;n} & K_{xy;n} \\ K_{xy;n} & K_{yy;n} \end{pmatrix}. \tag{A.29}$$

The linear algebraic system of equations can be solved by truncating as before.

APPENDIX B

Results of the simulation of Chapter 5 for other values of the pattern speed

In this appendix, we show the results for values of the pattern speed that were not included in Chapter 5. While these were not strictly necessary to understand the chapter, they constitute a useful reference.

APPENDIX C

Tests with self-gravity

In this appendix we show the results of few simulations that include self-gravity. The methods and the potential are the same as in Chapter 3. The sound speed is assumed to be $c_{\rm s} = 10\,{\rm km\,s^{-1}}$ and the spatial resolution is ${\rm d}x = 10\,{\rm pc}$. We have disabled the recycling law ($\alpha = 0$) to isolate more clearly the effects of self-gravity. Fig. C.1 shows the gas density for different simulations at the same evolutionary time long after the bar is fully grown. The parameter γ quantifies the strength of self-gravity by defining how much mass should be attributed to each unit of gas mass in the simulation. The self-gravity of the two-dimensional gas

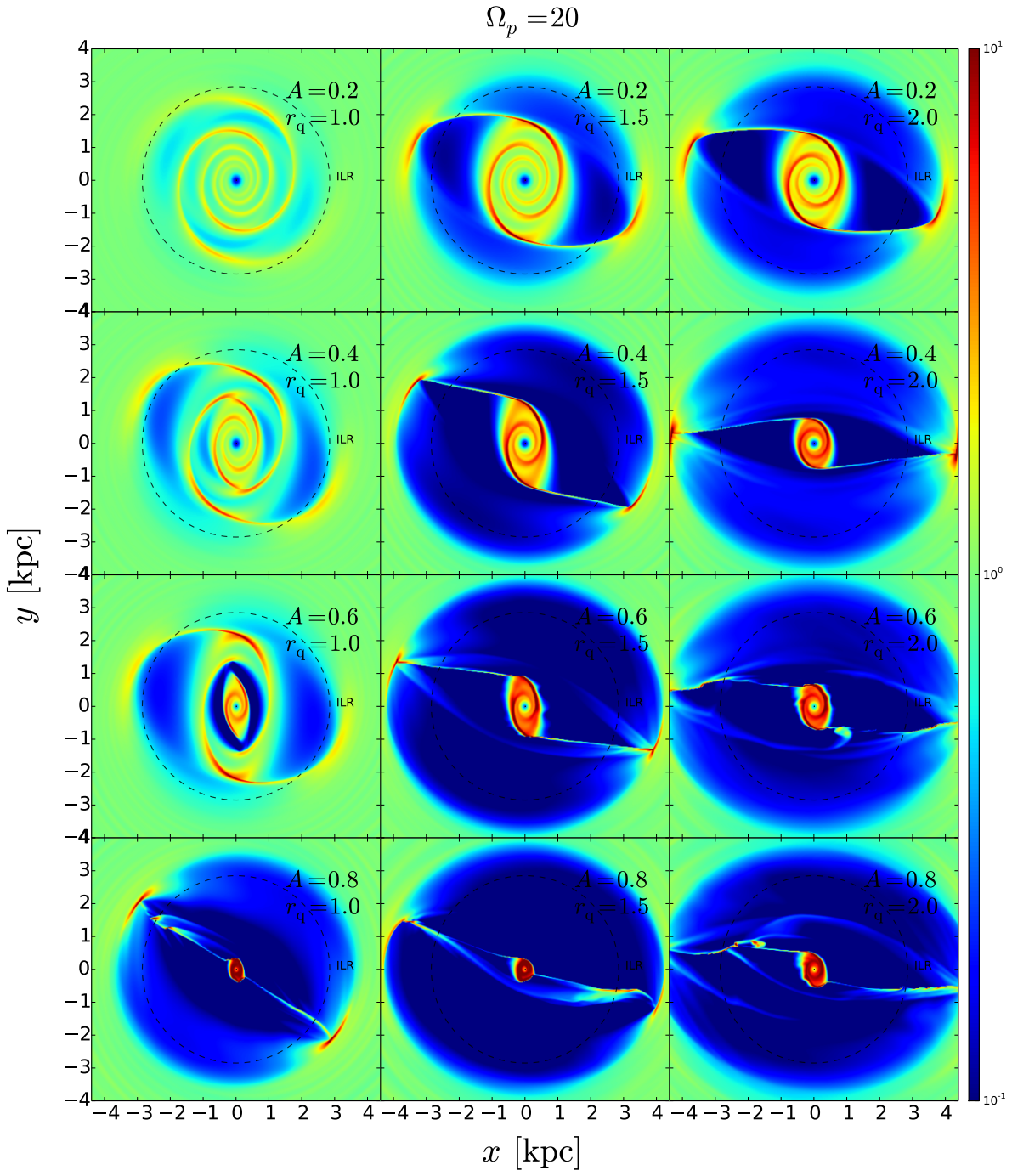


Figure B.1: Same as Fig. 5.6 but for a pattern speed $\Omega_\mathrm{p} = 20\ \mathrm{km\,s^{-1}\,kpc^{-1}}$.

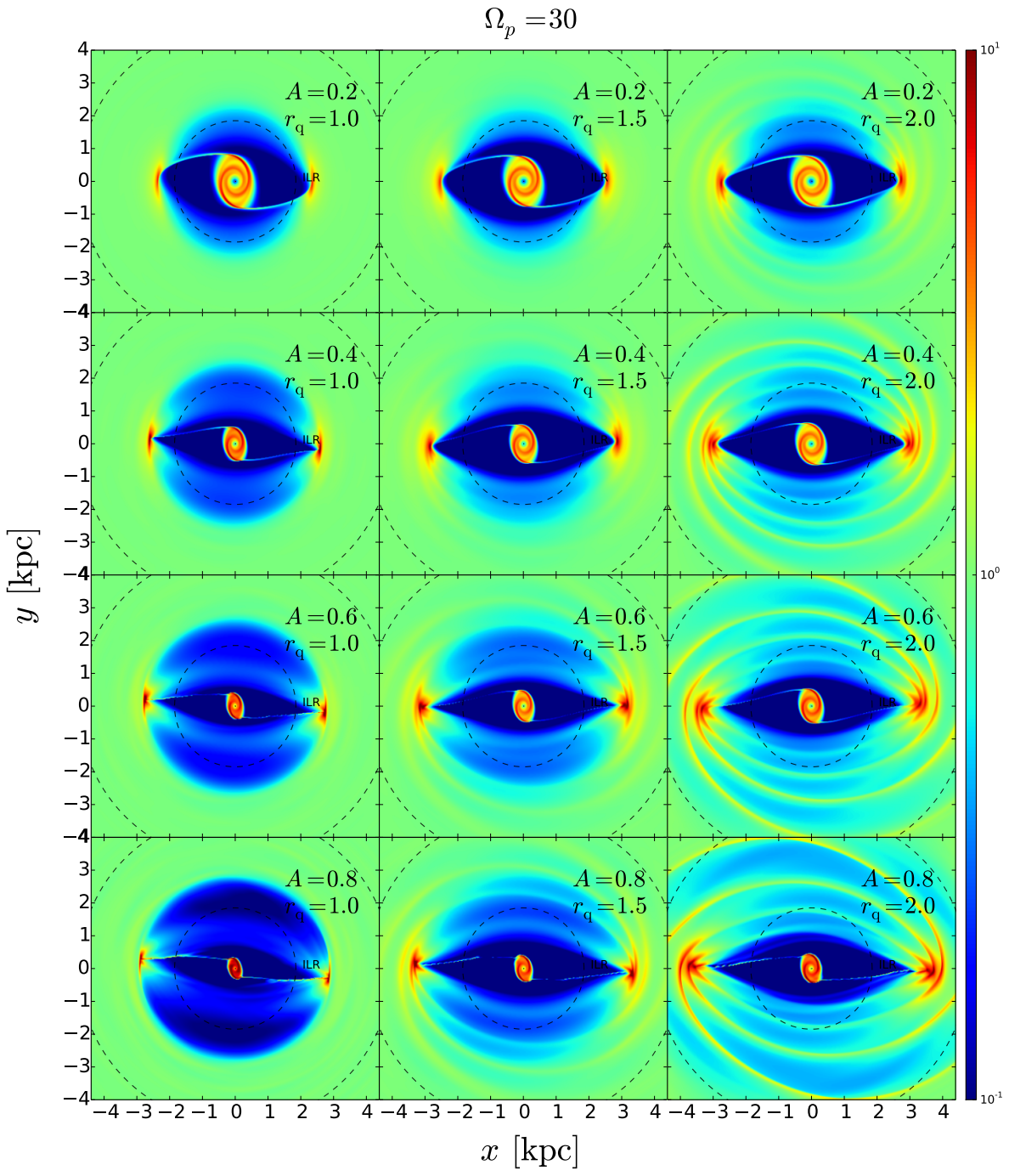


Figure B.2: Same as Fig. 5.6 but for a pattern speed $\Omega_\mathrm{p} = 30\ \mathrm{km\,s^{-1}\,kpc^{-1}}$.

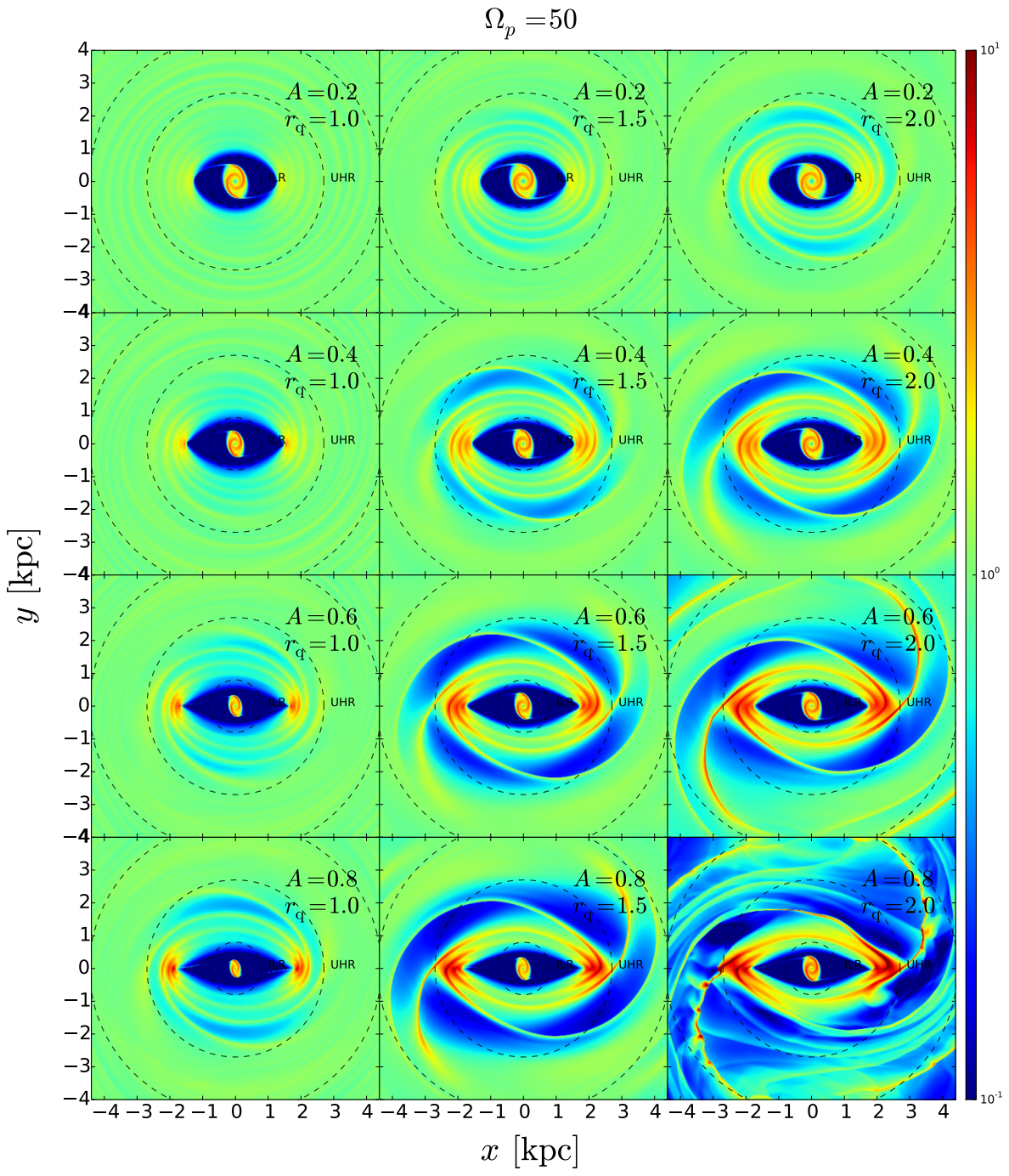


Figure B.3: Same as Fig. 5.6 but for a pattern speed $\Omega_\mathrm{p} = 50\ \mathrm{km\,s^{-1}\,kpc^{-1}}$.

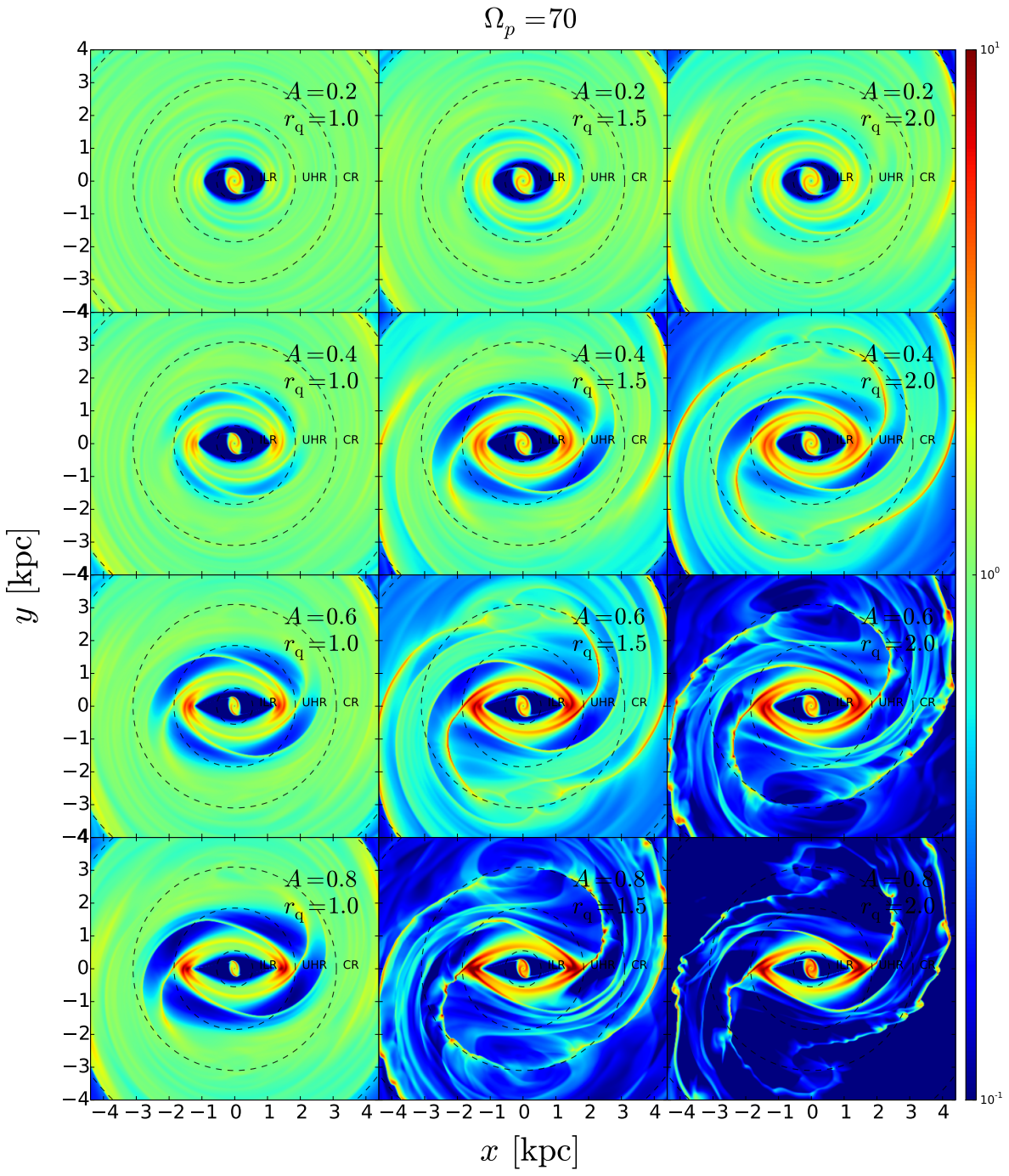


Figure B.4: Same as Fig. 5.6 but for a pattern speed $\Omega_\mathrm{p} = 70\ \mathrm{km\,s^{-1}\,kpc^{-1}}$.

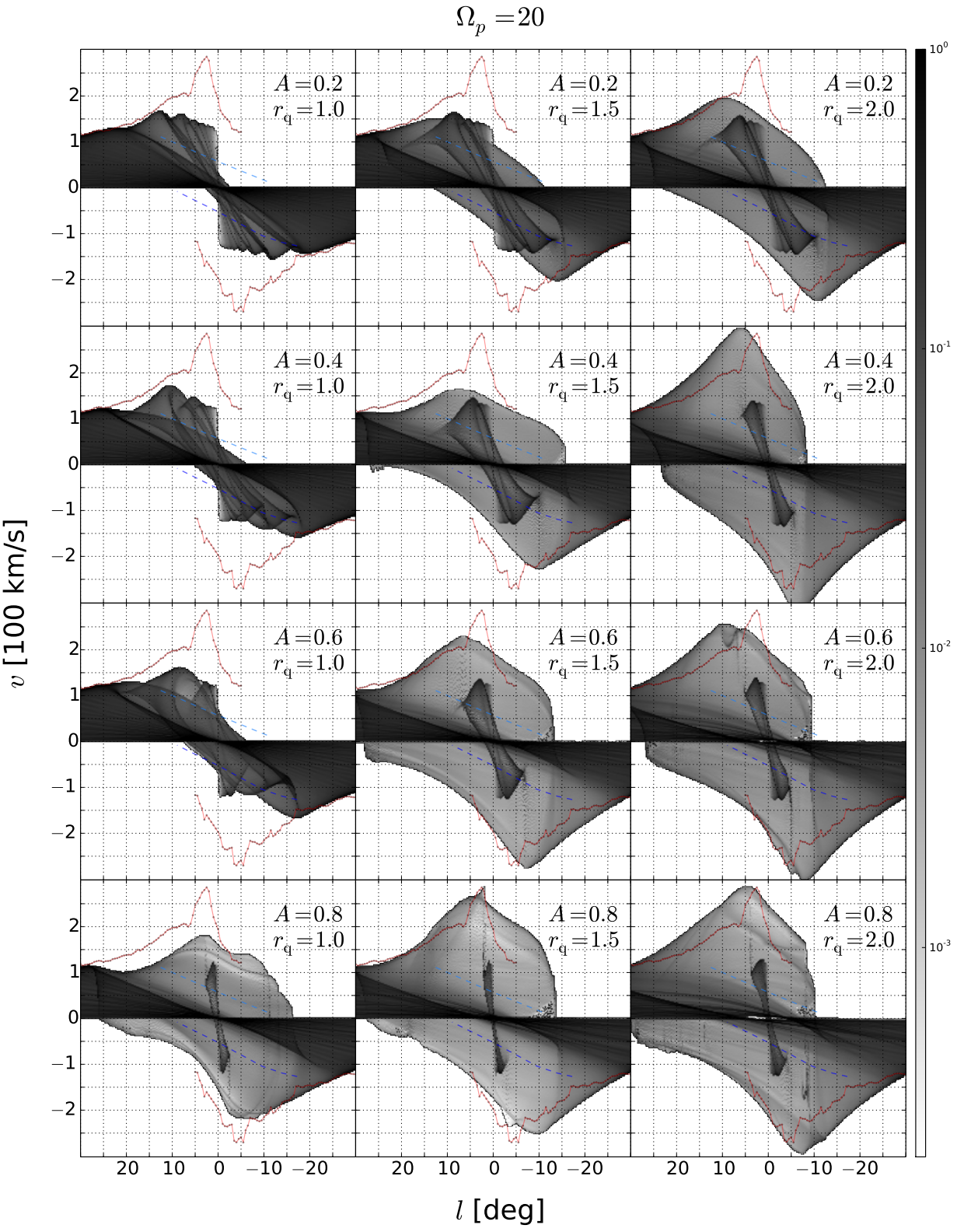


Figure B.5: Same as Fig. 5.7 but referring to Fig. B.1.

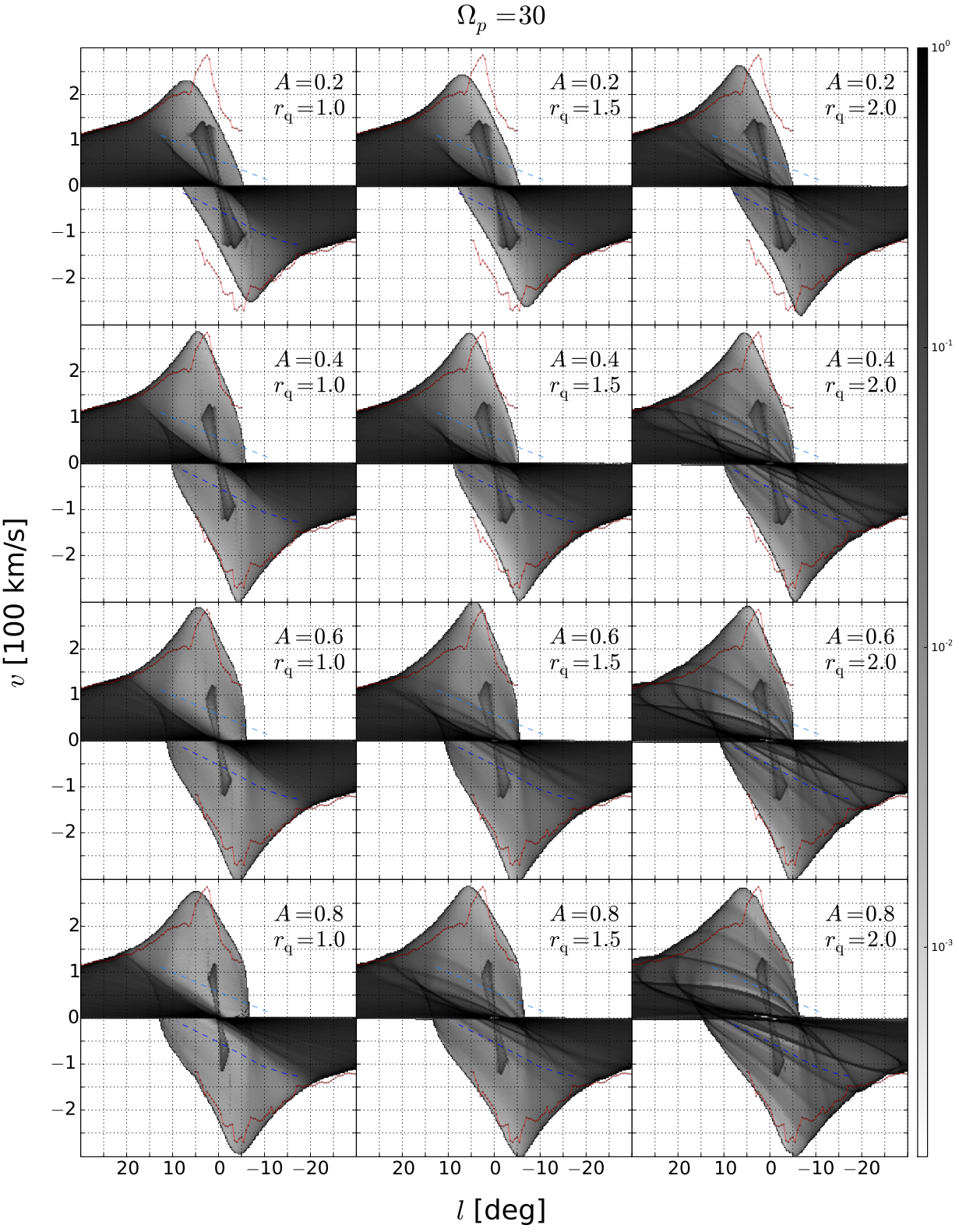


Figure B.6: Same as Fig. 5.7 but referring to Fig. B.2.

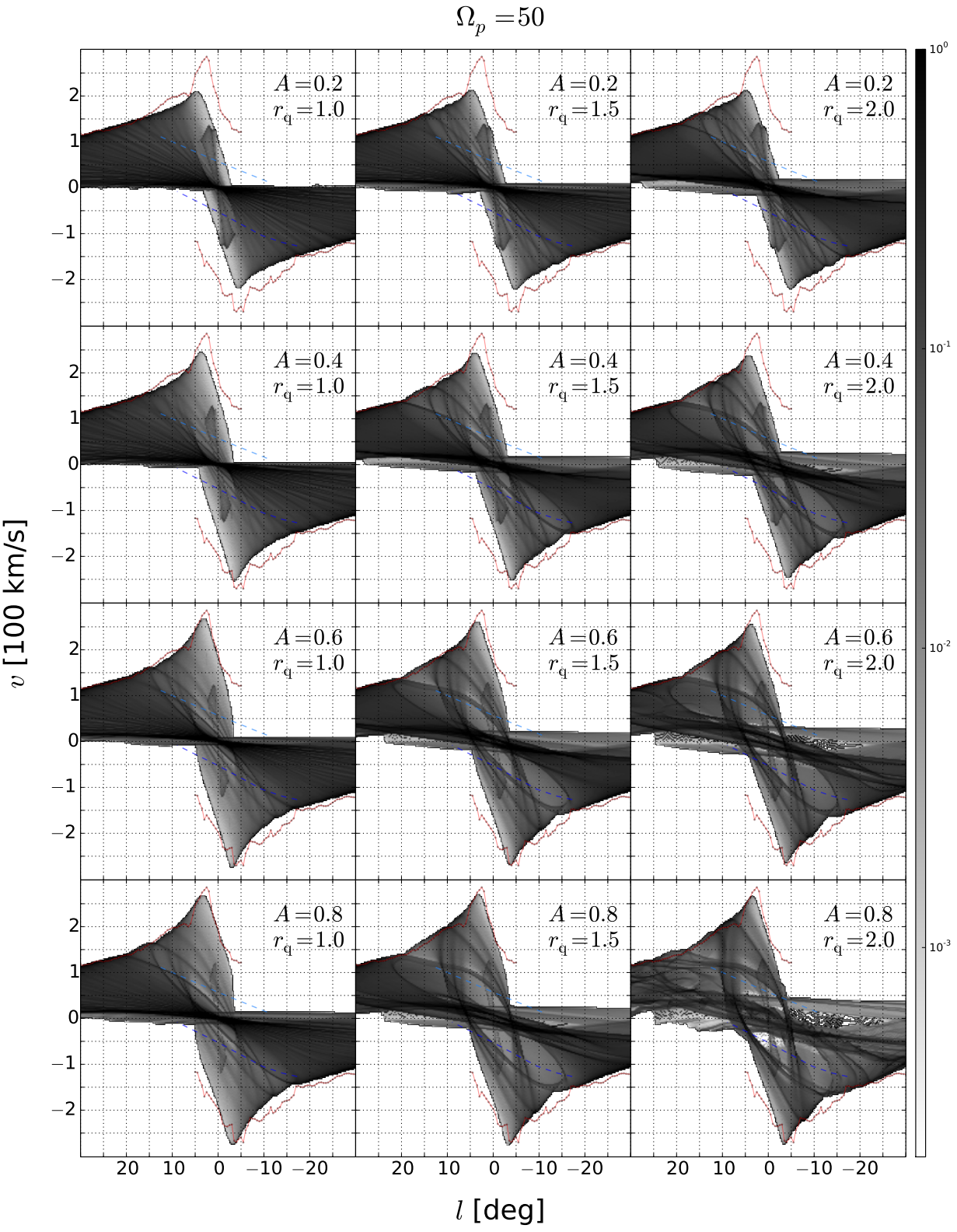


Figure B.7: Same as Fig. 5.7 but referring to Fig. B.3.

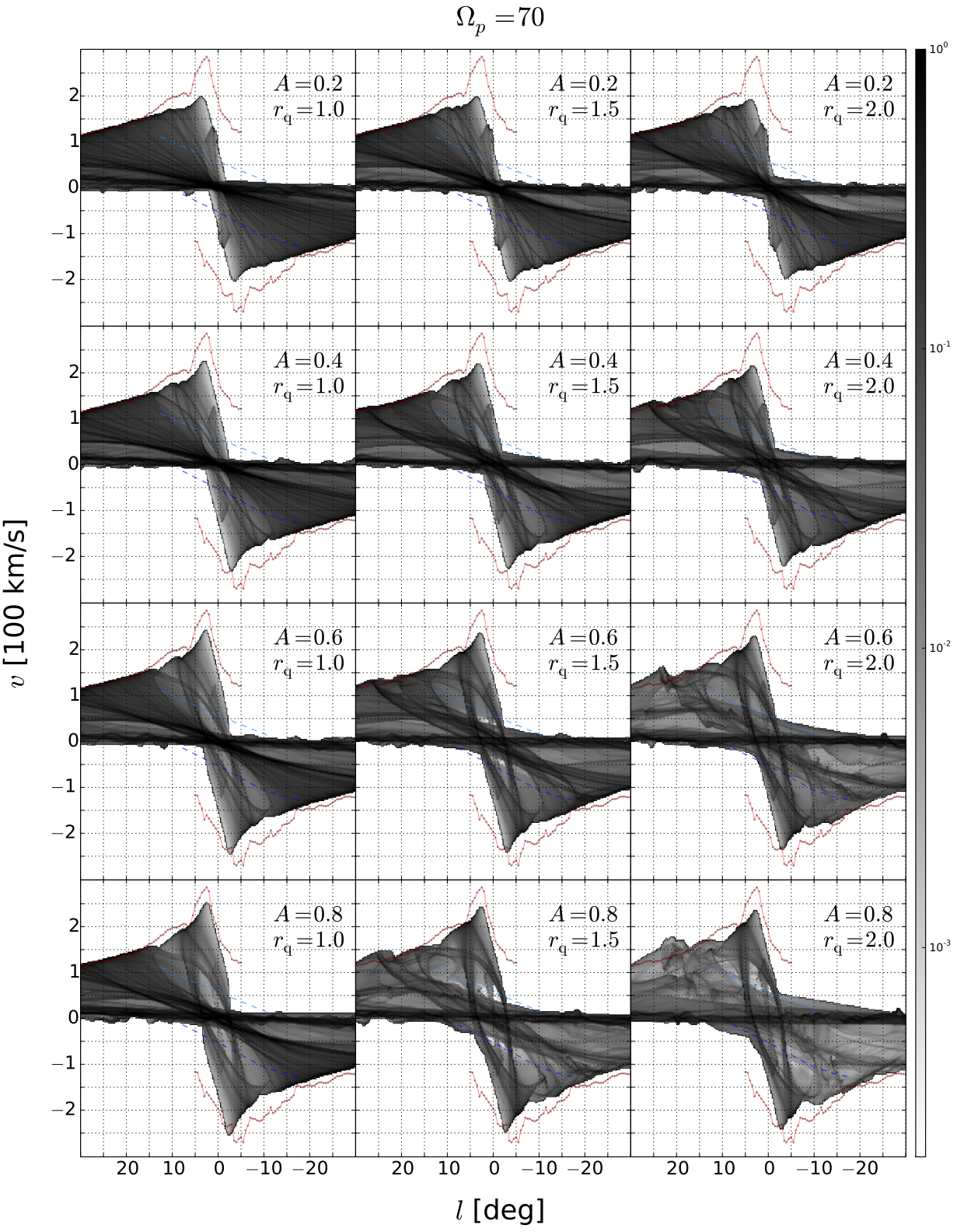


Figure B.8: Same as Fig. 5.7 but referring to Fig. B.3.

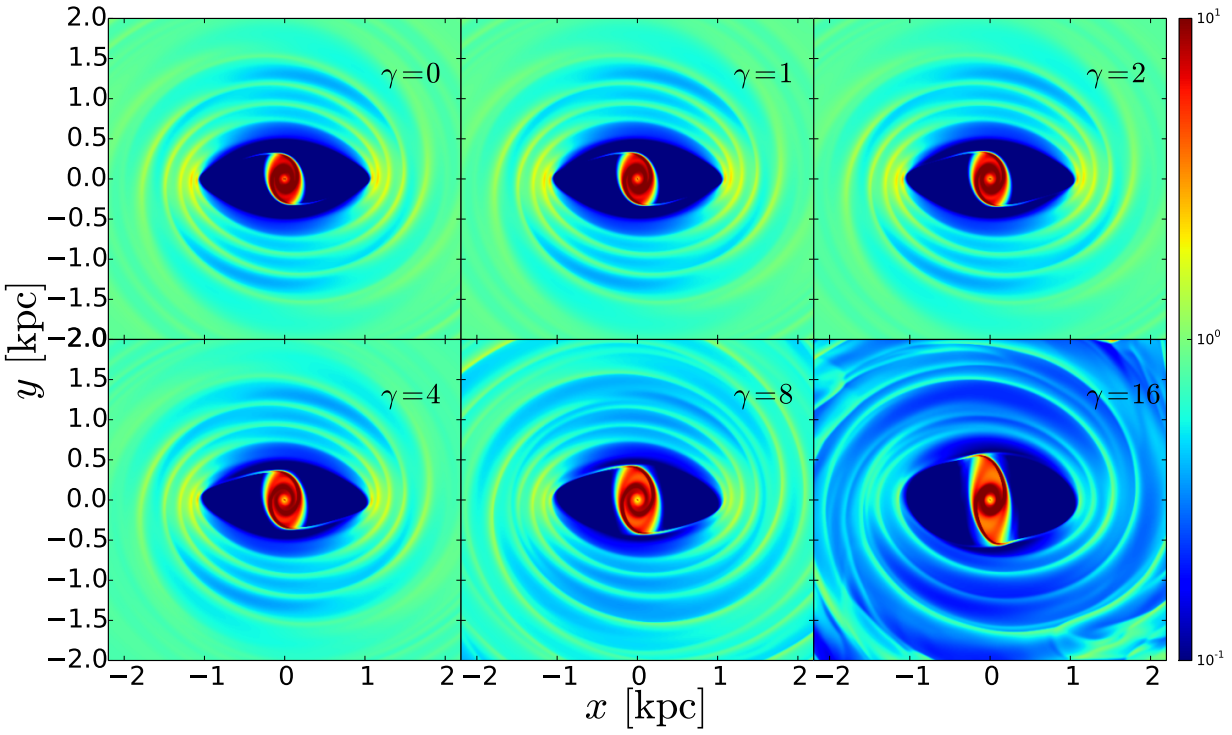


Figure C.1: Gas surface density in simulations that include the gas self-gravity in addition to the potential used Chapter 3. The parameters γ characterises the strength of self-gravity.

is calculated as if its surface density was given by the colorbar on the right in units of

$$\gamma\, M_\odot / \,\mathrm{pc}^2\,. \tag{C.1}$$

Thus, when $\gamma = 0$ self-gravity is disabled. The initial density in each simulations is assumed to be uniform at 1 in these units. The results show that for realistic values of gas density, $\gamma \lesssim 4$ (e.g. Binney & Merrifield, 1998) the effect of the self-gravity is very small on the gas flow, while it becomes important only for values of the self-gravity implausibly high. These simulations only refer to the large-scale effects of self-gravity; on smaller scales out of the scope of the present simulations it is known that self-gravity can be important in maintaining the structure of individual ISM clouds.

References

Binney J., Merrifield M., 1998, Galactic Astronomy. Princeton University Press

APPENDIX D

A potential used to test the methods of Chapter 6

Here are the details of the potential used in the generation of the mock data in Section 6.5.3. The potential is inspired by the models of Dehnen & Binney (1998). It is made by 3 components: bar, disk, halo. Table D.1 shows the value of the parameters used. The potential is steady in a frame that rotates with pattern speed $\Omega_{\rm p} = 48\ {\rm km\,s^{-1}\,kpc^{-1}}$.

- **Bar** The density distribution generating the potential of the bar is given by

$$\rho(a) = \rho_0 \left(\frac{a}{a_0}\right)^{-\alpha} \exp\left(-a^2/a_0^2\right)\ , \tag{D.1}$$

  where

$$a = \sqrt{x^2 + (y^2 + z^2)/q^2}.$$

  To fully specify the bar potential we therefore need 4 parameters: the central concentration ρ_0, inner slope α, major axis a_0 and axis ratio q. Equivalently, we can specify the total mass M instead of the central concentration ρ_0.

- **Disk** The density distribution of the disk is exponential. It has zero thickness, and the surface mass density is given by

$$\Sigma(R) = \Sigma_0 e^{-R/R_{\rm d}}\,. \tag{D.2}$$

  To fully specify the disk potential we need 2 parameters: the radius $R_{\rm d}$ and the central surface mass density Σ_0.

- **Halo** The halo potential is logarithmic,

$$\Phi_{\rm halo}(r) = \frac{1}{2} v_0^2 \log(r_{\rm h}^2 + r^2)\,. \tag{D.3}$$

Bar	$M = 0.5 \times 10^{10} M_\odot$	$a_0 = 2.5$ kpc	$\alpha = 1.8$	$q = 0.4$
Disk	$\Sigma_0 = 0.07 \times 10^{10} M_\odot\,\mathrm{kpc}^{-2}$	$R_\mathrm{d} = 2.5$ kpc		
Halo	$v_0 = 185\ \mathrm{km\,s}^{-1}$	$r_h = 5.0$ kpc		
Ω_p	$48\ \mathrm{km\,s}^{-1}\,\mathrm{kpc}^{-1}$			
ϕ	35°			

Table D.1: The parameters for the model used for the test in Sect. 6.5.3.

To fully specify the halo potential we need 2 parameters: the radius r_h and the circular velocity at infinity v_0.

References

Dehnen W., Binney J., 1998, MNRAS, 294, 429

APPENDIX E

Earth Mover Distance

The earth-mover distance (EMD) is a way of quantifying the dissimilarity of two distributions. Intuitively, given two distributions, one can be seen as a collection of piles of earth spread in space, the other as a collection of of holes in the same space. The amount of earth at each point can be any positive real number. The EMD measures the minimal amount of work needed to fill in the holes with earth taken from the piles. A unit of work corresponds to transporting a unit of earth by a unit of ground distance, which in our case would be a metric suitably defined in the (l, v) plane. The earth contained in one pile can be shared among many different holes if this solution requires less work than other alternatives. More details on this distance can be found for example in Rubner et al. (2000).

As noted in Section 6.4.2, EMD is an option that we initially found intuitively appealing for comparing features. In this case, the idea is to apply the EMD to binary images such as panel (h) in Fig. 6.4. The amount of earth is 1 at pixels corresponding to features, and zero otherwise. The dissimilarity between model features and data features is quantified by the minimal amount of work needed to turn the model features into data features (or vice versa). When used in this way, the EMD turned out to underperform the

much simpler SMHD; if anything, the EMD was actually too clever in matching features from one image to the other, with the result that the variation of EMD with ϕ and $\Omega_{\rm p}$ had spikes and false minima, and was much noisier than either the SMHD or envelope distances.

Nevertheless, if one were faced with carrying out simulations aimed at reproducing the detailed intensities at each point in the (l, v) diagrams, the EMD used in a qualitatively different way might be reconsidered as an alternative to χ^2. The idea in this case would be to use the EMD to compare two full (l, v) distributions, such as panel (a) in Fig. 6.4. The amount of earth would then be the brightness temperature at each pixel, which would *not* be constrained to be either 0 or 1. The potential advantage of EMD used in this way would be that it avoids one of χ^2 main problems, namely that of ignoring cross-bin information. We therefore suspect that it might be useful in cases where one is trying to fit the details of the chemistry and small scales physics as well as the potential. In exploratory tests we found that EMD applied to the full distributions performed well in retrieving parameters when the rule (6.1) used to project the models was identical to that used to generate the mock data. As the EMD used in this way is strongly dependent on the intensities at each point, it *requires* that a model able to reproduced the detailed (l, v) intensities for the species under consideration, however.

References

Rubner Y., Tomasi C., Guibas L. J., Guibas L. J., 2000, International Journal of Computer Vision, 99