P   :   -

by

Lawrence Russell Mudryk

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy,
Graduate Department of Astronomy and Astrophysics,
in the University of Toronto

Abstract

Planetary system evolution: planet-disk interactions and planet ejection from binary systems

Lawrence Russell Mudryk

Doctor of Philosophy

Graduate Department of Astronomy and Astrophysics

University of Toronto

2007

Throughout the evolution of a planetary system, planets, especially those newly formed, interact by several means with a variety of the system's constituents. In particular, the influence of the most massive planets is expected to govern much of the long-term evolution of the system. In early stages of this evolution, the gas disk that provided the material from which the planets formed also acts to couple the planets to its own dynamics. In part I of this thesis, I describe a new hydrodynamic code that I have developed, tuned to study these interactions. Using this code, I explore the formation of hydrodynamic structures within the disk, such as jets and eddies, that arise from the influence of the planets on the overall flow. I show that while the formation of vortices is damped in disks with a large enough viscosity, jet formation is more robust in this sense and jet structures form even in viscous flows. I further propose that these jets may affect the amount of material transport that occurs in the flow in a manner similar to that found in the Earth's atmosphere and in the weather layers of the Jovian planets. In order to qualify this claim, I perform preliminary numerical experiments that aim to establish this relationship.

Even after the removal of the gas disk, the gravitational influence of massive planets—or stellar companions in the case of multiple systems—severely limits the range of stable orbits of the system's lesser planets. In part II of this thesis, I examine the physical mechanisms responsible for planet ejection from unstable orbits previously observed in numerical experiments. I determine the instability is due to overlap of subresonances lying within mean-motion resonances.

Acknowledgements

I want to thank my supervisor, Norman Murray: first, for suggesting a research topic that managed to hold my interest even at the times when I hated it and second, for continuing to fund this research during the last two years of its development. I appreciated his clear and discerning physical insight.

There are many people at the University of Toronto, without whom I could not have finished this work. The person I put first on this list is Jonathan Dursi. His wry optimism, tireless ego-boosting, and pointed, useful suggestions guided me out of a morass of numerics, in which I would otherwise still be ensnared. I am also grateful to Yanqin Wu, for suggesting a little project which became the second part of this thesis, for being a member of my thesis committee in all but name, and mostly for allowing me to see how engaging research could be at a time when I really needed it. Chris Matzner, I thank for his encouragement of my abilities, for being there to bounce ideas off and for editing many rough drafts of thesis chapters. I also want to thank my external examiner, Phil Armitage, for providing a careful read of my thesis as well as constructive comments on the work.

I thank Marc Goodman for making all of the university's bureaucratic details as pleasant as possible and Hugh Merz and Chris Loken, for fulfilling countless computing requests.

There are many people in my life who supported this thesis on a more personal level, by supporting me throughout the undertaking. I am grateful to all of these people, in particular:

my parents and my sister, for supporting me and continuing to support me into the future,

my fellow students, for making the moments in between work more bearable, especially
Carrie, Kris, Kevin, and Preethi for Star Trek, cards, lunches, punches, movies and more,
Zino and Kaitlin, for reminding me to take breaks from working on my mind, to work on my body from time to time,
my darling, Silvia, for her generous and demanding spirit that captures the hearts of everyone,

my friends outside of the department and from back home, especially Tyler, Alby, Mark, and my Ma,

Rosemary, for holding my hand through the first steps,

and Bob, for holding my hand through everything after.

Contents

List of Tables

List of Figures

Preface

The first extrasolar planet orbiting a main sequence star, 51 Pegasi, was announced by Mayor & Queloz (1995). As of this writing, 248 extrasolar planets in 202 planetary systems have been discovered (http://vo.obspm.fr/exoplanetes/encyclo/encycl.html). The most striking contrast between these gas giants and those of our own solar system is that many of them have exceedingly short-period orbits (down to a day) while others have relatively high eccentricities. Such knowledge has forced the original theories of solar system formation and evolution, based solely on the planets within our own solar system, to be rewritten, and the task of developing new mechanisms and theories that can explain the observed differences has occupied theorists for the past decade.

Planet formation can be considered the late stage of star formation, which begins millions of years earlier. According to standard theory, the formation of a solar nebula starts with diffuse material in molecular clouds. By processes still debated, overdense clumps of gas within the molecular cloud form; within such a clump there may be several centrally concentrated regions of gas which are infalling. These centrally concentrated regions will eventually become the protostars of loosely grouped stellar systems. While the infalling gas is able to dissipate energy, it cannot easily do the same for its angular momentum, especially at longer distances from the central concentration. A flattened disk of material forms, and infalling material accretes towards the central protostar through this gas disk. At this stage of formation, the system is still enshrouded by envelopes of dusty gas, however the existence of such disks has been observationally inferred around young low-mass stellar objects from their infrared excesses (Adams et al., 1987). Direct images of disks have also been obtained at later stages of system development (Bally et al., 2000) after most of the surrounding gas and dust has been removed. The process by which planets form in this enshrouded early solar system are currently debated and observations can currently provide no direct evidence for a particular theory.

Bodenheimer et al. (2000) argues that it is very unlikely that planets with short-period orbits could have formed where they presently lie. Gravitational fragmentation has been proposed as a mechanism to form planets within a few dynamical times if the disk gas is cold and dense enough, but reasonable disk parameters suggest that disks are very stable and not prone to self-gravitating instabilities except at large radii (>100 AU). Furthermore, the dynamics of such processes tend to produce only massive objects (>$10M_J$). Alternatively (see for example, Lissauer (1993)) one can form planets through the conglomeration of small dust particles, a process which yields planetesimals, up to a kilometer in size over $0.1 - 1$Myr. Once

this size, the planetesimals' gravity will cause them to loosely group together, reaching larger sizes up to a few tens of an Earth mass. By this final stage, the protoplanet is massive enough that runaway gas accretion occurs, leading to a gas giant with a mass of order $10 - 1000 M_\oplus$. This so-called core accretion process does not favor the formation of planets at small distances from the parent star either. The temperature must be cool enough at the formation location for the dust grains to condense out. For many observed extra-solar systems this location lies well outside the current location of the protoplanets. Furthermore, even assuming the formation of planetesimals was possible at small radii, most reasonable distributions of gas density do not have enough material at the inner annuli of a disk to build up Jupiter-mass planets once runaway accretion starts. By contrast, the total mass available per logarithmic radius increases further out in the disk.

Based on such arguments, most current scenarios posit formation of protoplanets at $5 - 20$ AU from the star, and expect that in a given system these planets will migrate inwards or outwards to different radii. For this work it is assumed that reasonable mechanisms exist to form protoplanets at these radii without specific consideration of their formation process.

Regardless of the formation mechanism, one or more giant planets existing within the gas disk will exert a marked influence on the subsequent evolution of the disk. A planet's gravity launches spiral density waves in the disk (Goldreich & Tremaine, 1979), leading inwards and trailing outwards from the planet's location. These spiral density waves can interact with the disk and subsequently back on the planet itself. They are physical manifestations of torques between the planet and the disk which can cause the planet to migrate (usually inwards towards the central star) or alter its eccentricity. In Part I of this thesis, I study various effects arising from the planet's couple to the gas disk using computational fluid dynamics to model such interactions.

During the late stages of solar system evolution, the final mass, orbital radius, and eccentricity of any gas giants within a solar system will affect the subsequent stability of any terrestrial planets within the system, the habitability of such planets, and the potential development of life on such planets. Within our own solar system, Jupiter positively affects the habitability of Earth as it helps to remove potential Earth-crossing comets. Such an influence need not be positive. In the extreme case where a second star rather than a gas giant exists within the system, the presence of the companion star will severely limit the range of stable orbital radii for any terrestrial planets. The possibility of a second star is important to consider as roughly 60% of the stars in our local neighborhood are in binary or higher-multiple systems (Duquennoy & Mayor, 1991). The planet formation process inside these systems could differ markedly from that around single stars, due to the dynamical influence of the second star. Nonetheless, the *in-situ* stability of previously formed planets is a starting point for examining the stability of planetesimals and gas existing during the process of formation. A planet orbiting around a star in a binary system will experience forces from both the parent star and the companion star, and it may be dislodged from its host star as a results of these dual forces. In part II of this thesis, I examine the physical mechanisms for such planet ejection previously seen in numerical experiments.

Part I

Protoplanet-Disk Interactions

Chapter 1

Introduction

The presence of a Jupiter-mass protoplanet within a protoplanetary disk has definite influences on the subsequent evolution of the disk and on any additional planets within that system. Such planets grossly affect the distribution of important fluid quantities such as mass, momentum and vorticity within the disk in a both non-linear and non-local manner.

Theoretical work on early solar-system evolution has focused on the interactions of a single protoplanet and a hydrodynamic disk. Authors such as Goldreich & Tremaine (1979, 1980); Ward (1986, 1996, 1997); Artymowicz (1993) and Tanaka et al. (2002) established that tidal interactions between a protoplanet and the disk generate spiral density waves as a manifestation of the torque exchange between the planet and the disk, and that in the linear regime, the net result of the torque exchange is to migrate the planet inward. This phenomenon is the leading explanation for the short-period orbits observed in extrasolar planetary systems.

Scenarios examining the behavior of a single protoplanet have been modelled computationally by Kley (1999); Bryden et al. (1999); Nelson et al. (2000). Their simulations capture the excitation of spiral density waves which form within the disks over a few dynamical times. The deposition of angular momentum from the planet near its orbit redistributes the surrounding gas in a manner which tends to clear a gap in the density profile. There is a competing torque due to the disk viscosity which acts to smooth out any density perturbations. Then a necessary condition for gap formation is that the planet is massive enough such that the tidal torque is greater than the viscous torque. To further ensure that the angular momentum transported by the planet is deposited in the planet's vicinity further requires that the Hill radius of the planet be on the order of the disk's scale height. This second condition enables tidal disturbances caused by the planet to induce local nonlinear shock dissipation. Simulations (de Val-Borro et al., 2006) indicate that gaps are just able to form for Neptune-sized planets in disks with viscosity on the low end of what observations suggest exist in a planetary disks. Disks with correspondingly larger viscosity require a more massive planet in order to form a gap.

The combined influence of multiple protoplanets in a disk is a less well-explored scenario. The effect that two protoplanets have on gap-clearing was examined briefly by Kley (2000), where he found that once gaps have formed around each respective planet, the remaining inter-orbital annulus of gas between the planets

is also cleared within a few hundred orbits. Although the interpretation of Kley's results was complicated by relative migration of the planets, a mechanism for the inter-orbital gas clearing was not proposed. In this thesis (Ch. 5), I present results which modify Kley's conclusions and propose an explanation for these new findings. These findings and additional studies (Crida & Morbidelli, 2007) suggest the important role that viscosity plays in the evolution of systems, depending on its magnitude and the details of its physical form (and implementation). The results of Crida et al. (2007) are based on a new code which embeds a one-dimensional viscous spreading model of the disk, covering a much larger domain than the two-dimensional grid surrounding the region of interest. The addition of the viscous solution accounts for the global evolution of the disk more consistently than otherwise, and the authors report instances where the typical inward migration of the protoplanets is often slowed and may even be even reversed.

Several proposed mechanisms exist to account for the observed eccentricities of extrasolar planets. Simulations including multiple planets (Kley, 2000) produce eccentricity growth when differential migration of multiple planets within such a system causes the planets to become locked within a strong mean-motion resonance (such as the 2:1 resonance). In such cases the eccentricity excitation comes from the resonance dynamics of the two planets. Using this mechanism to explain the high eccentricities of many extrasolar planets would require a second planet to have existed in each system and in many cases, to have subsequently been ejected. Goldreich & Sari (2003) and Ogilvie & Lubow (2003) proposed an alternative mechanism for exciting the eccentricity of a protoplanet via interactions with the gas disk, rather than possible additional planets within the system. Masset & Ogilvie (2004) and D'Angelo et al. (2006) computationally tested this scenario, and their results support the theoretical conjecture as a possible excitation mechanism. Nonetheless, Juric & Tremaine (2007) recently advanced the idea that the dynamical relaxation of a system of multiple planets over 10^7–10^8 yrs could instead explain the observed distribution of eccentricities.

Finally, a string of recent work has examined the distributions and production of potential vorticity within the disk based on theoretical work by Papaloizou & Lin (1989); Sheehan et al. (1999); Lovelace et al. (1999); Godon & Livio (2000) and Li et al. (2000). Li et al. (2001) and Klahr & Bodenheimer (2003) studied the generation of Rossby wave instabilities and vortices in baroclinic disks through forced turbulence. de Val-Borro et al. (2007) and Ou et al. (2007) considered the growth of vortices due to such effects when there is an embedded protoplanet.

Potential vorticity is an important quantity to consider. In two-dimensional barotropic flows, the quantity is a conserved tracer of fluid elements. Even when not explicitly conserved, the quantity is a useful diagnostic in balanced flows. Furthermore, its long history of use in atmospheric dynamics, which involves many analogous structures and models, means that many analyses and results can be borrowed from previous work in that context.

Despite the qualitative agreement between simulations and theoretical predictions on some of the points mentioned above, it is difficult to ascertain to what extent the finer details of simulations should be considered physically accurate. Many such results are exceedingly sensitive to the specifics of the simulation and can vary considerably from code to code. The authors, de Val-Borro et al. (2006), initiated a comparison

between different incarnations of several different hydrodynamics codes on the single-protoplanet problem. Such a systematic comparison of results is essential to be able to rely on further outcomes from numerical experiments.

In this thesis I describe the development of a numerical code based on an algorithm by Pen (1998) and Trac & Pen (2003) previously used on cosmological problems. I modify the code for use in the simulation of planet-disk interactions and perform a systematic comparison of results depending on the details of the implementation similar to de Val-Borro et al. (2006). Based on this careful examination, I use the code to consider the dynamics in the region between two forming protoplanets. I am particularly interested in the vorticity distribution and the effects that it has on material transport. I confirm several of the results discussed in Ou et al. (2007) and de Val-Borro et al. (2007) and extend the study to disks with multiple protoplanets. I also discuss analogous finding from atmospheric physics and their possible relation to the protoplanetary-disk system.

In Chapter 2 I provide the necessary theoretical background required to study protoplanetary disks. Details of the numerical code are left for Chapter 3, where I also compare its performance on a suite of standard shock problems. In Chapter 4, I detail the results of the code in simulations involving a single protoplanet embedded in a gas disk. Results from simulations of multiple protoplanets are discussed in Chapter 5, and conclusions are presented in Chapter 6.

Chapter 2

Protoplanet-Disk Dynamics

2.1 Introduction

This chapter provides the necessary theoretical background to study protoplanetary disks. I discuss the basics of Eulerian fluid dynamics, introducing the necessary equations in Section 2.2. In Section 2.3 I discuss the standard steady-state models of protoplanetary disk structure, including typical treatments of viscosity and transport. Section 2.4 describes the relevant interactions that occur between disks and protoplanets that can cause planetary migration and eccentricity growth. Finally in Section 2.5 I discuss potential vorticity and its role as a transport barrier for use in Chapter 5.

2.2 Eulerian hydrodynamics

2.2.1 Mass and momentum evolution

An Eulerian formulation of the standard fluid dynamics equations may be obtained by first considering conservation of mass and momentum within a control volume (see Kundu, 1990, for example). The application of these conservation laws yields the time evolution of the fluid's density, ρ, and velocity, $\boldsymbol{u}$:

$$\frac{\partial \rho}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla} \rho = -\rho \boldsymbol{\nabla} \cdot \boldsymbol{u} \tag{2.1}$$

$$\rho \frac{\partial \boldsymbol{u}}{\partial t} + \rho \boldsymbol{u} \cdot \boldsymbol{\nabla} \boldsymbol{u} = -\rho \boldsymbol{\nabla} \phi + \boldsymbol{\nabla} \cdot \boldsymbol{\tau} - 2\boldsymbol{\Omega} \times \rho \boldsymbol{u} + \rho \Omega^2 \boldsymbol{R}, \tag{2.2}$$

where $\boldsymbol{u} \cdot \boldsymbol{\nabla}$ is the advective operator, $\boldsymbol{u} \cdot \boldsymbol{\nabla} = \sum_i u_i \frac{\partial}{\partial x_i}$. The term $\rho \boldsymbol{\nabla} \phi$ represents any conservative body force (gravity, say) and $\boldsymbol{\tau}$ is the stress tensor for the fluid. The Coriolis acceleration, $-2\boldsymbol{\Omega} \times \boldsymbol{u}$, and the centrifugal acceleration, $\Omega^2 \boldsymbol{R}$, only appear in frames which are rotating with respect to an inertial frame. The quantity, Ω, is the rotation rate of the frame and the vector, $\boldsymbol{R}$, is the cylindrical radius perpendicular to the rotation direction.

These equations express the evolution of a conserved fluid quantity, q (either ρ or u_i), due to the advection

of that fluid quantity ($\boldsymbol{u} \cdot \boldsymbol{\nabla} q$). The right-hand side term in the density equation represents any divergence of fluid through the boundary of the control volume, and in the momentum equations the right hand side includes any existing forces acting on the control volume. These equations are valid for any solid or liquid which may be treated as a continuum, no matter the relationship between the stress tensor, $\boldsymbol{\tau}$, and the fluid response.

2.2.2   Stress and strain rate tensors in the Newtonian limit

In order to determine a closed form of the above equations one needs to determine a more explicit form for the stress tensor, $\boldsymbol{\tau}$. To start, one breaks the stress tensor into an isotropic and a nonisotropic component: $\tau_{ij} = -p\delta_{ij} + \sigma_{ij}$. The isotropic component, $p\delta_{ij}$, exists even when the fluid is at rest and represents the thermodynamic pressure (see Kundu, 1990, §4.10). The nonisotropic term, σ_{ij}, results from stresses due to fluid motion. Assuming a Newtonian fluid whereby the stress tensor is linearly related to the strain rate, one expects the nonisotropic component of the tensor to be linearly proportional to the velocity gradient. This velocity gradient may be decomposed into symmetric and asymmetric parts:

$$\frac{\partial u_i}{\partial x_j} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) + \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i}\right). \tag{2.3}$$

The second term on the right-hand side does not generate stress. The first term does so and is defined to be the *strain rate tensor*, $e_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$. Kundu (1990) determines the relationship between the stress and the strain rate tensors in the Newtonian limit to be

$$\sigma_{ij} = 2\mu e_{ij} - \frac{2}{3}\mu(\boldsymbol{\nabla} \cdot \boldsymbol{u})\delta_{ij}, \tag{2.4}$$

where μ is the viscosity coefficient. The momentum evolution equation, equation (2.2), may thus be recast as

$$\rho\frac{\partial u_i}{\partial t} + \rho u_j\frac{\partial u_i}{\partial x_j} = -\frac{\partial p}{\partial x_i} - \rho\frac{\partial \phi}{\partial x_i} + \frac{\partial}{\partial x_j}\left[2\mu e_{ij} - \frac{2}{3}\mu(\boldsymbol{\nabla} \cdot \boldsymbol{u})\delta_{ij}\right]. \tag{2.5}$$

The molecular viscosity coefficient, μ, may be written in terms of the kinematic coefficient, ν, using the relation, $\nu\rho = \mu$. Specific components of stress tensor are written out for Cartesian and cylindrical coordinates in Appendix A.

2.2.3   Energy equations

In order to determine a *kinetic* energy equation, one may take the scalar product of equation (2.2) with $\boldsymbol{u}$ and integrate by parts to obtain,

$$\rho\frac{\partial(\frac{1}{2}u_i^2)}{\partial t} + \rho u_j\frac{\partial \frac{1}{2}(u_i^2)}{\partial x_j} = -\rho u_i\frac{\partial \phi_t}{\partial x_i} + u_i\frac{\partial \tau_{ij}}{\partial x_j}, \tag{2.6}$$

where $\phi_t = \phi - \frac{1}{2}\Omega^2 R^2$ and $\boldsymbol{u}\cdot(\boldsymbol{\Omega}\times\rho\boldsymbol{u}) = 0$. Kundu (1990) describes increases in the fluid's *internal* energy, ε, due to deformation of the fluid elements and the addition of heat flux, $\boldsymbol{\psi}$, by the equation

$$\rho\frac{\partial\varepsilon}{\partial t} + \rho u_j\frac{\partial\varepsilon}{\partial x_j} = \tau_{ij}\frac{\partial u_i}{\partial x_j} - \frac{\partial\psi_i}{\partial x_i}. \tag{2.7}$$

Equations (2.6) and (2.7) may be added, yielding

$$\rho\frac{\partial(\varepsilon + \frac{1}{2}u_i^2)}{\partial t} + \rho u_j\frac{\partial(\varepsilon + \frac{1}{2}u_i^2)}{\partial x_j} = -\rho u_i\frac{\partial\phi_t}{\partial x_i} + \frac{\partial(\tau_{ij}u_i)}{\partial x_j} - \frac{\partial\psi_i}{\partial x_i}. \tag{2.8}$$

Adding $\frac{1}{2}u_i^2$ times equation (2.1) to the above equation, one obtains

$$\frac{\partial e}{\partial t} + \boldsymbol{\nabla}\cdot(\boldsymbol{u}e) = -\rho\boldsymbol{u}\cdot\boldsymbol{\nabla}\phi_t + \boldsymbol{\nabla}\cdot(\boldsymbol{\tau}\cdot\boldsymbol{u}) - \boldsymbol{\nabla}\cdot\boldsymbol{\psi}, \tag{2.9}$$

where $e = \rho\varepsilon + \frac{1}{2}\rho u_i^2$ is the total energy density.

In order to close these equations, one requires an equation of state relating the internal energy to other known quantities. I consider two equations of state within the present research context. For an adiabatic gas ($\boldsymbol{\psi} = 0$ in eq. [2.9]) the internal energy density depends only on the pressure and density as $\rho\varepsilon = \frac{1}{\gamma-1}p$, for adiabatic index, γ. I also consider the simpler locally isothermal equation of state (equivalent to $\gamma \to 1$ in the adiabatic case) where the pressure and energy are related by the local sound speed, which itself is a prescribed function of the disk radius. Using $p = \rho c_s^2$, I set the sound speed to be proportional to the Keplerian velocity via $c_s = m_H v_K$. The ratio, $m_H = H(r)/r$, is a property of the disk and is taken to be constant. It describes the *thickness* of the disk where larger values correspond to a hotter (thicker) disk. Its value is typically chosen to be 0.05 and unless otherwise indicated we will adopt this fiducial value. Such parametrizations are further discussed in Section 2.3.

2.2.4   Flux form of the Euler equations

It is possible to obtain a form of the Euler equations where the conservation of mass, momentum and energy is more explicitly apparent. Beginning with equation (2.1), one can rewrite it using the vector identity, $\boldsymbol{\nabla}\cdot(a\boldsymbol{u}) = \boldsymbol{u}\cdot\boldsymbol{\nabla}a + a\boldsymbol{\nabla}\cdot\boldsymbol{u}$, as

$$\frac{\partial\rho}{\partial t} + \boldsymbol{\nabla}\cdot\rho\boldsymbol{u} = 0. \tag{2.10}$$

Similarly one rewrites the momentum equations by taking the density terms within the derivatives to yield

$$\frac{\partial\rho\boldsymbol{u}}{\partial t} - \boldsymbol{u}\frac{\partial\rho}{\partial t} + \boldsymbol{\nabla}\cdot(\rho\boldsymbol{u}\boldsymbol{u}) - \boldsymbol{u}\boldsymbol{\nabla}\cdot(\rho\boldsymbol{u}) = -\boldsymbol{\nabla}p - \rho\boldsymbol{\nabla}\phi + \boldsymbol{\nabla}\cdot\boldsymbol{\sigma}, \tag{2.11}$$

where $\rho\boldsymbol{u}\,\boldsymbol{u} \equiv \rho u_i u_j$ is a direct product yielding a matrix. Note that the terms multiplied by the vector, $\boldsymbol{u}$, are identical to those of mass conservation and vanish. The pressure and stress tensor may be taken into the

divergence term yielding

$$\frac{\partial \rho \boldsymbol{u}}{\partial t} + \boldsymbol{\nabla} \cdot [\rho \boldsymbol{u} \boldsymbol{u} + p \boldsymbol{I} - \boldsymbol{\sigma}] = -\rho \boldsymbol{\nabla} \phi, \tag{2.12}$$

where $\boldsymbol{I}$ is the identity matrix. The adiabatic energy equation may also be rearranged as

$$\frac{\partial e}{\partial t} + \boldsymbol{\nabla} \cdot [\boldsymbol{u} e - \boldsymbol{u} \cdot \boldsymbol{\tau}] = -\rho \boldsymbol{u} \cdot (\boldsymbol{\nabla} \phi). \tag{2.13}$$

When written in flux-conservative form in terms of the solution vector, $\boldsymbol{q} = (\rho, \rho u_1, \rho u_2, \rho u_3, e)$, flux tensor, $\boldsymbol{F}(\boldsymbol{q})$, and source vector, $\boldsymbol{S}$, these equations all have the same formally simple form,

$$\frac{\partial \boldsymbol{q}}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{F} = \boldsymbol{S}. \tag{2.14}$$

The form of equation (2.14) is the one I model computationally. For reference, I expand this form of the Euler equations into its separate components in both Cartesian and cylindrical coordinates in Appendix A.

Equation (2.14) may be integrated over the entire domain volume, V, to yield

$$\int_V \frac{\partial \boldsymbol{q}}{\partial t} dV + \int_V \boldsymbol{\nabla} \cdot \boldsymbol{F} dV - \int_V \boldsymbol{S} dV = 0. \tag{2.15}$$

Reversing the derivative and integral in the first term and rewriting the second term using Gauss' Law yields a useful integral form of the equation:

$$\frac{\partial}{\partial t} \int_V \boldsymbol{q} dV + \int_{\partial V} \boldsymbol{F} \cdot \boldsymbol{n} dA = 0. \tag{2.16}$$

where $\boldsymbol{n}$ is the normal vector to the domain boundary, ∂V. Unless external source terms are introduced at the domain boundary (representing influxes and effluxes of the fluid quantities) the integral of the flux over the domain boundary vanishes as the flux vector is necessarily tangential to the domain boundary. Because the source term is explicitly given as $\boldsymbol{S} = (0, -\rho \frac{\partial \phi}{\partial x_1}, -\rho \frac{\partial \phi}{\partial x_2}, -\rho \frac{\partial \phi}{\partial x_3}, -\rho u_i \frac{\partial \phi}{\partial x_i})$, all of its components are perfect integrals when integrated over the fluid volume and it too vanishes as long as the source of the gravitational potential is completely confined within the total fluid volume which is being considered. In such cases the integral, $\frac{\partial}{\partial t} \int_V q_i$, is zero and quantities that satisfy equations of the above form are not only locally conserved, but also globally conserved. Thus, the Euler equations globally conserve the domain-integrated quantities, $\int_V q_i$, specifically, the total mass, Cartesian momenta, and energy of the fluid. Global conservation will not strictly hold when there is an external potential which is located outside of the fluid volume. In such cases the external potential (say that of a central star at the center of an accretion disk which is being modelled as an annular region of fluid surrounding the star) represents a possible source/sink of energy and momentum which is only being accounted for over part of the fluid volume.

2.2.5 Angular momentum

Equation (2.12) describes the conservation of linear momentum. In cylindrical coordinates, it is not possible to write a strictly conservative form of the linear momentum equation involving the natural cylindrical momentum components ρu_r and ρu_θ, because these quantities are not *individually* conserved; rather the linear Cartesian components, ρu_x and ρu_y are each conserved. Instead, it is a fluid's *angular* momentum along the rotation axis, which is conserved in a rotating system, and as such, it is a more natural physical variable to work with. Numerically, its use yields better results as well (see §4.5). An angular momentum conservation equation may be found by taking the cross-product of equation (2.2) with the vector $\boldsymbol{R}$, defined as the components of the position vector, $\boldsymbol{x}$, which are perpendicular to the rotation of the frame (so that $\boldsymbol{\Omega} \times \boldsymbol{x} = \boldsymbol{\Omega} \times \boldsymbol{R}$). Defining the angular momentum density along the rotation axis as $h = \rho \boldsymbol{u} \times \boldsymbol{R} \cdot \hat{\boldsymbol{\Omega}}$, one has

$$\frac{\partial h}{\partial t} + \boldsymbol{\nabla} \cdot h\boldsymbol{u} + \rho \boldsymbol{u} \cdot [\boldsymbol{\nabla}(\Omega R^2)] = \boldsymbol{\nabla} \cdot \boldsymbol{\tau} \times \boldsymbol{R} \cdot \hat{\boldsymbol{\Omega}} - \rho \boldsymbol{\nabla}\phi \times \boldsymbol{R} \cdot \hat{\boldsymbol{\Omega}}. \tag{2.17}$$

The third term on the left-hand side represents the contribution to the angular momentum balance due to the rotation of the frame. By making use of the continuity equation, this term may be rewritten as $\rho \boldsymbol{u} \cdot [\boldsymbol{\nabla}(\Omega R^2)] = \boldsymbol{\nabla} \cdot [R^2 \Omega \rho \boldsymbol{u}] - R^2 \Omega \boldsymbol{\nabla} \cdot \rho \boldsymbol{u} = \boldsymbol{\nabla} \cdot [R^2 \Omega \rho \boldsymbol{u}] + R^2 \Omega \partial \rho / \partial t$. The conservative form then becomes

$$\frac{\partial H}{\partial t} + \boldsymbol{\nabla} \cdot [H\boldsymbol{u} + Rp\boldsymbol{I} - R\boldsymbol{\sigma}] = -(\rho \boldsymbol{\nabla}\phi \times \boldsymbol{R}) \cdot \hat{\boldsymbol{\Omega}}, \tag{2.18}$$

where $H = h + \rho R^2 \Omega$ is the absolute angular momentum density. This equation is written out in cylindrical coordinates in Appendix A. As per the discussion in subsection 2.2.4, both the angular momentum flux and the source term, $(\rho \boldsymbol{\nabla}\phi \times \boldsymbol{R}) \cdot \hat{\boldsymbol{\Omega}}$, vanish assuming a bounded fluid volume, so that equation (2.18) forms a global conservation equation for the angular momentum. I compare results solving for two different sets of solution variables, $(\rho, \rho u_r, \rho u_\theta)$ and $(\rho, \rho u_r, H)$, in Section 4.5.

2.3 Protoplanetary disk models

The presence of planetary disks has been inferred around about half of all T Tauri stars from their infrared excesses (see Adams et al., 1987, for example). Direct images of disks have also been observed at later stages of system development once much of the surrounding gas and dust have been removed (Bally et al., 2000). Masses have been estimated at $M_d \sim 10^{-2 \pm 1} M_\odot$ with sizes from $20 - 100$ AU (Beckwith & Sargent, 1996). Observed IR excesses (Adams et al., 1987) are consistent with the disk undergoing steady-state accretion, however, they do little to constrain the mass-infall rates. Studies of optical excess emission from young stellar objects (White & Hillenbrand, 2004) suggest time-averaged mass-accretion rates of a few $\times 10^{-8} M_\odot$/yr. Luminosities of emission-lines assumed to originate from infalling magnetospheric flows (Hartmann et al., 1994; Muzerolle et al., 1998a) suggest similar rates (Muzerolle et al., 1998b; Doppmann et al., 2005). These accretion rate estimates along with the decreasing fraction of disks observed around

late-stage T Tauri stars both suggest disk lifetimes on the order of 10^7 yrs.

2.3.1   Temperature and density distributions

Following Nelson et al. (2000), Bryden et al. (2000), and the like, I typically model accretion disks using locally isothermal conditions. The temperature and corresponding sound speed as well as the background density are prescribed functions of radius. Once these prescriptions are chosen one can determine the local pressure using the ideal gas law, $p = \rho c_s^2$.

The prescription for the temperature is determined assuming vertical hydrostatic equilibrium: to first order, the vertical pressure gradient is balanced by the vertical component of the gravitational force, $\partial p / \partial z = -\rho g \sin \theta$, where $\sin \theta \approx z/r$ for a thin disk is the ratio of mid-plane height to radius. The gravitational acceleration is given by $g = GM_*/r^2$, which written in terms of the Keplerian velocity is $g = v_K^2/r$. One can approximate the pressure gradient as $\partial p / \partial z = c_s^2 \partial \rho / \partial z$, assuming the vertical change in density away from the midplane is substantially greater than that of the temperature. Solving for the density yields an exponential atmosphere, $\rho(z) = \rho_0 \exp(-z^2/H^2)$, in terms of a scale height, $H/r = \sqrt{2} c_s/v_K$.

Observations yield disk temperatures of $T \sim 10^3 - 10^1$ K corresponding to distances from the disk center of $R \sim 10^{-1} - 10^2$ AU (Adams et al., 1988). Using the above scale height relations, these observations suggest such disks are supersonic with $H/r \sim 0.03 - 0.1$. For the purposes of this investigation, I assume a flat disk with constant value of this ratio, taken to be $H/r = 0.05$ unless otherwise indicated.

While observations can help to estimate the masses of planetary disks, their density distributions are poorly constrained. Estimates based on the heavy-metal composition of the planets within our own solar system suggests a density profile which varies as $r^{-3/2}$ between Venus and Neptune. While I have used this distribution in certain simulations, I typically model disks with initially uniform density distribution with a total mass of $0.002 M_*$ out to the orbit of the nearest planet. The results I present are not sensitive to the choice of distribution.

2.3.2   Alpha-viscosity models

Molecular viscosity is not large enough by several orders of magnitude to account for observed mass-infall rates. Diffusive processes transport fluid quantities on a timescale l^2/D, where D is the relevant diffusion coefficient. For a typical microphysical kinematic viscosity of $v = 10^5 \text{cm}^2/\text{s}$ in a disk, it would take the lifetime of the disk to transport material only $\sim 10^{-3} AU$.

To account for the high mass-infall rates observed in black-hole accretion disks, Shakura & Sunyaev (1973) postulated an anomalous turbulent viscosity which would explain the efficient transport. Their treatment and the subsequent treatment of Lynden-Bell & Pringle (1974) modelled the anomalous viscosity simply as an enhanced version of a vertically averaged Navier-Stokes viscosity. Using cylindrical coordinates, for a Keplerian disk, this assumption allows the $r - \theta$ component of the stress tensor, appearing in the

radial derivative of the angular velocity equation (see eq. [A.13]) to be written as

$$\sigma_{r\theta} = r\rho\nu_{turb}\frac{d\Omega}{dr}.$$ (2.19)

Under such an assumption, the stress depends on the local angular-velocity gradient.

It is further assumed that the enhanced turbulent viscosity can parametrized as, $\nu_{turb} = \alpha c_s H$, in terms of the dimensionless parameter, $\alpha < 1$. The rationale behind this parameterization is that turbulence can generate eddies only as large as the scale height of the disk, H, with a turnover time no faster than H/c_s. Replacing this parameterization in equation (2.19), one has

$$\sigma_{r\theta} = \alpha p\frac{d\ln\Omega}{d\ln r}$$ (2.20)

for a thin disk where $H/r \sim c_s/v_K$. The above assumptions thus relate the efficiency of transport directly to the vertically averaged local pressure. In modern literature such a treatment is termed an alpha-viscosity model.

Many physical mechanisms have been proposed to explain the efficient transport observed in accretion disks(for a review, see Papaloizou & Lin, 1995). The magneto-rotational instability (MRI) appears to be one of the few nontransient mechanisms which can generate transport rates that are large enough ($\alpha \sim 0.01$, Hawley & Balbus (1991)) to match those that are observed ($\alpha \sim 10^{-2}$,Hartmann et al. (1998)); however, its functionality in planetary disks may be limited by low levels of ionization (Gammie, 1996).

Regardless of the specifics of the physical process, equating transport due to turbulent instabilities with an alpha-viscosity prescription is accurate only in a space and time-averaged sense, at best. While authors such as Stone et al. (2000) caution against such simple prescriptions, global simulations are currently too computationally expensive to dynamically account for turbulent transport in other than such a simplified way.

When discussing viscosity present in the simulations, I adopt the approach of most authors performing global disk simulations, and use the standard alpha-prescription. Estimates in Chapter 4 suggest that numerical viscosity limits the value of alpha to be $\alpha \gtrsim 10^{-3.5}$ without an explicitly added viscosity. Larger values of α may be obtained by modelling an additional physical viscosity of a given value within a simulation.

In Chapter 5 I calculate local values of transport caused by hydrodynamic structures such as jets and vortices generated within the disk by the planet. I show that the time-averaged effective transport due to such structures is substantial and correlates with the potential vorticity distribution.

2.4   Planet-Disk interactions

Protoplanets are able to interact with a variety of solar system constituents likely to exist during the early stages of the system's evolution (during the mid- to final-stages of its formation). The catalogue of possible interaction mechanisms is often referenced to provide explanations for many of the characteristics of our

own solar system as well as the differing characteristics thus-far evident in extra-solar systems, namely the existence of gas-giant planets with short-period orbits and high eccentricities. I briefly describe some of these mechanisms.

Early solar systems are likely to be littered with small asteroid-sized planetesimals that have not yet grown large enough to begin grouping with other planetesimals or accrete gas. The ejection of a substantial fraction of these objects lying within the orbit of a protoplanet, would cause the planet to migrate inward (Murray et al., 1998). To move a Jupiter-mass planet by several AU, this mechanism would require a very massive disk of initial material to produce enough planetesimals. A gas disk of such a size would likely be unstable. However this scenario is a more plausible explanation for the migration of the less massive ice giants such as Neptune and Uranus, and possibly even Saturn (Levison et al., 2007).

If there exist multiple planets within the same system, resonant interactions can drive up the eccentricity of one or more of the objects. Rasio & Ford (1996); Weidenschilling & Marzari (1996) and Ford et al. (2001) examined this scenario with two planets, but such interactions result in an eccentricity distribution which is too low when compared with observed values (Ford et al., 2001; Goldreich & Sari, 2003). More recently Juric & Tremaine (2007) have shown that resonant interactions involving more than two planets may better explain the observed eccentricity distributions.

Protoplanets may also tidally interact with the gas disk in which they are embedded. The planet and disk are able to exchange angular momentum at resonance locations within the disk, where the orbital harmonics of the gas and the protoplanet are commensurate. The resulting gravitational torques can cause the protoplanet to migrate or alter its eccentricity, depending on the particular resonances involved. I discuss these two possible effects in more detail below.

2.4.1   Resonance locations

Goldreich & Tremaine (1979) performed a perturbation expansion of the mass and momentum conservation laws (eqs. [2.1] and [2.2]) in an inertial frame. They expanded these equations for an inviscid cylindrical disk of fluid (where the stress tensor contributes only a thermodynamic pressure term) subject to a central gravitational potential as well as an external perturbing mass. Within the context of planetary systems, the role of the central potential is played by the central star and the protoplanet provides the perturbation potential. Their analysis showed that the solution to such an expansion allows spiral density wave propagation to occur within the disk except at specific resonance locations where strong interactions occur between the protoplanet and the disk. Away from these resonance locations, the spiral waves transport angular momentum within the disk, but do not exchange momentum between the disk and the planet. Nonetheless the spiral waves are visible manifestations of the interaction. I comment further on them in Chapter 4 when examining the results of my simulations.

Resonance locations are where the frequency harmonics of the planet's potential, $\Omega_{\ell,m}$, are commensurate with either the fluid's mean orbital frequency, Ω, (so-called corotation resonances) or its epicyclic frequency, κ, (so-called Linblad resonances). For a resonance of order, m, these locations are defined by

Resonance	Amplitude	Frequency of Potential	Frequency of Fluid	Location	$(\frac{1}{a})(\frac{da}{dt})_{\text{res}}$	$(\frac{1}{e})(\frac{de}{dt})_{\text{res}}$
Principal OLR	$\phi_{m,m}(r,e_p^0)$	Ω_p	$\Omega + \kappa/m$	$(m+1)/m$	$--$	$+$
Principal CR	$\phi_{m,m}(r,e_p^0)$	Ω_p	Ω	1	?	?
Principal ILR	$\phi_{m,m}(r,e_p^0)$	Ω_p	$\Omega - \kappa/m$	$(m-1)/m$	$++$	$-$
Fast 1st-Order OLR	$\phi_{m+1,m}(r,e_p^1)$	$\Omega_p + \kappa_p/m$	$\Omega + \kappa/m$	1	$-$	$-$
Fast 1st-Order CR	$\phi_{m+1,m}(r,e_p^1)$	$\Omega_p + \kappa_p/m$	Ω	$m/(m+1)$	$-$	$--$
Fast 1st-Order ILR,	$\phi_{m+1,m}(r,e_p^1)$	$\Omega_p + \kappa_p/m$	$\Omega - \kappa/m$	$(m-1)/(m+1)$	$+$	$++$
Slow 1st-Order OLR	$\phi_{m-1,m}(r,e_p^1)$	$\Omega_p - \kappa_p/m$	$\Omega + \kappa/m$	$(m+1)/(m-1)$	$-$	$++$
Slow 1st-Order CR	$\phi_{m-1,m}(r,e_p^1)$	$\Omega_p - \kappa_p/m$	Ω	$m/(m-1)$	$+$	$--$
Slow 1st-Order ILR	$\phi_{m-1,m}(r,e_p^1)$	$\Omega_p - \kappa_p/m$	$\Omega - \kappa/m$	1	$+$	$-$

Table 2.1 List of properties of resonances including their location and their action on the planet's semi-major axis and eccentricity. Locations are listed in relative units of f given by $r = r_p f^{3/2}$ assuming $\kappa = \Omega$ as for a pressureless disk and $\kappa_p = \Omega_p$ as for a central point-mass potential. Leading-order effects are denoted by double signs. (After Masset & Ogilvie, 2004; Goldreich & Sari, 2003).

$m|(\Omega_{\ell,m} - \Omega)| = 0$ for corotation resonances (CRs) and by $m|(\Omega_{\ell,m} - \Omega)| = \pm\kappa$ for Linblad resonances (LRs). The term, Outer Linblad resonances (OLRs), distinguishes LRs that occur at physical locations in the disk which are further out from the central star relative the corresponding Inner Linblad resonances (ILRs).

The harmonics of the planet's potential may be considered by decomposing the perturbing potential into a Fourier series of angular modes, $\cos(\theta - \Omega_{\ell,m}t)$, each with amplitude, $\phi_{\ell,m}(r,e_p)$. The value of $\Omega_{\ell,m}$ corresponds to the pattern frequency of the mode. In the case of a planet moving on a circular orbit, an expansion of its potential only contains terms involving its mean orbital frequency so that $\Omega_{\ell,m} = \Omega_p$. For planets on elliptical orbits, the potential may be expanded to order, ℓ, in moments of the planet's epicyclic frequency as $\Omega_{\ell,m} = \Omega_p + (\ell - m)\kappa_p/m$. Considering components up to first order in the planet's eccentricity ($\ell = m \pm 1$), each mode, m, contributes three potential components, each of which may be commensurate with the gas disk at three possible locations. For later reference, some of the properties and physical effects of the interactions at these nine locations are summarized in Table 2.1.

2.4.2   Torquing formulas

The work of Goldreich & Tremaine (1979) developed two formulas for the torque exchange between a protoplanet and a gas disk at either LRs or CRs, which I quote here as

$$T_L = -m\pi^2 \left| \frac{\rho}{rdD/dr} \right| \left[r\frac{d\phi_{\ell,m}}{dr} + \frac{2\Omega\phi_{\ell,m}}{\Omega - \Omega_{\ell,m}} \right]^2 \Bigg|_{r_L} \tag{2.21}$$

$$T_C = \frac{1}{2}m\pi^2 \frac{\phi_{\ell,m}^2}{d\Omega/dr} \frac{d}{dr}\left(\frac{\rho}{B}\right)\Bigg|_{r_c} \tag{2.22}$$

where $D = \kappa^2 - m^2(\Omega - \Omega_{l,m})^2$ and $2B = 2\Omega + rd\Omega/dr$. All variables are to be evaluated at the location of the resonance.

Equations (2.21) and (2.22), show that the strength of the Linblad and corotation resonances increases with order, m, while at the same time, the resonance locations approach the planet (Table 2.1). If the pressure gradients in the gas were zero, the locations of the Lindblad resonances would indeed move closer to the location of the planet. However, the effect of a pressure gradient is to bias the position of the LRs away from their nominal positions according to

$$m \mid \Omega - \Omega_{\ell,m} \mid = \pm\kappa[1 + \xi^2]^{1/2}, \tag{2.23}$$

where $\xi = mH/r_p$. This bias leads to a limiting position for the high-order modes at $r_L = r_p \pm 2H\sqrt{1 + \xi^2}/3\xi$ such that $r_L \rightarrow r_p \pm 2H/3$ as $\xi \rightarrow \infty$. As the resonance locations and the planet become separated on the order of (H/r) for large m, the exerted torques become negligible. This result is referred to as the "torque cut-off" (Goldreich & Tremaine, 1980; Artymowicz, 1993).

An evaluation of the Linblad torque formula (see Goldreich & Tremaine, 1980; Ward, 1986) reveals that interactions with ILRs cause the planet to gain momentum while interactions with OLRs remove angular momentum from the planet. For the principle LRs this exchange may be be understood as follows. The perturbations of the planet on gas particles just inside its orbit cause those particles to lose angular momentum to the planet since they orbit more quickly than the planet. Likewise the particles just outside of the planet's orbit gain angular momentum from the planet. The effect of these reciprocal interactions is to clear the gas surrounding the planet.

If this clearing proceeds on a timescale that is faster than that over which viscosity can resmooth such perturbations, a gap may form in the disk. In such regimes, the density in Formulas (2.21) and (2.22) can no longer be approximated by its unperturbed value, and the interactions become nonlinear. This division between linear and nonlinear regimes based on the formation of a gap is discussed in the next two sections.

2.4.3   Type I migration

Were the exchanges of angular momentum due to principle LRs symmetrical, the planet's orbital radius would remain constant. In a gas disk, pressure gradients again exert their influence here as they shift the locations of all resonances inward. This shift occurs because pressure gradients provide additional support to the fluid at a given location in the disk, allowing the fluid's orbital frequency to be slower at a given location than without such support. This inward shift means that the OLRs couple more strongly to the planet's location than do the ILRs leading to a net loss of momentum from the planet and causing its inward migration.

When the planet's mass is small enough such that the perturbation exerted is small, linear theory can be used to examine the effect on the planet. The torque in equation (2.21) is dependent on the gas density and protoplanet mass according to $T_L \propto \rho M_p^2$. This relationship leads to a migration timescale inversely

proportional to M_p (Ward, 1997):

$$\tau \sim 2 \times 10^4 \left(\frac{M_p}{M_\oplus}\right)^{-1} \left(\frac{M_d}{0.002 M_*}\right)^{-1} \left(\frac{H/r}{0.05}\right)^3 \Omega^{-1}. \tag{2.24}$$

Migration in this regime is called type I migration.

2.4.4 Type II migration

When the perturbation becomes nonlinear, the planet begins to affect the structure of the gas disk around its orbit, altering the local gas density, ρ. The non-linear waves (carrying the angular momentum flux) are presumed to grow very quickly in amplitude and to begin to break very soon after they are launched, depositing their momentum in the vicinity of the planet. This process will try to open a gap surrounding the protoplanet. This gap is maintainable if the timescale on which the waves transfer angular momentum to the gap is shorter than the timescale on which viscous diffusion smooths out such density perturbations. With a cleared gap, the planet effectively becomes locked into the viscous evolution of the disk and will migrate inwards or outwards on a viscous timescale. Only the very outer parts of the disk are diffusing outward for typical disk models (see Pringle, 1981, for example). The planet is thus likely to migrate inwards on a timescale given by (Ward, 1997)

$$\tau \sim 4 \times 10^4 \left(\frac{\alpha}{10^{-2}}\right)^{-1} \left(\frac{H/r}{0.05}\right)^{-2} \Omega^{-1}. \tag{2.25}$$

This timescale has been parametrized in terms of the alpha-viscosity (see §2.3). Migration in this regime is referred to as Type II migration.

2.5 Potential vorticity and transport barriers

The potential vorticity (PV) is a useful quantity to consider in many balanced flows, from three-dimensional stratified systems where the bulk motion is layerwise to rotating systems where the rotation acts to "two-dimensionalize" the system. The potential vorticity has a long history of use in atmospheric physics, since Rossby (1940) and Ertel (1942) both introduced several different forms of the quantity for both two-dimensional and three-dimensional flows. In astronomy, only its two-dimensional form tends to be considered (Papaloizou & Lin, 1989; Ogilvie & Lubow, 2006), where it is has been termed the *vortensity*.

I introduce the quantity for the present work because in atmospheric contexts, strong gradients of PV have been numerically and observationally linked to transport barriers—surfaces or contours of a flow through which passively advected material elements do not pass. I wish to exploit this connection in my own results. After introducing potential vorticity, I discuss some work in atmospheric and planetary physics contexts which has explored this connection.

2.5.1　Potential vorticity evolution

The absolute vorticity, q, in any reference frame is given as the curl of the relative velocity within the frame, $\omega = \nabla \times u$, plus any contribution from the motion of the frame. At a given location in a frame rotating at the Keplerian frequency, $\Omega(r) = \sqrt{GM_*/r^3}$, the absolute vorticity is thus, $q = \nabla \times (u + r\Omega\hat{\theta}) = (\omega + 2\Omega)\hat{z}$, where the last equality assumes a two-dimensional system, rotating around the z-axis. In balanced flows without large-scale contributions from wave phenomena (inertial or acoustics waves, for example), strong PV gradients in one direction, imply the presence of a jet with a flow direction perpendicular to and with its axis along the direction of the PV gradient. This conclusion depends only on the mathematical relation between the PV and the curl of the velocity. Determining the velocity field based on the PV distribution is termed "PV inversion" in the atmospheric literature.

The equation for vorticity evolution in a corotating frame may be obtained by taking the curl of equation (2.2), to yield

$$\frac{\partial q}{\partial t} + u\cdot\nabla q = q\cdot\nabla u - q\nabla\cdot u$$
$$+ \frac{1}{\rho^2}\nabla\rho \times \nabla p + \nu\nabla \times \nabla^2 u, \qquad (2.26)$$

where some vector algebra is used to write $\nabla \times (u\cdot\nabla u) + \nabla \times 2\Omega \times u = \nabla \times q \times u = u\cdot\nabla q - q\cdot\nabla u + q\nabla\cdot u$, and the stress tensor is represented as simple Laplacian diffusion. Continuity allows us to write

$$\frac{\partial Q}{\partial t} + u\cdot\nabla Q = Q\cdot\nabla u + \frac{1}{\rho^3}\nabla\rho \times \nabla p + \frac{\nu}{\rho}\nabla^2\omega. \qquad (2.27)$$

where the *potential vorticity*, Q, is defined as $Q = \rho^{-1}q$ in astronomy.

For the two-dimensional disks I consider, the potential vorticity (PV) reduces to a scalar quantity. Also for two-dimensional flows, the first term on the right-hand side which represents vortex stretching and tilting must be zero. The second term on the right-hand side represents contributions from baroclinity; for barotropic equations of state, this terms also vanishes. In *locally* isothermal disks, the temperature gradient is a prescribed function of radius, while the density can be independently altered by compressibility, leading to baroclinic generation of vorticity. Ou et al. (2007) discuss this effect. The last term in equation (2.27) represents diffusion of relative vorticity due to dissipative processes.

Ignoring any PV generation or diffusion due to the right-hand side of equation (2.27), in two dimensions it reduces to an expression of two-dimensional vortex kinematics, and the PV measured in the inertial frame is conserved following the fluid motion. Thus, in a rotating frame, if a fluid parcel's density increases, it must begin to rotate negatively with respect to the frame's motion, in order to conserve PV.

In atmospheric physics similar kinematics govern the (2D) shallow-water equations. For more compli-cated three-dimensional systems, the *Rossby-Ertel potential vorticity*, $Q_{RE} = q\cdot\nabla\Theta$, where Θ is the potential temperature (the temperature of a given fluid parcel when adiabatically brought to a reference pressure-surface), is introduced. Rossby-Ertel potential vorticity is conserved for inviscid fluid elements along isen-

tropes (even for baroclinic fluids). In such contexts the full Rossby-Ertel potential vorticity is often referred to simply as the potential vorticity, but it should not be mistaken with the simpler, two-dimensional fluid quantity as defined in astronomy.

2.5.2   PV gradients as transport barriers

Numerous studies in atmospheric and planetary physics have examined the link between strong gradients of PV, jets and eddy-transport barriers. I provide a brief overview of some of these results and some of the current theoretical arguments which have been put forth to explain the connection.

Juckes & McIntyre (1987) performed numerical studies of polar vortex models (the polar vortices are roughly circular caps of high-PV air on either of the Earth's poles, separated from the surrounding low-PV air by a sharp PV gradient and a corresponding eastward jet). In their experiments the high-PV regions behaved as material entities, withstanding large-scale disruptions from the surrounding fluid. Although the air within the high-PV region was susceptible to small-scale erosion of material, it was almost impervious to incursions of the surrounding low-PV air. Chemical tracer data and analyses of effective diffusivities (Nakamura, 1996) from the Arctic and Antarctic polar regions confirm the high degree of isolation of this polar air and the low levels of mixing (transport) from surrounding air into these regions (McIntyre, 1989; Haynes & Shuckburgh, 2000). Marshall et al. (2006) performed a similar study of the Antarctic circumpolar current, finding low effective diffusivities along the jet axis of the current and high values on the equatorward flank. In this study regions of high and low effective diffusivity were shown to be correlated with regions of, respectively, weak and strong isentropic PV gradients.

Sommeria et al. (1989) performed laboratory experiments demonstrating low levels of transport across jets formed in a rotating tank. In these experiments, a radially inward fluid flow is setup in a rotating cylindrical tank with a sloped bottom (acting as a background PV gradient). At sufficiently large radial flow rates an eastward (prograde) jet forms with a strong gradient of PV across its axis. The strength of the jet is stated to depend upon the balance between the torque arising from the Coriolis force on the radially pumped fluid and that arising from viscosity. Away from the center of the jet, the PV gradients are relatively weak and the transport shown to be much stronger.

Finally work on the upper weather layer of Jupiter and the other gas giants suggests a similar link between jets and lowered transport. Data taken from the Cassini spacecraft (Porco et al., 2003) is consistent with Jupiter's system of belts and zones forming a "PV staircase", where broad regions of almost uniform PV lying on broad, westward jets are separated by sharp PV gradients, collocated on strong, narrow eastward jets. In his original review, Marcus (1993) reached this conclusion based on models of vortex interactions. A more recent analysis by Beron-Vera et al. (2007) demonstrates that both the eastward and westward jets of such a staircase structure should function as meridional transport barriers, explaining the relative chemical isolation of the belts and zones from one another.

Despite the large number of studies linking jets and transport barriers, a clear physical mechanism which explains the connection has not been fully developed. The relationship becomes apparent if one accepts *a*

priori the tenet that PV gradients inhibit mixing. Perturbations to such a system will amplify any existing or produced PV gradients—in regions with relatively weak PV gradients, mixing will be strong, while in regions with relatively strong PV gradients, mixing will be weak. The combined effect is reduce gradients where they are already weak and increase them where they are strong. PV inversion naturally implies a narrow, eastward jet centered on the sharp gradients.

Alternatively, if there are hydrodynamic processes distinct from two-dimensional vorticity kinematics that cause inhomogeneous mixing, sharp PV gradients and jet should develop in the regions defined by weak mixing. For further explanations of the possible mechanisms at work see works by (Baldwin et al., 2007, and references within) and Rypina et al. (2006, 2007).

Chapter 3

Computational Fluid Dynamics

3.1 Introduction

In this chapter I discuss the basics of computational fluid dynamics and summarize the numerical method based on the original Total Variation Diminishing (TVD) code by Trac & Pen (2003). I describe my modifications to the original code—the adaptation of the algorithm for use on a cylindrical grid and the implementation a fast advection scheme in the azimuthal direction based on the work of Masset (2000). In this chapter I also describe the performance of the code on a suite of four standard hydrodynamical tests.

3.2 CFD basics

In computational fluid dynamics, one generally discretizes a continuous problem by one of two ways. The first method, the Eulerian approach, discretizes the spatial domain into volumes termed grid cells, between which the fluid moves. By contrast, in the Lagrangian approach the fluid is discretized into fluid elements (or 'particles') which can then move freely according to their initial velocities and only their interactions need to be modelled.

For problems such as those in accretion-disk dynamics where the fluid velocities are dominated by a large background flow, Lagrangian methods may initially seem more intuitive. In such situations, Eulerian methods will spend the bulk of their time advecting fluid between cells, requiring hundreds of time steps per orbit simply to model the background flow, with the computational inefficiency and the accumulated numerical error implied by such numerous iterations. However, the usual Lagrangian method, SPH, does not easily allow the higher spatial accuracy that grid methods can employ, nor does it capture shocks as accurately as grid methods. Both of these properties are necessary to examine subtle effects like the interaction of resonances and to accurately calculate the torquing due to spiral density waves on the protoplanet. Furthermore, SPH codes preferentially resolve high-density regions, meaning that in planet-disk simulations, where the formation of a gap in the disk is an important part of the dynamics, these codes will under-resolve such regions (de Val-Borro et al., 2006). Given these considerations, I use an Eulerian approach, along

with an implemented method that increases algorithm efficiency in the presence of a large background flow (discussed in §3.10).

After discretizing the spatial domain into finite volumes, an Eulerian approach must calculate changes in fluid quantities over a specified *time step*. The numerical values at points on the grid are usually interpreted to be the cell-averaged quantities of the fluid variable. In order to calculate new cell-averaged quantities after a finite amount of time, one writes equation (2.15) in one dimension and for a single conserved fluid quantity so that $q(x,t) \equiv q(x,t)$. Ignoring source terms this reduction yields

$$\frac{\partial}{\partial t} \int_{x_1}^{x_2} q(x,t)dx + \frac{\partial}{\partial x} \int_{x_1}^{x_2} F(x,t)dx = 0. \tag{3.1}$$

For a single cell at location, x_i, the spatial boundaries of the cell are at $x_1 = x_{i-1/2}$ and $x_2 = x_{i+1/2}$. The integrals $\int_{x_1}^{x_2} qdx$ and $\int_{x_1}^{x_2} F(x,t)dx$ represent the cell-averaged fluid quantities, q_i and F_i, allowing one to write the discretized equation,

$$\frac{q_i^{t+\Delta t} - q_i^t}{\Delta t} + \frac{F_{i+1/2}^t - F_{i-1/2}^t}{\Delta x} = 0, \tag{3.2}$$

where superscripts reference the specific time step and subscripts reference the spatial cell. Note that in order to compute $q_i^{t+\Delta t}$ for any grid cell, one requires the value of that cell's flux, $F_{i\pm1/2}$, at its boundaries. This value for the flux must be interpolated from the known cell-averaged values at the cell centers. The manner in which a given finite-volume fluid algorithm interpolates this flux determines the order and accuracy of the algorithm.

The interpolation of the cell-boundary fluxes and the subsequent reconstruction of fluid quantities as discretized in equation (3.2) are both determined using only a finite number of neighbouring cells. This truncation of the information used in the reconstruction process introduces errors which may be diffusive or dispersive in nature. Diffusive errors result from excessive clipping and averaging occurring during the reconstruction process, resulting in the smearing of an initially sharp profile. Such errors are unavoidable in computational codes, but can be minimized. Dispersive errors result from spurious over- and undershoots occurring during the reconstruction process. They result in ringing-type oscillations occurring near sharp discontinuities.

I illustrate these two types of errors by advecting both a square wave and a sinusoidal wave form. Figure 3.1 shows the initial waveform and the resulting waveform after advection once through a periodic domain. The diffusive scheme has decreased the overall variation of both the smooth and the square waveforms and has broadened (smeared) their widths as well. The dispersive scheme does a good job of advecting the smooth wave form with little loss of amplitude, but the square wave form displays the ringing oscillations, characteristic of dispersive schemes.
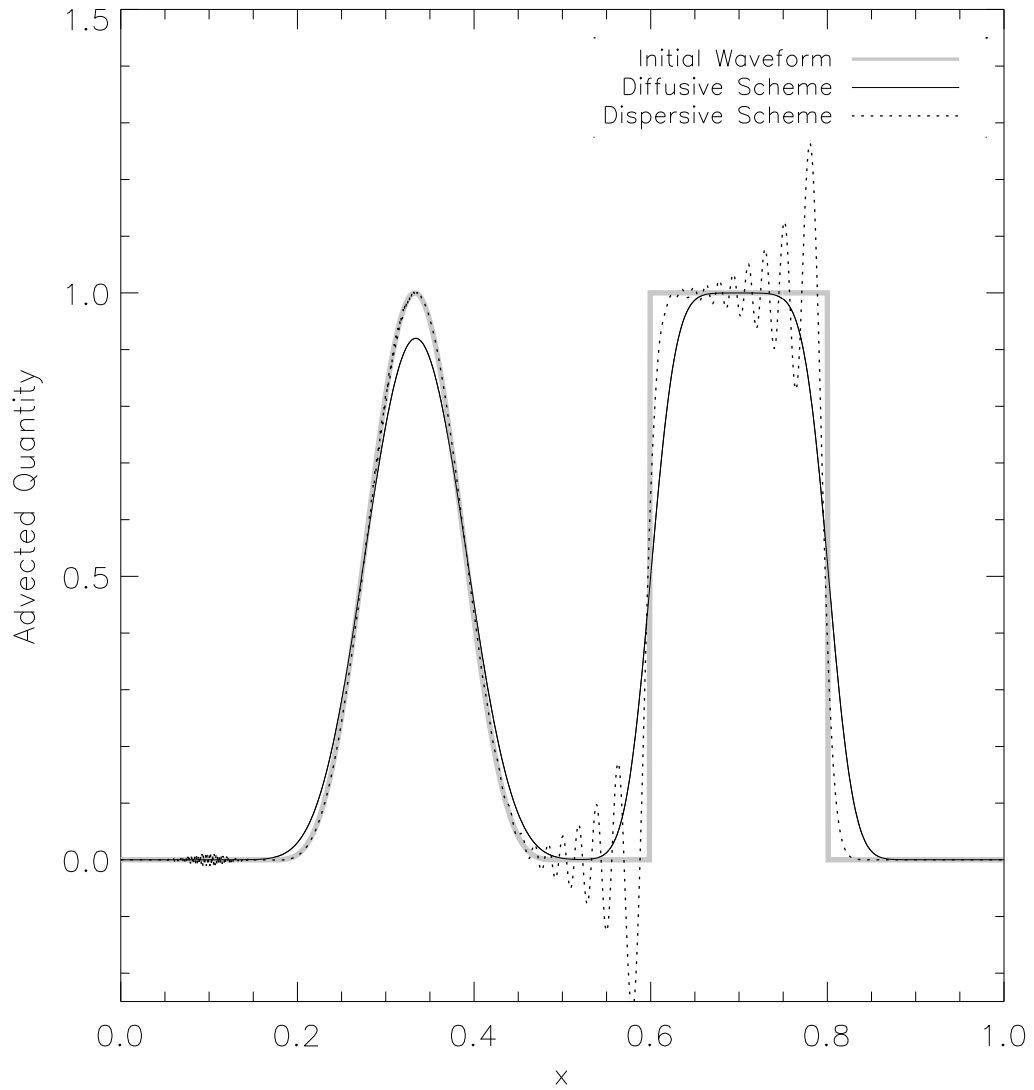
Figure 3.1 Initial and advected wave forms illustrating diffusive and dispersive errors. The thick, light-grey line is the initial wave form. The solid line is the wave advected once through the periodic box using a diffusive scheme. The dotted line is the same initial form advected once through the box with a dispersive scheme.

3.3   Algorithm summary

The TVD code is a finite-volume method, second-order accurate in time and space. It uses a Total Variation Diminishing (TVD) spatial scheme (described in §3.4), which helps to control spurious oscillations, to solve the advective terms in the Euler equations. Time-stepping is accomplished through a standard second-order Runge-Kutta scheme. For the first half-time step of the Runge-Kutta scheme, the fluxes at the cells boundaries, $F_{i\pm1/2}$, are interpolated using the first-order upwind method (described briefly in §3.4.1). For the full-time step, the second-order limiter method is used (§3.4.2). Non-viscous source terms are treated separately using a second or higher-order Runge-Kutta scheme and the viscous terms are approximated as second-order finite differences.

3.4   The Total Variation Diminishing algorithm

Full details of the TVD algorithm are provided in Trac & Pen (2003); I will merely summarize the method here. The TVD algorithm is a second-order, nonlinear algorithm that restricts the *total variation* of the discrete solution's flux at a given time step,

$$2(\sum F^t_{max} - \sum F^t_{min}), \tag{3.3}$$

to be less than or equal to that at the previous time step. The variables F^t_{max} and F^t_{min} refer to local maxima and minima in the flux's solution set. Spurious oscillations increase the number of extrema and thus increase the total variation.

3.4.1   The upwind method

Any flux-interpolation scheme which satisfies condition (3.3) is called a TVD method. A first-order example of such a method is the upwind method that I use to interpolate the cell-boundary fluxes for the half time step. It assigns fluxes based on the direction in which fluid is advecting by assuming that most of the flux at a given cell boundary comes from cells upwind of the boundary location. Considering a one-dimensional flow where flow to the right (larger indices) is positive, the upwind scheme can be described by

$$\begin{aligned} F^U_{i+1/2} &= F_i, & u_i > 0 \\ F^U_{i+1/2} &= F_{i+1}, & u_{i+1} < 0. \end{aligned} \tag{3.4}$$

This algorithm is highly diffusive. Godunov (1957) discusses the properties of TVD algorithms, noting that the upwind method is the only first-order algorithm which is TVD; and therefore, less diffusive algorithms must necessarily be nonlinear.

3.4.2 Second-order schemes

It is possible to improve upon the upwind scheme by considering second-order corrections to the assigned fluxes. For flow to the right there are two second-order corrections:

$$
\begin{aligned}
\Delta F^L_{i+1/2} &= \frac{F_i - F_{i-1}}{2}, \\
\Delta F^R_{i+1/2} &= \frac{F_{i+1} - F_i}{2}.
\end{aligned}
\tag{3.5}
$$

These two corrections consider the influence of flux from the cells further to the left and to the right of the ith (upwind) cell. In a similar manner for flow to the left, there are two second-order corrections, which consider the influence of flux from cells to the left and right of the $(i+1)$-th (upwind) cell:

$$
\begin{aligned}
\Delta F^L_{i+1/2} &= \frac{F_{i+1} - F_i}{2}, \\
\Delta F^R_{i+1/2} &= \frac{F_{i+2} - F_{i+1}}{2}.
\end{aligned}
\tag{3.6}
$$

To determine the actual value of the correction to the purely upwind flux, I apply a *limiter*, $\phi(\Delta F^L, \Delta F^R)$, which determines the relative weight of the two corrections. For a given limiter, the second-order boundary flux is determined to be $F^U_{i+1/2} + \Delta F_{i+1/2}$, where $\Delta F_{i+1/2} = \phi(\Delta F^L_{i+1/2}, \Delta F^R_{i+1/2})$.

I consider four established limiters: Minmod, Van Leer, MC and Superbee, all designed to satisfy the TVD condition (eq. [3.3]) as well as two newly designed schemes. The Minmod limiter determines the flux correction according to

$$
\phi_{MM}(a, b) = \begin{cases} \min(|a|, |b|), & ab \geq 0 \\ 0, & ab < 0 \end{cases}.
\tag{3.7}
$$

The Van Leer limiter takes the harmonic mean of the two corrections:

$$
\phi_{VL}(a, b) = \begin{cases} \frac{2ab}{a+b}, & ab \geq 0 \\ 0, & ab < 0 \end{cases}
\tag{3.8}
$$

while the MC limiter determines the flux as,

$$
\phi_{MC}(a, b) = \begin{cases} \min\left[\frac{|a|+|b|}{2}, \min(2|a|, 2|b|)\right], & ab \geq 0 \\ 0, & ab < 0 \end{cases}.
\tag{3.9}
$$

The superbee limiter is defined as

$$
\phi_{SB}(a, b) = \begin{cases} \max(\min(2|a|, |b|), \min(|a|, 2|b|)), & ab \geq 0 \\ 0, & ab < 0 \end{cases}.
\tag{3.10}
$$

It is possible to examine these schemes graphically in terms of the ratio of leftward and rightward cor-

rections, $\xi = \Delta F^L / \Delta F^R$, and the corresponding magnitude of the chosen flux correction, ΔF. They are displayed in Figure 3.2. Note that all the above limiters are zero when the flux corrections are of opposite sign (where ξ is negative) as occurs near an extremum. This feature prevents growth of the extremum. They are also all symmetric under exchange of a and b. Note that the exchange, $a \to b$ corresponds to $\xi \to 1/\xi$; hence, the definition $\xi = \Delta F^L / \Delta F^R$ rather than $\xi = \Delta F^R / \Delta F^L$ is arbitrary.

The region defined by $\Delta F < \max(0, \min(2, 2\xi))$ satisfies non-linear stability conditions determined to be in the general class of TVD-stable limiters (Sweby, 1984). Limiters that satisfy the above condition and that are also second-order accurate are found within the area bounded by the superbee and minmod limiters.

The Minmod limiter is the most diffusive because it always takes the minimum value of the correction possible which is still second order; thus, any flux not assigned by the second-order reconstruction ends up being smeared out over more than one grid cell. The Van Leer, and MC, AS and mixed schemes (see below) are all progressively less diffusive, as they assign more and more of the possible flux correction to a definite cell. Superbee is the least diffusive second-order limiter possible, but at the cost of increased instability. These tradeoffs are discussed further in Section 3.11.

Also shown on the graph are the AS and MB limiters I designed which sometimes exhibit better compromises between stability and higher-order accuracy. The MB limiter is a normalized linear combination of the MC and SM schemes. In practice I usually weight the scheme as 80% MC and 20% SB as drawn in Figure 3.2. The AS limiter is asymmetric with respect to the left- and rightward flux corrections; thus, it is defined differently for $\xi_1 = \Delta F^R / \Delta F^L$ and $\xi_2 = \Delta F^L / \Delta F^R$. It can, however be defined without reference to either flux ratio as

$$\phi_{as}(a,b) = \begin{cases} \max\left[0, \min\left\{\frac{3(|a|+|b|)}{5}, \min(|a|, 2|b|)\right\}, \min\left\{\frac{(5|a|-3|b|)}{3}, \min(2|a|, |b|)\right\}\right], & ab \geq 0 \\ 0, & ab < 0. \end{cases} \tag{3.11}$$

I discuss more details of their performance in subsections 3.11.2 and 3.11.3.

3.5 Time step restrictions

In nature, information of various kinds is limited by physical laws to propagate at finite speeds. The speed of sound, for example, limits the speed at which pressure disturbances can propagate in a given medium. Numerically, information propagates on a grid because neighbouring cells are used to determine a given cell's reconstruction at the next time step. Restrictions equivalent to the speed of sound must be imposed in a numerical code to insure that numerical information does not propagate at physically unrealistic speeds. In practise one limits the value used for the time step so that the types of waves being modelled cannot affect grid cells which are further away than the waves could have reached were their equivalent physical speeds so-limited. Such conditions are called *Courant-Friedrichs-Lewy* (CFL) conditions, and they ensure CFL stability. Failure to regulate the time step in this way will cause the code to be massively unstable and highly inaccurate.
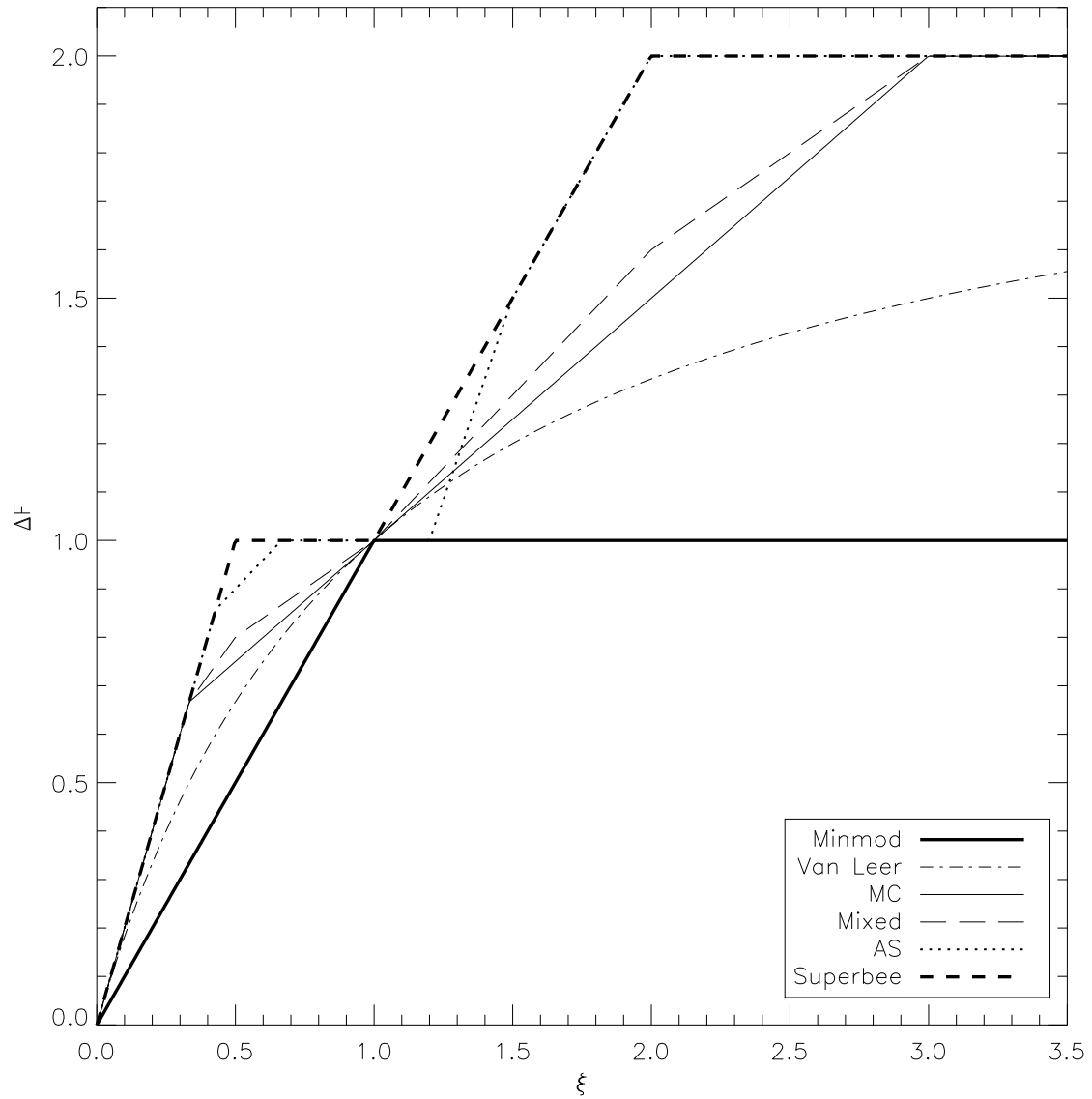
Figure 3.2 Flux corrections for the considered limiter schemes: Minmod (heavy solid), Van Leer (dot-dashed), MC (fine solid), mixed (fine dashed), AS (dotted), and Superbee (heavy dashed).

Consider the one dimensional Euler equations. They support three types of waves: entropy waves which move at the fluid's flow speed, u, and two types of acoustic waves which travel "rightward" and "leftward" at the speed of sound relative to the flow speed, $u + c_s$ and $u - c_s$, respectively. Then, at a given spatial location, the maximum and minimum values of these three speeds determines the speed at which information can travel to the right and to the left, respectively. More generally one can limit the time step due to the two global speed extrema. These two speeds are equivalent to the largest and smallest eigenvalue of the flux Jacobian, $\partial F(q)/\partial q$ as demonstrated in Laney (1998).

The time step may be further limited by other physical restrictions such as the CFL condition imposed by viscosity. Physical and numerical viscosity are further discussed in Section 4.6. The fast advection algorithm (§3.10) places a further restriction on the time step in order to ensure that differential rotation does not cause two adjacent annuli of fluid to shear past one another by more than a half grid cell. In simulations of protoplanetary disks, the restriction due to the acoustic waves is usually the most stringent.

3.6   Wave splitting

In the discussion of the TVD method I have implicitly assumed that it is possible to determine the fluid's direction of flow. When solving the Euler equations, this determination is not always straightforward as the pressure can influence the total velocity in a cell through the sound speed. In order to determine a set flow direction for each cell, the equations are split into leftward and rightward-travelling components and the coupled system,

$$\frac{\partial \boldsymbol{q}}{\partial t} + \frac{\partial}{\partial x}(c\boldsymbol{w}) = 0 \tag{3.12}$$

$$\frac{\partial \boldsymbol{w}}{\partial t} + \frac{\partial}{\partial x}(c\boldsymbol{q}) = 0, \tag{3.13}$$

is solved, where the definitions, $\boldsymbol{q} = \boldsymbol{q}^R + \boldsymbol{q}^L$, and $\boldsymbol{w} = \boldsymbol{F}/c = \boldsymbol{q}^R - \boldsymbol{q}^L$, define the solution variables in terms of the leftward and rightward-travelling waves. Equation (3.13) represents a separate equation for the evolution of the normalized flux vector, $\boldsymbol{w}$. The variable, c, is a positive-definite function which has the interpretation of a speed associated with a particular grid cell. The solution is TVD for values of c greater than or equal to the largest eigenvalue of the flux Jacobian, $\partial F(q)/\partial q$. Because the the waves are split into separate rightward and leftward components, the maximum eigenvalue of the Jacobian is satisfied for both components by the value $c_i = |u_i| + c_s$ where c_s is the sound speed for the cell. Substituting these definitions into equations (3.12)–(3.13) decouples the system and yields

$$\frac{\partial \boldsymbol{q}}{\partial t} + \frac{\partial \boldsymbol{F}^R}{\partial x} - \frac{\partial \boldsymbol{F}^L}{\partial x} = 0, \tag{3.14}$$

where $\boldsymbol{F}^L = c\boldsymbol{q}^L$ and $\boldsymbol{F}^R = c\boldsymbol{q}^R$. The original coupled system, equations (3.12) and (3.13), is then equivalent to the solutions of the two separate leftward- and rightward-moving waves given in equation (3.14). It is

now possible to separately solve for each of the travelling waves using the second order Runge-Kutta/TVD scheme and add results to determine the full solution.

3.7   Operator splitting

The above description applies to the Euler equations in one dimension. Multiple dimensions and additional source terms present in the full Navier-Stokes equations are accounted for by using the operator splitting technique of Strang (1968). A full time step is performed as a double sweep through an ordered sequence of operators comprising the full equation, first in forward sequence, then in reverse. To illustrate this process for a single double sweep, I write equation (2.14) using operators as

$$\frac{\partial q_j}{\partial t} + \sum_{i}^{N_D} L_i[q_j] - S_{g,\phi}[q_j] - V[q_j] = 0, \tag{3.15}$$

where N_D is the number of physical dimensions being modeling. The $L_i[q_j]$ terms represent the update of q_j in a single direction due to advective terms, $\sum_{i}^{N_D} u_i \partial q_j / \partial x_i$. The operators, $S_{g,\phi}$ and V represent additional routines which differ numerically from the TVD algorithm and account for the effects of gravity, source terms due to cylindrical geometry and viscosity. I have lumped the source terms due to gravity and geometry into the same operator as they are performed in the same subroutine. The first half of the double sweep is then performed in the sequence,

$$q_j^{t+\Delta t} = V S_{g,\phi} L_3 L_2 L_1 [q_j], \tag{3.16}$$

and a second sweep is then performed using the same time step, Δt, to yield the completely updated solution

$$q_j^{t+2\Delta t} = L_1 L_2 L_3 S_{g,\phi} V V S_{g,\phi} L_3 L_2 L_1 [q_j]. \tag{3.17}$$

As discussed in Strang (1968) this procedure ensures second-order accuracy.

3.8   Alterations for cylindrical grids

In order to determine the changes required to solve the Euler equations on cylindrical rather than Cartesian grids, I examine the equations in Appendix A for zero viscosity ($\mu = 0$) and no external potential ($\phi = 0$). In Cartesian coordinates they may be written as

$$\frac{\partial q}{\partial t} + \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} = 0 \tag{3.18}$$

compared to their form in cylindrical coordinates:

$$\frac{\partial \boldsymbol{q}}{\partial t} + \frac{1}{r}\frac{\partial r\boldsymbol{F}_r}{\partial r} + \frac{1}{r}\frac{\partial \boldsymbol{F}_\theta}{\partial \theta} + \frac{\partial \boldsymbol{F}_z}{\partial z} = \boldsymbol{S} \qquad (3.19)$$

where $\boldsymbol{S} = (0, \rho u_\theta^2/r, \rho u_r u_\theta/r, 0, 0)$ for the solution set $\boldsymbol{q} = (\rho, \rho u_r, \rho u_\theta)$ and $\boldsymbol{S} = (0, \rho u_\theta^2/r, 0, 0, 0)$ for the solution set $\boldsymbol{q} = (\rho, \rho u_r, H)$.

Equations (3.18) and (3.19) are computationally equivalent for the substitutions $\boldsymbol{q}(x, y, z) \rightarrow \boldsymbol{q}(r, \theta, z)$, $\boldsymbol{F}(x_i, y_i, z_i) \rightarrow \boldsymbol{F}(r_i, \theta_i, z_i)$ and $(\Delta x, \Delta y, \Delta z) \rightarrow (r_i \Delta r, r_i \Delta\theta, \Delta z)$ if the effects of the source term $\boldsymbol{S}$ are also included. I account for this term using a second-order Runge-Kutta scheme implemented with operator splitting as discussed in Section 3.7. When performing the radial sweep the fluxes are multiplied by the additional factor of r as required. Equivalently: while the azimuthal advection operator acts directly on the solution vector as $L_\theta[\boldsymbol{q}]$, the radial advection operator acts on the solution vector scaled by r as $L_r[r\boldsymbol{q}]$.

3.9 Boundary treatments

Special boundary conditions are required only in the non-angular directions. The angular coordinate is numerically treated as periodic by directly mapping the $(N_\theta+1)$-th cell to the 1st cell when calculating the fluxes.

In the radial direction, I use $n_b = 2^{N_D}$ ghost cells on the inner and outer edges of the computational domain. This number of additional cells prevents the effects of lower-order flux interpolations occurring at the first and last cells from propagating in towards the center of the domain. After the solution quantities are updated at the end of each double sweep, the values of these boundary cells are redetermined, depending on the physical effect being employed. I consider three treatments here. The first simply re-initializes the quantities in the ghost cells to the initial conditions or some known, prescribed solution. The second re-assigns the cell values according to

$$\begin{aligned} q_{n_b-i} &= w q_{n_b+i+1}, && i = 0, n_b - 1 \\ q_{N_r-n_b+i+1} &= w q_{N_r-n_b-i}, && i = 0, n_b - 1 \end{aligned} \qquad (3.20)$$

where $w = 1$ for all variables except the radial velocity when $w = -1$. This boundary treatment approximates reflecting boundary conditions for which all scalar variables are symmetric around the boundary while vector variables are anti-symmetric. The third treatment re-assigns the cell values according to the prescription

$$\begin{aligned} q_{i+1} &= q_{n_b+i+1}, && i = 0, n_b - 1 \\ q_{N_r-n_b+i} &= q_{N_r-2*n_b+i}, && i = 0, n_b - 1 \end{aligned} \qquad (3.21)$$

for all variables. This treatment approximates a free-streaming outflow boundary.

In addition to the boundary treatments discussed above, I sometimes implement wave-damping condi-

tions near the boundaries, but still inside the solution domain proper. These wave-damping conditions are described by

$$q(r, t) = q(r, 0) + [q(r, t) - q(r, 0)]e^{|r - r_b|t/(\tau R_0)}. \tag{3.22}$$

Such a treatment damps any perturbations about the initial equilibrium solution that are within distance, r_b, of the boundary, on a spatial scale, R_0 and on an orbital time scale, τ.

3.10 Fast angular advection routine

I discuss in Section 3.2 how accurately predicting the physics of accretion disk dynamics requires the use of a grid code, but that such an approach is computationally inefficient and numerically diffusive for flows with a large background component. The large background flow speed lowers the time step allowed by the CFL stability condition, requiring hundred of iterations per orbit, much of which goes into simply translating the fluid in the direction of the background flow.

In order to increase the efficiency of the code in such situations, I adopt the fast advection algorithm described by Masset (2000) for use in two-dimensional polar grid codes. The algorithm takes advantage of the near-Keplerian velocity profile of accretion disks in order to obtain a significant computational speed-up as well as a significant reduction in the numerical viscosity. The algorithm's underlying strategy is to subtract off the bulk background flow, which can be considered simply a translation of grid quantities in the angular direction, leaving the dynamically important residual velocity.

I discuss the velocity decomposition scheme of the algorithm below and the methods used to transport the fluid variables by each of the different components.

3.10.1 Method summary

In decomposing the velocity, the cylindrical grid is broken up into a series of annuli and an averaged, background velocity in the azimuthal direction, $u_{AVG}(r)$, is calculated for each annulus. The first velocity component corresponds to the residual amount, $u_{RES}(r, \theta)$, by which the total velocity differs from its azimuthally averaged value, $u_{AVG}(r)$. The averaged, background velocity is further decomposed as, $u_{AVG}(r) = u_{SH}(r) + u_{CR}(r)$. The former component is constructed to correspond to the largest possible whole-number shift of grid cells in the azimuthal direction and the second to the remaining partial-cell shift. Neither of these components depend on the angular variable. The whole-number shift is rounded to the nearest integer so that the partial-cell shift may be positive or negative, but will always correspond to a shift magnitude less than or equal to half a grid cell. The total velocity at any cell on the grid is then given as

$$u(r, \theta) = u_{SH}(r) + u_{CR}(r) + u_{RES}(r, \theta). \tag{3.23}$$

The transport of fluid quantities due to the u_{SH} component is easily accomplished by numerically shifting

the fluid variables by the appropriate number of cells in the azimuthal direction. Because the shift is integral, the process does not introduce any numerical diffusion. Also, since it represents the advection of a balanced, steady-state fluid flow, the process does not limit the size of time step permitted by the CFL condition as such flows do not support wave phenomena.

Transport of the partial-cell component is accomplished by shifting a second-order interpolation of the fluid variables by the appropriate azimuthal amount. The details of the second-order interpolation used are laid out in the subsection below. While this treatment does introduce some numerical viscosity via the interpolation, per the same reasoning as above, it does not limit the time step allowed by the CFL condition.

Finally, transport of fluid quantities due to the residual azimuthal component is performed by the TVD algorithm, just as for the radial velocity sweep. This transport step contributes to the numerical viscosity and also lowers the size of the time step allowed for stability. However, if the flow is near-Keplerian, the residual velocity will be small and the time step will be much larger than that permitted by the full azimuthal velocity. In practice, the allowed time step is increased by a factor of $5 - 10$ times that allowed without removing the background flow. I provide further comparisons and discussion of this improvement in Section 4.4.

3.10.2   Transporting fluid quantities via second-order interpolations

Transport of a fluid quantity along an annulus due to a velocity which is independent of the grid cell may be accomplished by interpolating the function and redetermining the interpolation at a shifted location. This process is not unstable; therefore, it does not decrease the CFL-allowed time step, but it does introduce some diffusion as peaks in the interpolated quantity are shifted by fractions of a grid cell and must then be redistributed among more than one cell.

Consider a known distribution along a single direction of cell-averaged values, $\overline{q}(x_i)$, corresponding to a given fluid quantity. The cell-averaged values of the quantity can be related to an assumed underlying continuous quantity, $q(x)$, by the relation

$$\overline{q}(x_i) = \frac{1}{x_{i+1/2} - x_{i-1/2}} \int_{x_{i-1/2}}^{x_{i+1/2}} q(x)dx \qquad (3.24)$$

Assuming a single advection direction, u_{ad}, one can determine an upwind two-cell interpolation for the continuous function, $q(x)$, as

$$q(x) = m_-(x - x_{i-1}) + b_-, \quad u_{ad} > 0 \qquad (3.25)$$

$$q(x) = m_+(x - x_{i+1}) + b_+, \quad u_{ad} < 0 \qquad (3.26)$$

where $m_- = (\overline{q}_i - \overline{q}_{i-1})/(x_i - x_{i-1})$, $m_+ = (\overline{q}_{i+1} - \overline{q}_i)/(x_{i+1} - x_i)$, $b_- = \overline{q}_{i-1}$ and $b_+ = \overline{q}_{i+1}$. Notice that the definition of $\overline{q}(x_i)$ in equation (3.24) is self-consistent for the above interpolations. For a fractional cell shift of width, $f\Delta x = u_{CR}\Delta t$, where $u_{CR} > 0$ and $\Delta x = (x_{i+1/2} - x_{i-1/2})$ material advected into the ith cell was

previously between $x_{i+1/2} - |f|\Delta x$ and $x_{i-1/2} - |f|\Delta x$. Then, the updated cell-averaged value at x_i is

$$
\begin{aligned}
\overline{q}'(x_i) &= \frac{1}{\Delta x} \int_{x_{i-1/2}-|f|\Delta x}^{x_{i+1/2}-|f|\Delta x} [m_-(x - x_{i-1}) + b_-]dx \\
&= \overline{q}_i(1 - |f|) + \overline{q}_{i-1}|f|.
\end{aligned} \tag{3.27}
$$

For a shift where $u_{CR} < 0$, the cell-averaged fluid quantity is updated due to advection of material previous located between $x_{i+1/2} + |f|\Delta x$ and $x_{i-1/2} + |f|\Delta x$ so that

$$
\begin{aligned}
\overline{q}'(x_i) &= \frac{1}{\Delta x} \int_{x_{i-1/2}+|f|\Delta x}^{x_{i+1/2}+|f|\Delta x} [m_+(x - x_{i+1}) + b_+]dx \\
&= \overline{q}_i(1 - |f|) + \overline{q}_{i+1}|f|.
\end{aligned} \tag{3.28}
$$

The above two-cell interpolation may be extended to higher order. In the code I implement a centered, three-cell interpolation (of order f^2) to update the fluid quantities by the constant fractional velocity, u_{CR}, in the azimuthal direction.

3.11 Hydrodynamical tests

I perform a suite of four hydrodynamic tests in order to ascertain the accuracy and stability of the code: a Sedov-Taylor blast wave, a two-dimensional oblique shock at three different angles, a Kelvin-Helmholtz (KH) instability test and a cylindrical bow-shock test. All except the KH instability test contain supersonic fluid flow. The first three tests are performed in Cartesian coordinates and the last test is performed in cylindrical coordinates. The results of these tests are compared to those from the piecewise parabolic method (PPM, Colella & Woodward, 1984) code as implemented in *VH-1* (`http://wonka.physics.ncsu.edu/pub/VH-1/`), and when possible with analytical solutions. When referring to results from the TVD code, the limiter scheme used for the simulation will be placed in parentheses.

3.11.1 Sedov-Taylor blast wave

A Sedov-Taylor blast wave is a strong, pressure-driven explosion resulting in a spherical shock front with self-similar subsonic flow in the post-shock region. Landau & Lifshitz (1959) detail the analytical solution. In order to simulate such an explosion, I initialize the computational grid with a fluid of uniform density, ρ_0, and pressure, p_0, except for the central grid cell where the pressure is set to be $p_c = 10^5 p_0$. The gas is unmoving initially and I set $\gamma = 5/3$.

Figure 3.3 compares the resulting density, momentum and pressure using 128^3 grid cells after a time where the shock wave has travelled 57 grid cells. A random selection of 1/100th of the data is plotted. Values have been normalized to their expected analytical values immediately behind the shock front which
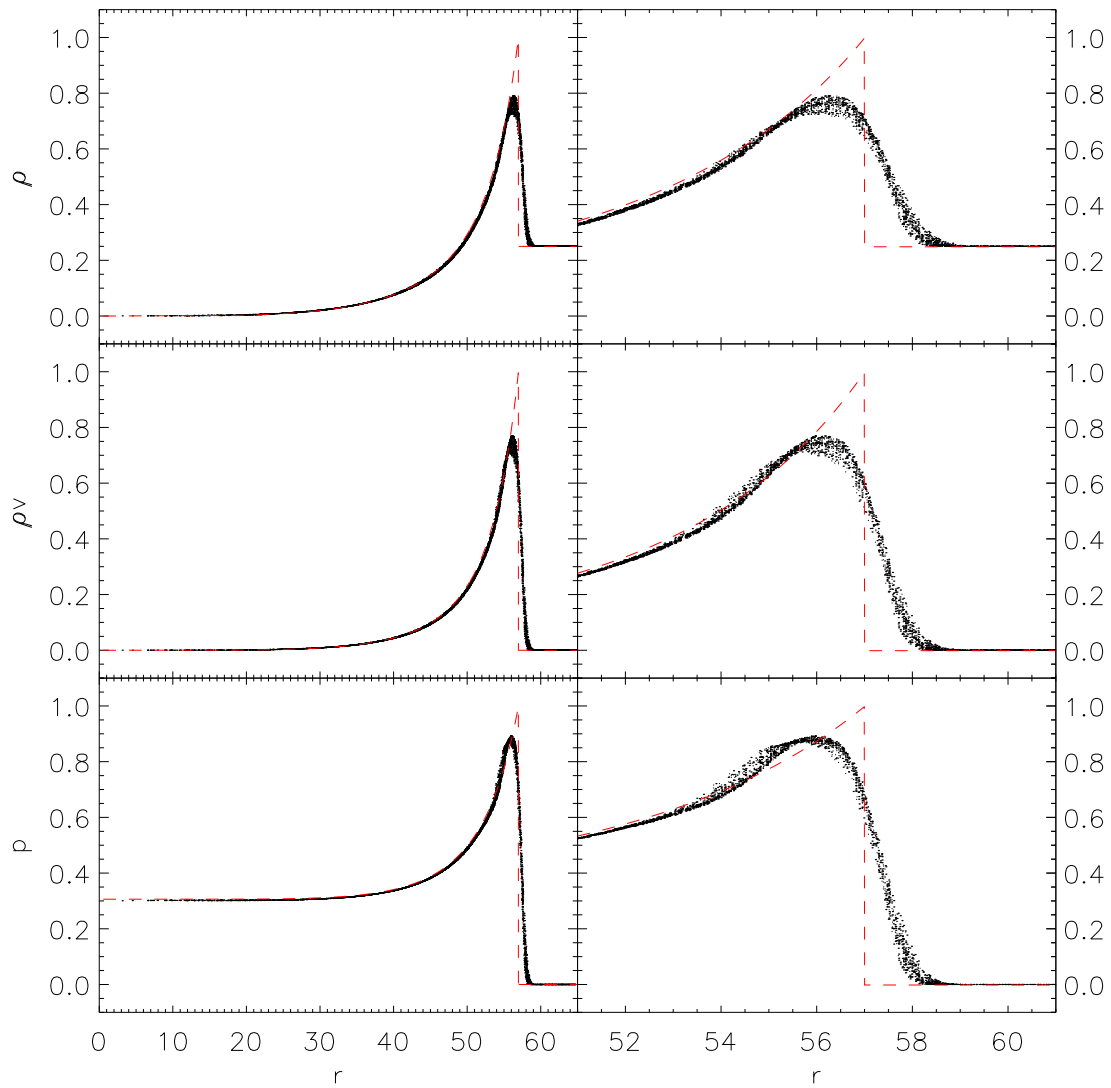
Figure 3.3 Density, total momentum and pressure of a Sedov-Taylor explosion after the initial shock front has travelled 57 grid cells. The analytical solutions are drawn with dashed lines. All values for the quantities have been normalized to their expected analytical values immediately behind the shock front.
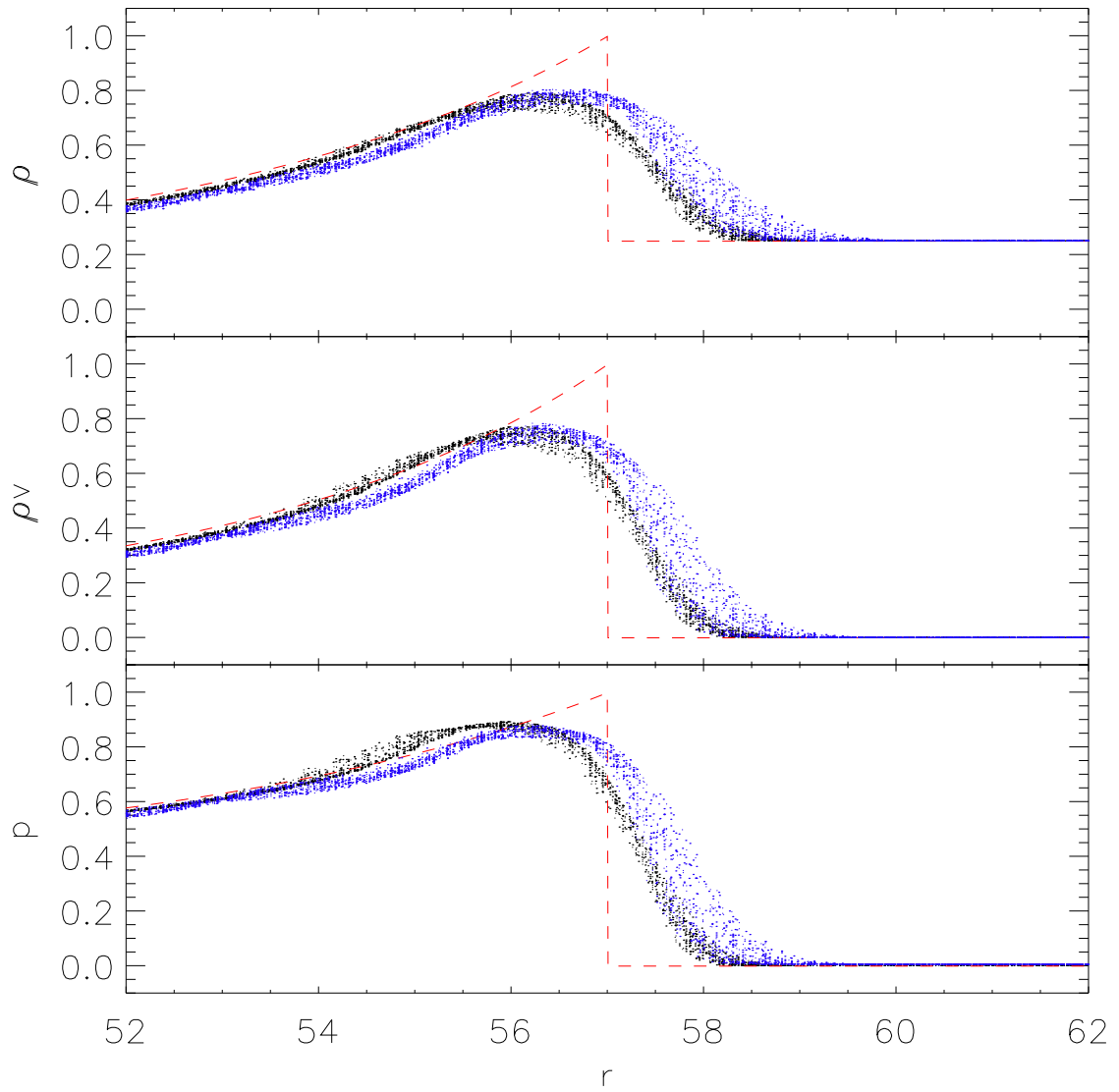
Figure 3.4 Density, total momentum and pressure after the shock front has travelled 57 grid cells. TVD(VL) results are shown in black and PPM results are shown in blue. The analytical solutions are drawn with red, dashed lines.

I quote from Landau & Lifshitz (1959) as:

$$\rho_s = \left(\frac{\gamma + 1}{\gamma - 1}\right)\rho_0 \tag{3.29}$$

$$v_s = \left(\frac{2}{\gamma + 1}\right)v_{fr} \tag{3.30}$$

$$p_s = \left(\frac{2}{\gamma + 1}\right)\rho_0 v_{fr}^2 \tag{3.31}$$

where ρ_0 is the density of the ambient medium and v_{fr} is the velocity of the shock front. These values are determined by the Rankine-Hugoniot jump conditions and can be applied to shocks in general. Note that for an adiabatic gas with $\gamma = 5/3$ the post-shock density can be at most 4 times larger than that of the ambient medium. For an isothermal shock, it is unconstrained.

The code does a good job of reproducing the analytical results of all three quantities in the post-shocked region, resolving the shock front over $2 - 3$ grid cells and reproducing about 80% of the expected shock strength. Figure 3.4 compares the resolution of the shock front as captured by the TVD(VL) and PPM codes. The resolution of the front is slightly better with the TVD code and the shape of the front differs slightly between the two codes. Besides these differences the codes perform similarly.

3.11.2 Two-dimensional oblique shock

The two-dimensional oblique shock is a version of the one-dimensional Sod shock tube (see Landau & Lifshitz, 1959, for example) set-up in a two-dimensional box at an angle to the box boundaries. In the 1-D Sod shock a jump discontinuity is initialized where the pressure and density of two regions of fluid are initially disparate (say, separated by a membrane). The two regions of fluid may or may not initially be moving relative to one another. In the 2-D oblique case, the fact that the initial density and pressure discontinuities are at an angle causes the resulting shock front to propagate obliquely across the box. The specific initial conditions I implement are given by

$$\rho = \begin{cases} 1, & x < x_0 \\ 0.125, & x \geq x_0 \end{cases} \tag{3.32}$$

$$p = \begin{cases} 1, & x < x_0 \\ 0.1, & x \geq x_0 \end{cases}, \tag{3.33}$$

with no initial velocity. For the two dimensional test, I set $x_0 = 0.25$ along the vertical and horizontal box boundaries so that $x_0 = \sqrt{2}(0.25)$ along the central diagonal.

Figure 3.5 shows the density and pressure of the TVD(VL) and PPM codes measured along the central diagonal of the box and the corresponding analytical solution at the elapsed time. Both codes agree well with the analytical solution, capturing the shock front and backward travelling rarefaction wave in both density and pressure as well as the contact discontinuity in the density profile near $x_{diagonal} = 0.5$.
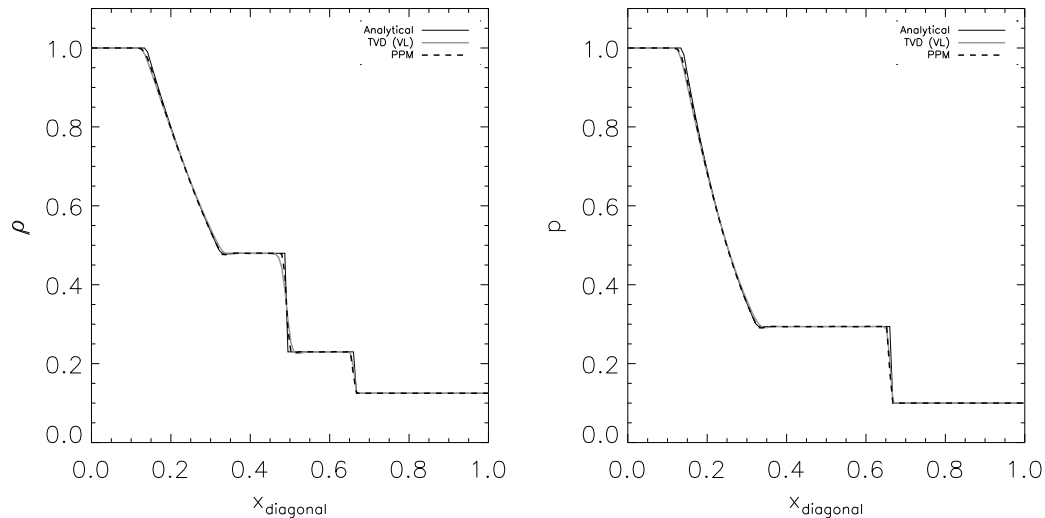
Figure 3.5 Density and pressure when the shock has reached $x_{diagonal=0.67}$ Results from the TVD(VL) are drawn in the heavy, grey line, results from the PPM simulation are shown in the heavy, dashed line and the analytical solution is drawn in the fine, solid line.

In order to compare the results between the different TVD limiter schemes, Figure 3.6 shows the difference between the computed density and that of the analytical solution for the five TVD limiter schemes as well as the PPM code (differences between the computed and analytical pressure show analogous results). The PPM code does better than any of the TVD codes at minimizing the diffusion near shock fronts and discontinuities, while the diffusion of the TVD codes varies considerably at these points. The superbee limiter has the least diffusion of the TVD codes, and is close to the PPM results except that it displays some spurious oscillations before the shock fronts. These oscillations are indicative of instabilities and in practice, if the simulation has too many shocks, the superbee limiter can force the size of the time step allowed by CFD conditions too small to be of practical use. The asymmetric limiter scheme has only slightly more diffusion than the superbee scheme and correspondingly has smaller pre-shock oscillations and higher stability. In practice the asymmetric scheme has not forced the time step size to be too small and has proved a good compromise between stability and lowered diffusion. The MC limiter scheme (not shown) has almost identical results to the Van Leer scheme which shows more diffusion than the previous schemes and the minmod scheme shows the most diffusion. The PPM code does not show oscillations because it places further conditions on the dynamics that tend to flatten gradients before and after a discontinuity.

Because the Sod Shock Tube is propagating at an angle, it proves a useful examination of any differences caused by dimensional splitting. Figure 3.7 shows contour plots of the density from the TVD and PPM codes run with $N_x \times N_y = 180x180$ resolution and outflow conditions on all boundaries. For most of the length of the initial jump discontinuity the shock front is straight and propagates at the same speed along the original diagonal. As one gets close to the sides of the front, the fluid is able to "bleed" away towards the open sides (and eventually out of the box once it reaches the boundaries). As a result the shock front begins to diffract
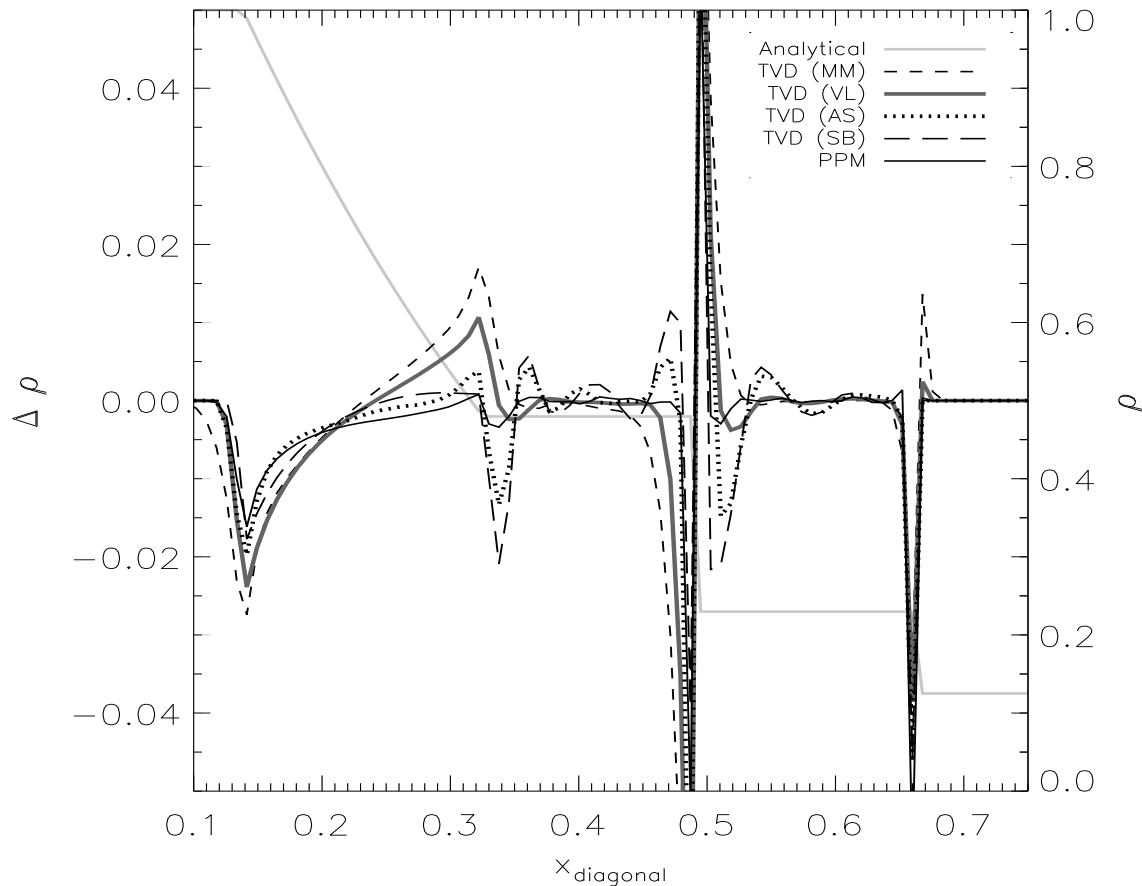
Figure 3.6 Density differences between numerical and analytical solutions for the different codes. Devia-
tions from the analytical solution are measured by the scale on the left. The analytical solution is shown in
the heavy, pale grey line, measured by the scale on the right.

at the edges; this bending increases as the front evolves in time.

3.11.3   Kelvin-Helmholtz instability

In any two fluids with shear (such as that caused by differential rotation), if the velocity gradient in the
shear region is strong enough compared with the stratification, the Kelvin Helmholtz (KH) instability should
develop. This instability is difficult to capture accurately in numerical simulations. Codes like the PPM code
can overproduce the instability's small-scale turbulent structure for a given resolution (Dwarkadas et al.,
2004) while a code with too much diffusion will underproduce such structure. SPH codes also perform
poorly on such a test when there is a density jump across the shearing region. I find a wide range of
small-scale structure production from the different limiter schemes enabling me to control the amount of
small-scale structure in a simulation by choosing an appropriate limiter scheme.

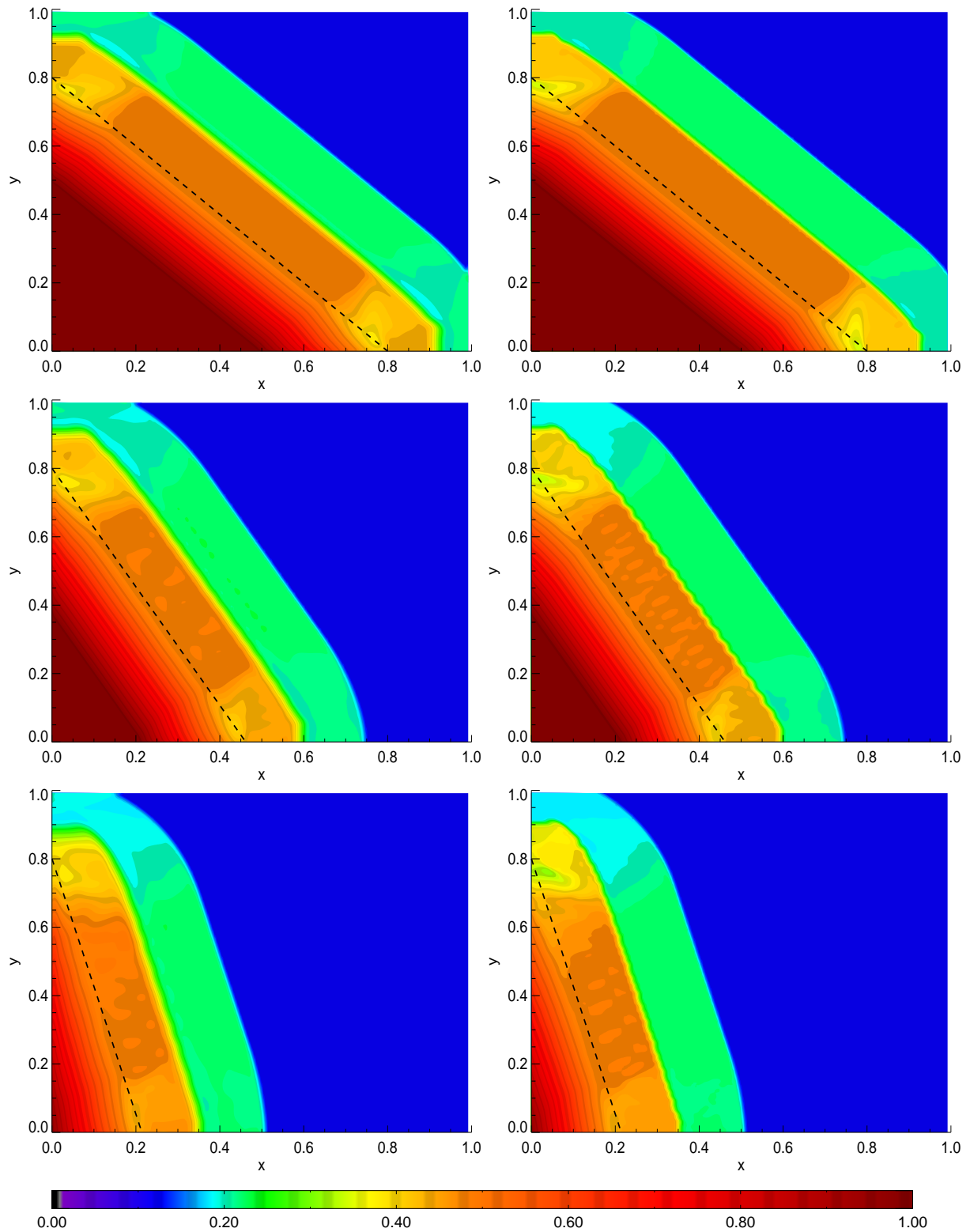The KH instability is initialized on a Cartesian grid of resolution $N_x \times N_y = 400 \times 400$ that is periodic

Figure 3.7 Density contour plots illustrating propagation of an oblique shock at angles, $\theta = 45°$, $\theta = 30°$ and $\theta = 15°$ from top to bottom, respectively. Figures on the left show the TVD(MC) results, figures on the right, the PPM results. The dashed lines show the initial location of the discontinuity.

in the x direction and has reflecting boundaries on the top and bottom. The fluid in the top half of the box is set moving to the left at about one thirteenth of the sound speed and the fluid on the bottom is set moving to the right with the same speed. The densities of the two regions of fluid are set to values of 0.9 and 1.1 on the top and bottom, respectively, in order to help visualize the instability. The interface between the two fluids is initialized to a sine wave in order to excite the instability. Figure 3.8 shows results from the TVD(MB,AS,MC,MM) and PPM codes. Each rows has three panels showing density contour plots, snapshots taken at $t = 0.3$, $t = 0.6$ and $t = 1.5$, respectively. Time is measured in units where the undisturbed fluid propagates completely across the box in unit time.

The superbee limiter proves too unstable in this test and shortly after $t = 0.3$ the time step becomes unreasonably small. The other different limiters produce a range of small-scale structure. The minmod limiter produces the least, developing only a single cusp along the interface and only a single, loose cat's eye structure. The MC limiter has less diffusion and shows hints of small scale instabilities at t=0.3 forming along the interface. The final cat's eye is more tightly wound and shows increased substructure as well. Finally the AS limiter produces as much substructure as the PPM code. Indeed at times, $t = 0.3$ and $t = 0.6$, the results look very similar. At time, $t = 1.5$, both codes show very complicated interfaces with much mixing between the two fluid layers. As in the oblique shock test, the PPM code shows evidence of flattening or clipping the density profile. The TVD runs exhibit larger variation in the density, not only near the interface, but also in the bulk regions of the fluids. The density of the PPM code is comparatively very uniform in the bulk of each fluid region, illustrated by the uniform yellow and blue colors in each region.

The KH instability test readily demonstrates the differences between the different limiters and codes. The reason that this test does so is that with an infinitely sharp interface the KH problem is formally ill posed, and the growth rate for infinitely small disturbances is infinitely fast. Because there exists no small-scale cutoff for the dynamics, the numerics themselves dictate the evolution on the smallest scales.

3.11.4 Supersonic flow around a cylinder

In order to test the implementation of the Euler equations on a cylindrical grid, I examined the formation of a bow shock caused by supersonic flow around a cylinder. I initialize a cylindrical grid spanning an annular region from $2 \leq r \leq 20$, $0 \leq q \leq 2\pi$ with resolution, $N_r \times N_q = 600 \times 150$, with a supersonic flow to the left at three times the sound speed. The initial density and pressure on the grid are uniform. The density is set equal to the adiabatic index, $\rho = 5/3$, and the pressure is set to $p = 1$. The inner boundary of the grid is reflecting, simulating a solid cylinder around which a bow shock forms. The outer boundary allows outflow. Figure 3.9 shows the results from the TVD(VL) and PPM simulations.

The two codes both produce a density peak just in front of the cylinder with a value close to 6 (about 6 and 6.4, respectively, for the PPM and TVD(VL) codes). These peak densities are about 90% and 96% of the post-shock density of $\rho_s = 4\rho_0$ for an adiabatic shock of index, $\gamma = 5/3$. Again the TVD code has small-scale oscillations in front of the shock, which the PPM code appears to have flattened. Both codes capture the smaller tail shocks produced at the rear of the cylinder.
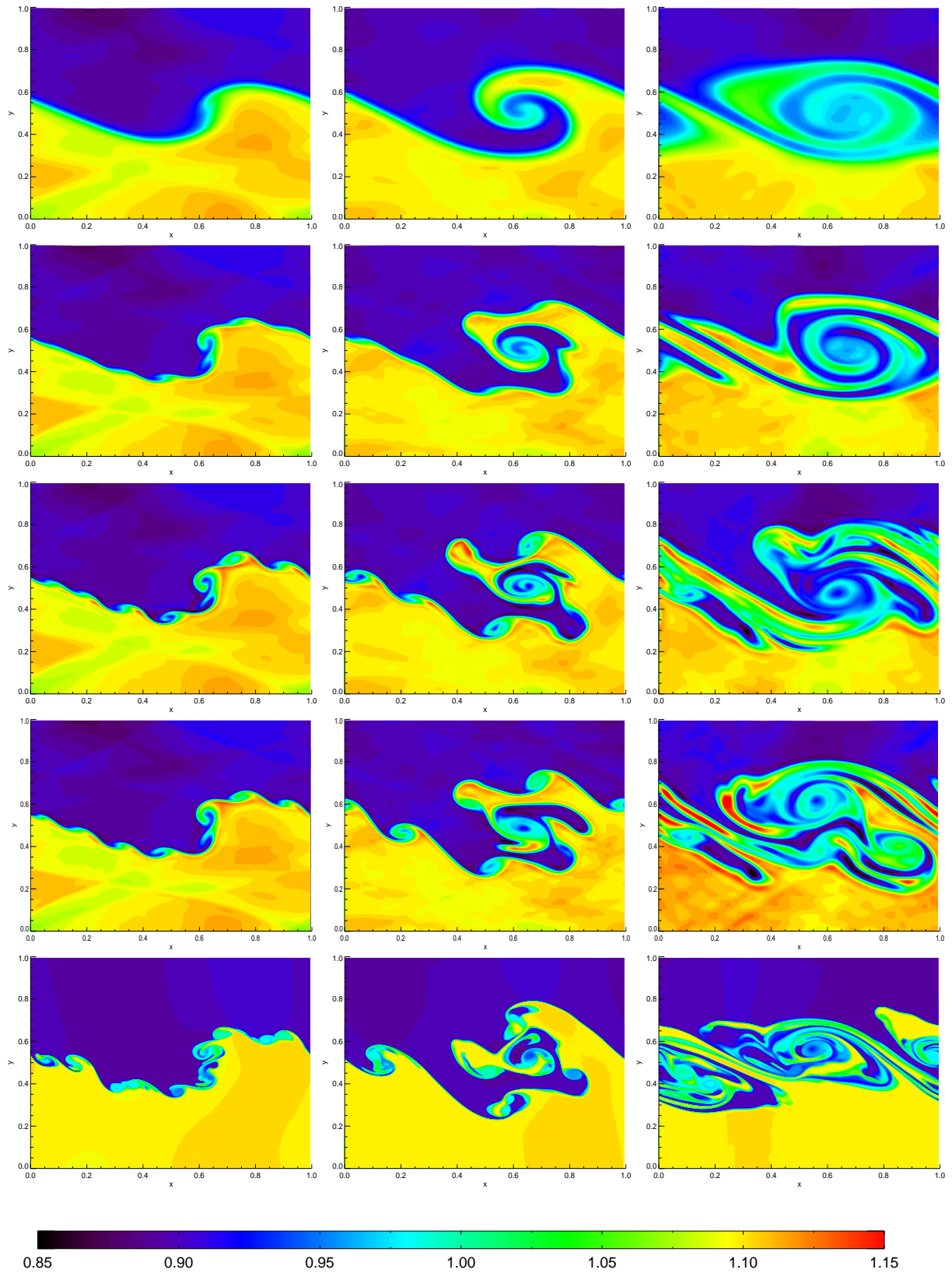
Figure 3.8 Density contours for a KH instability. Each row shows a contour plot of the density taken at $t = 0.3$, $t = 0.6$ and $t = 1.5$, respectively. The first four rows show results from the TVD code using the minmod, MC, AS and mixed limiters, respectively. The last row shows results from the PPM code.
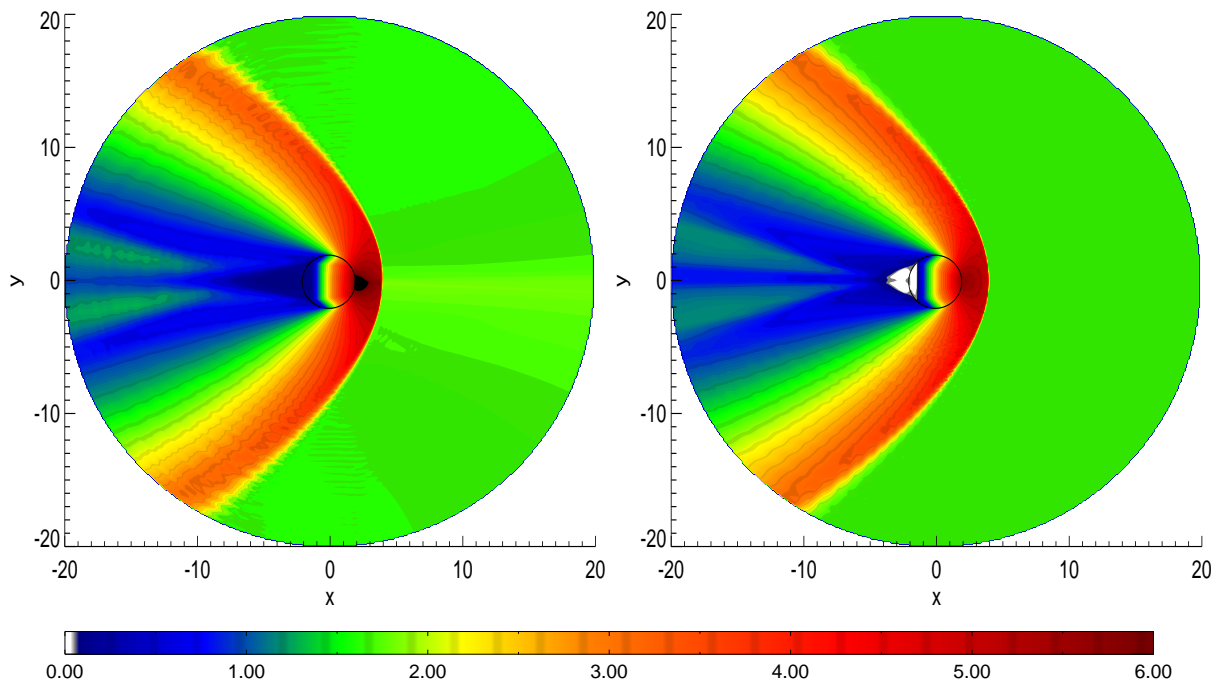
Figure 3.9 Density contour plots of the steady-state density. The image on the left shows results from the TVD(VL) code and the image on the right, that from the PPM code. The black circles represent the boundary of the cylinder.

Chapter 4

The Protoplanet Comparison Problem

4.1 Introduction

In this Chapter, I describe the standard setup where a single protoplanet is embedded within a gas disk. I provide a systematic comparison of results depending on several details and parameters of the implementation. The purpose of such a study is to be able to better determine which details observed within a run are physically realistic and consistent between runs and which are likely due to numerical artifacts. In Section 4.3, I present the reference results of the code on a "Jupiter-mass" planet. Section 4.4 discusses the speedup obtained through the use of the fast advection algorithm as well as differences in the results caused by its use. Differences caused by the choice of two different sets of solution variables are presented in Section 4.5. Sections 4.6, 4.7, and 4.8 examine the influence of viscosity (both numerical and physical viscosity), resolution, and choice of limiter, respectively. I summarize all these results in Section 4.9.

4.2 Problem setup

The details for the setup are the same as those used by de Val-Borro et al. (2006) and represent what is now a standard problem in accretion disk theory. A polar grid is setup with initial conditions (described below) using a given mesh resolution, $N_r \times N_\theta$. In this chapter, the standard run uses $N_r \times N_\theta = 384 \times 384$. The azimuthal range is always taken to be $[-\pi, \pi]$ and unless otherwise indicated, the radial range is $[0.4a, 2.5a]$, where a is the mean orbital radius of the protoplanet. The gravitational potential of the star is calculated for a grid position, $r = (r, \theta)$, in terms of the central star's mass, M_* and location, r_*, as $\phi_* = -GM_*/|r - r_*|$ while the protoplanet is represented through the softened potential,

$$\phi_p = \frac{-Gm_p}{\sqrt{|r - r_p|^2 + \epsilon^2}}. \tag{4.1}$$

The softening length, ϵ, is fixed in terms of the disk's scale height as $\epsilon = 0.6H(a)$. A "Jupiter-mass" planet is defined in terms of the system's mass ratio, $\mu = m_p/(m_p + M_*)$, to have $\mu = 10^{-3}$. Simulations with a

single planet are calculated in the frame corotating with the planet, with the origin at the system's center of mass. This choice of origin means that the star orbits a distance, μa, from the origin, the planet at a distance, $(1 - \mu)a$.

I use dimensionless units where the unit of mass is taken to be the total mass of the system, $M_* + m_p$. Length is measured in units of the planet's initial radial separation from the star, a, and I set the gravitational constant, G, to unity. Time is measured in units of the orbital time,

$$\tau = \sqrt{\frac{a^3}{G(M_* + m_p)}}. \tag{4.2}$$

With this definition, one orbital period takes 2π units of simulation time.

Fiducial initial conditions are those of a uniform-density Keplerian disk, which is the equilibrium solution for a single central potential of mass, M_*. I transform the azimuthal velocity into the corotating frame and correct for pressure support as $u_\theta = \sqrt{(GM_*)/r}[(1 - m_H^2)^{1/2} - (r/a)^{3/2}]$, where $m_H = H/r$ is the disk thickness. In order to allow the Keplerian disk to gradually adjust to the influence of any additional potential, such as that due to an orbiting planet, the potential of the planet is slowly "turned-on" according to the formula,

$$\phi_p(r, t) = \sin^2\left[\frac{t}{4\mathcal{N}\tau}\right]\phi_p(r), \tag{4.3}$$

where t is the simulation time and $\mathcal{N} = 10$ is the number of orbits over which the potential is turned on. After the prescribed number of orbits the full value of the potential, $\phi_p(r)$, is left constant.

4.3   Standard comparison run

Here I present a standard comparison run of the code for a Jupiter-mass planet, run for 300 orbits using the VL limiter scheme. The calculation is performed advecting the solution set $(\rho, \rho u_r, H)$, where $H = \rho(ru_\theta + r^2\Omega)$ is the absolute angular momentum (combined gas and frame) in the corotating frame.

In Figure 4.1 I show density contours after 5, 10, 20, 50, 100 and 300 orbits. Note the appearance of the spiral arms occurs very quickly—within a few dynamical times at most. There are two trailing arms outside the planet's orbital radius and three arms inside the orbital radius. They are close to steady-state in the sense that they occur at fixed locations within the disk when time-averaged over a few orbits, although they exhibit small spatial and temporal oscillations in simulations with low-viscosity.

As the simulation progress, the planet begins to clear gas from its orbit, but not uniformly. Gas is more readily cleared in two orbital tracks $1 - 2$ Hill radii, $R_H \equiv (\mu/3)^{1/3}$, to the inside and the outside of the planet's orbit. These locations are the approximate distance at which the averaged torque density due to neighbouring resonances peaks (around resonance order, $m = 10$; see Ward (1996) for details). There is a further asymmetry in the efficiency of the gas clearing for locations trailing and leading the planet as demonstrated in Figure 4.2. It shows the density, averaged along the full azimuth direction at five different
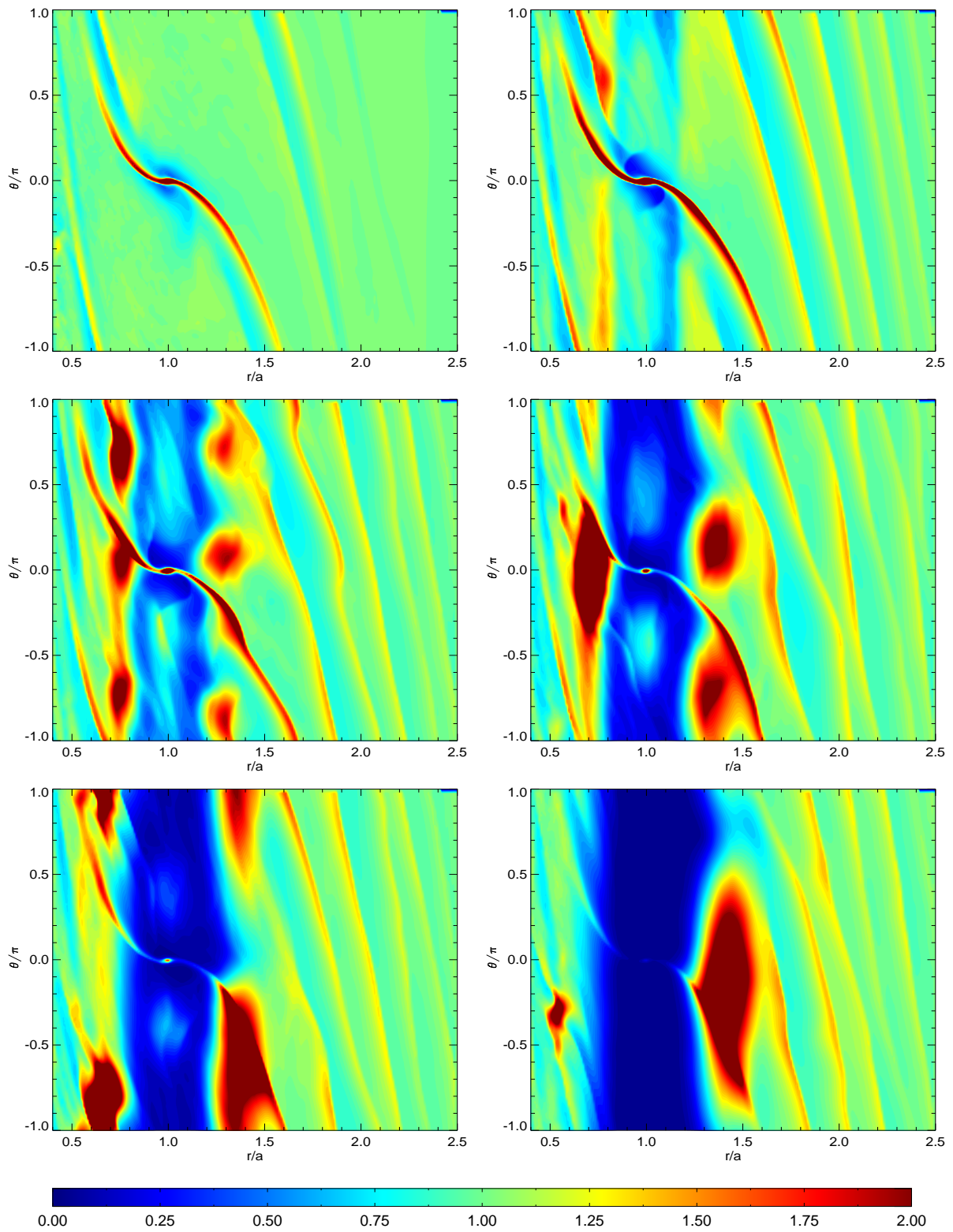
Figure 4.1 Density contours for a standard Jupiter mass run using the VL limiter scheme. Each plot in sequence shows a contour of the density taken after 5, 10, 20, 50, 100, and 300 orbits, respectively.
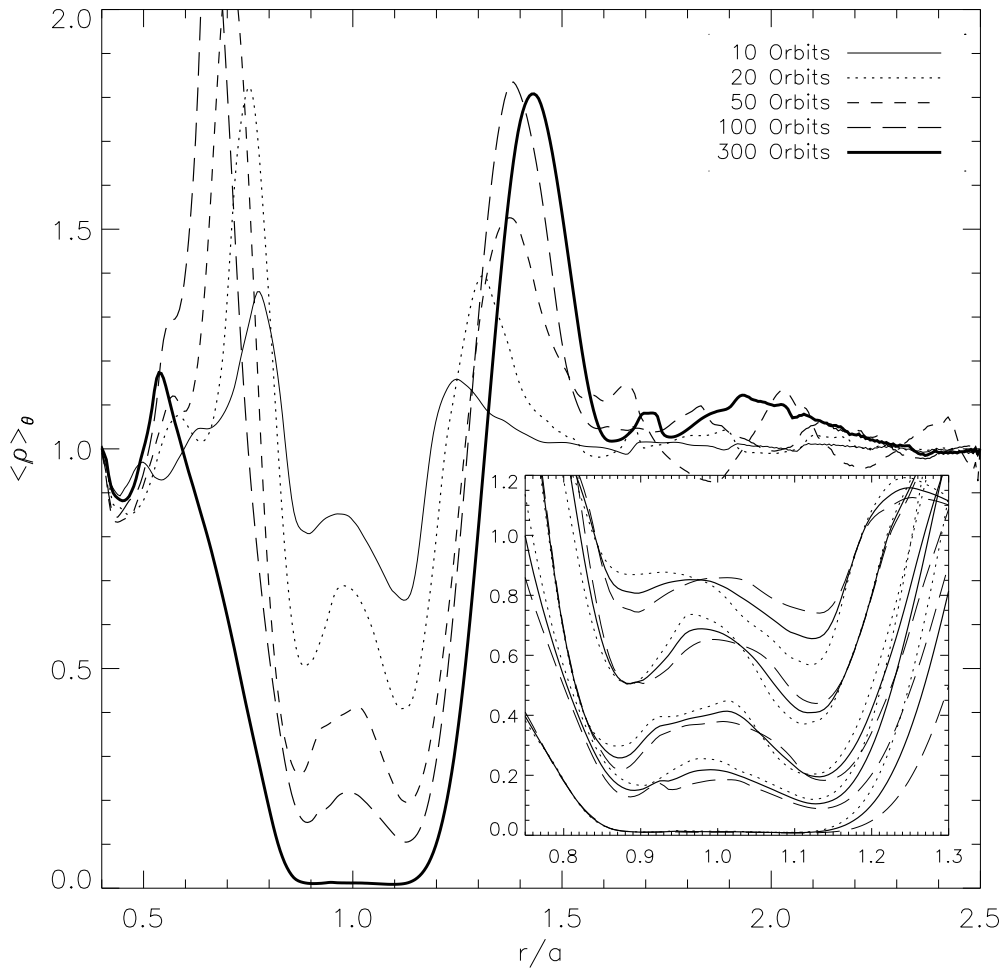
Figure 4.2 Azimuthally averaged density profile of the standard Jupiter mass run using the VL limiter scheme. The inset displays the same averaged densities for the 5 different times in the solid curves as well as averages for $\theta > 0$ in the dotted curves and $\theta < 0$ in the dashed curves.

times during the simulation. Especially during the initial formation of the gap, the total density in the track outside the planet's orbit is much lower relative to the density in the track inside its orbit. In the inset, solid lines show the averaged density as before, while the broken lines show the averages separated into halves for $\theta > 0$ (dashed) and $\theta < 0$ (dotted). While the gap region leading the planet seems to clear approximately equally inside and outside the planet's orbit, the region trailing the planet to the outside clears more quickly than elsewhere and the region trailing the planet to the inside clears more slowly.

Note that there are regions of fluid within the gap which persist over time. These regions surround the L_4 and L_5 Lagrangian equilibrium points located at $\theta = \pm\pi/3$. In Figure 4.1 these are the circularly shaped overdensities which remain within the gap. I illustrate the evolution of these regions in Figure 4.3. The density plotted has been radially averaged at each azimuth from $r = 0.9a$ to $r = 1.1a$. Again there is a
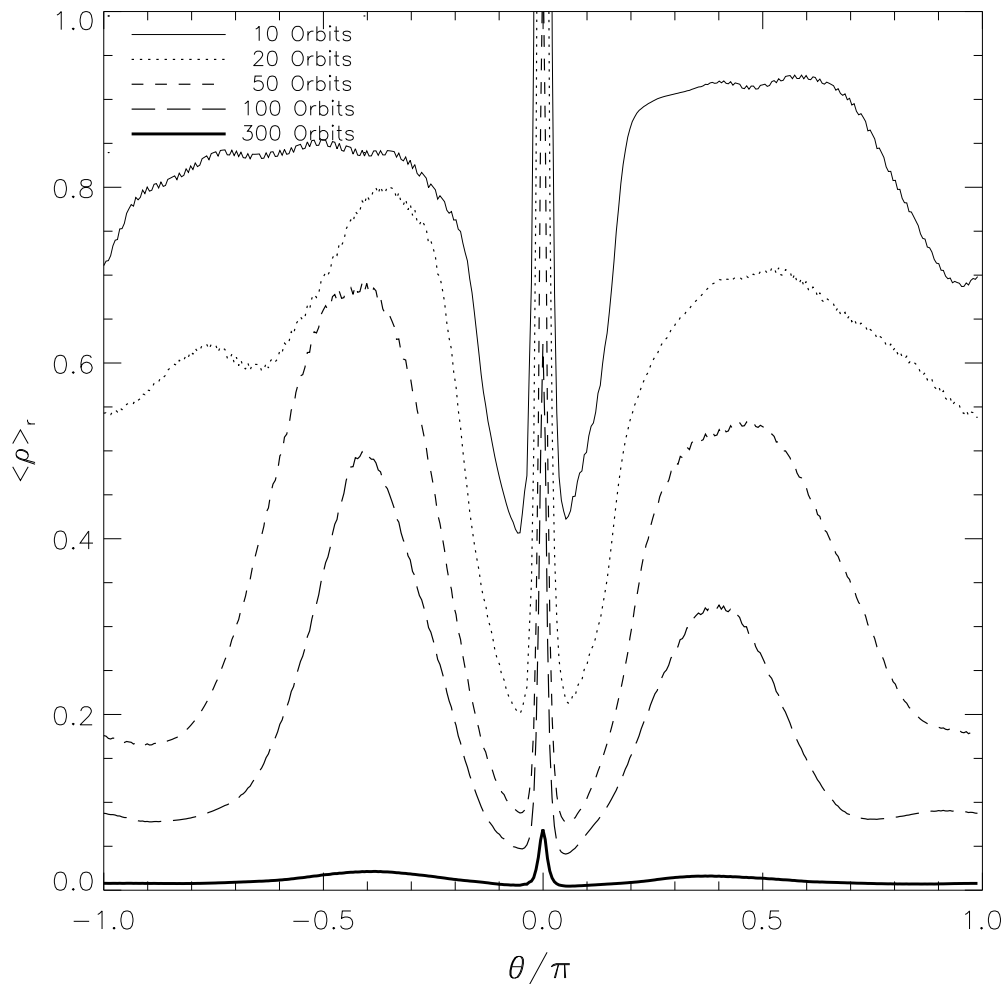
Figure 4.3 Average density within the gap region for the standard Jupiter-mass run. The density within 10% of the planet's orbital radius has been radially averaged. Note the asymmetry between the L_4 and L_5 equilibrium points.

leading-trailing asymmetry. This asymmetry has been observed in most planet codes to varying degrees (see de Val-Borro et al., 2006).

In addition to the above libration islands at the L_4 and L_5 points, there are large over-densities which begin to develop due to the generation of vortices to either side of the gap region. Initially several small vortices develop at roughly the same radii, but spaced in azimuth. These vortices appear as roughly concentric overdensities in the density contour plots (in Fig. 4.1 at 20 orbits there are three such regions at both $r/a = 0.75$ and $r/a = 1.3$). As the simulation progresses, the vortices grow and begin to merge, depending on their radial location within the disk. In disks with large viscosity, ($\nu \gtrsim 10^{-5}$), the vortices do not form. Similar structures have been observed in other codes at low viscosity. The vortices are likely the result of Rossby-wave instabilities (see Papaloizou & Lin, 1989; Lovelace et al., 1999; Li et al., 2000, 2001, and
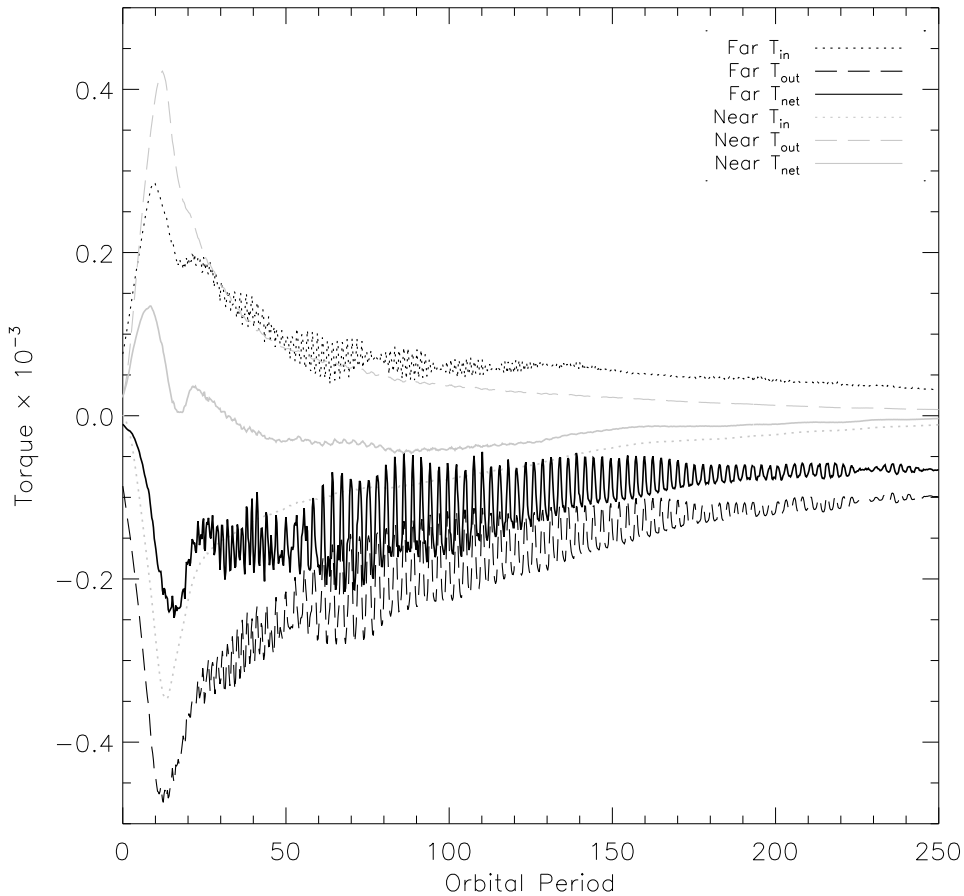
Figure 4.4 Time evolution of the torque on the planet's orbit due to the disk. The torque is broken up into components from the disk material inside (dotted) and outside the planet's orbits (dashed). Black lines show the two components and the net torque from all the disk material more than one Hill radius, $R_H = (\mu/3)^{1/3}$, away from the planet. Grey lines show the two components and the net torque due to material between $0.5R_H$ and R_H.

references therein). I discuss their generation and evolution more thoroughly in Section 5.2.

In Figure 4.4 I present the total torque summed over various regions of the disk, showing its evolution over the course of the simulation. A running average over a period of 10 orbits has been performed to smooth out some of the oscillations. As per the treatment in de Val-Borro et al. (2006), material within the Hill sphere is considered separately, mimicking the effect of the torque-cutoff. The gas within this region feels the softened gravitational potential of the planet, rather than the long-range singular potential. As theoretically predicted (Goldreich & Tremaine, 1980; Ward, 1986), the torque inside the planet's orbit is positive (transferring angular momentum to the planet), while that outside the planet's orbit is negative (angular momentum is transferred from the planet to the exterior disk). Also as predicted, there is an asymmetry in the magnitudes of these torques (Ward, 1996). The net torque on the planet is negative and

would cause it to migrate inwards, were its orbit not held fixed.

The initially smooth, broader scale oscillations of the total torque (on a time scale of ~ 20 orbits) are consistent with Phase I evolution, as described by Koller et al. (2003). The frequency of the subsequent rapid variations which develop matches the inverse period of the large vortex outside the planet's orbit as measured in the frame of the planet.

4.4   Speed-up due to the fast advection algorithm

Using the fast-advection algorithm discussed in Section 4.4 decreases the number of iterations required to perform a simulation. This result is accomplished by subtracting off the average background flow in the azimuthal direction, and advecting fluid quantities via the TVD method by only the residual azimuthal velocity atop of the background flow. The subsequent transport of these quantities due to the background flow may be performed without limiting the time step allowed by the CFL condition. The simulation time is reduced approximately by the ratio of the residual azimuthal velocity to the full azimuthal velocity. This reduction comes at the expense of increased diffusion introduced by the transport of quantities by the background flow. In order to determine the effect of this diffusion on my results, I compare the standard run with and without the fast advection routine implemented.

In Figure 4.5 I show density plots taken after 20, 100 and 300 orbits comparing a run which uses the fast-advection routine and one which does not (I refer to the latter as the pure TVD run). The run which uses the routine requires 7.8 times fewer iterations to reach 100 orbits and finishes 6.6 times faster. It is evident from the density plots that the algorithm increases the diffusion of the simulation; nonetheless, the important structures of the flow are still apparent. The spiral arms show nonsteady-state behavior and the libration islands of fluid are present in both runs. Vortices form on the same radial lines in both runs and eventually merge (the two separate vortices present in the pure TVD run have merged by 400 orbits). Both results also show narrow tendrils of gas threading the gap region.

I compare results for the azimuthally averaged density and the radially averaged density in Figure 4.6. Again the increased diffusion is apparent, but the results are otherwise similar.

4.5   Dependence on the chosen solution set

Kley (1998) has shown that advecting the absolute angular momentum, $\rho r(u_\theta + r\Omega)$ (that of both the fluid and the frame), as written in equation (2.18) produces better results than just advecting the quantity $\rho r u_\theta$. In practise this distinction is between accounting for the Coriolis and centripetal accelerations terms using the hydro portion of the code (the TVD method in my case) or as source terms where they are solved via the Runge-Kutta method. I perform a comparison between a simulation which uses the solution variables $(\rho, \rho u_r, H)$ and one which uses the solution variables $(\rho, \rho u_r, H/r)$, that is between one which solves for the absolute specific angular momentum and one which solves for the absolute specific angular velocity. As one
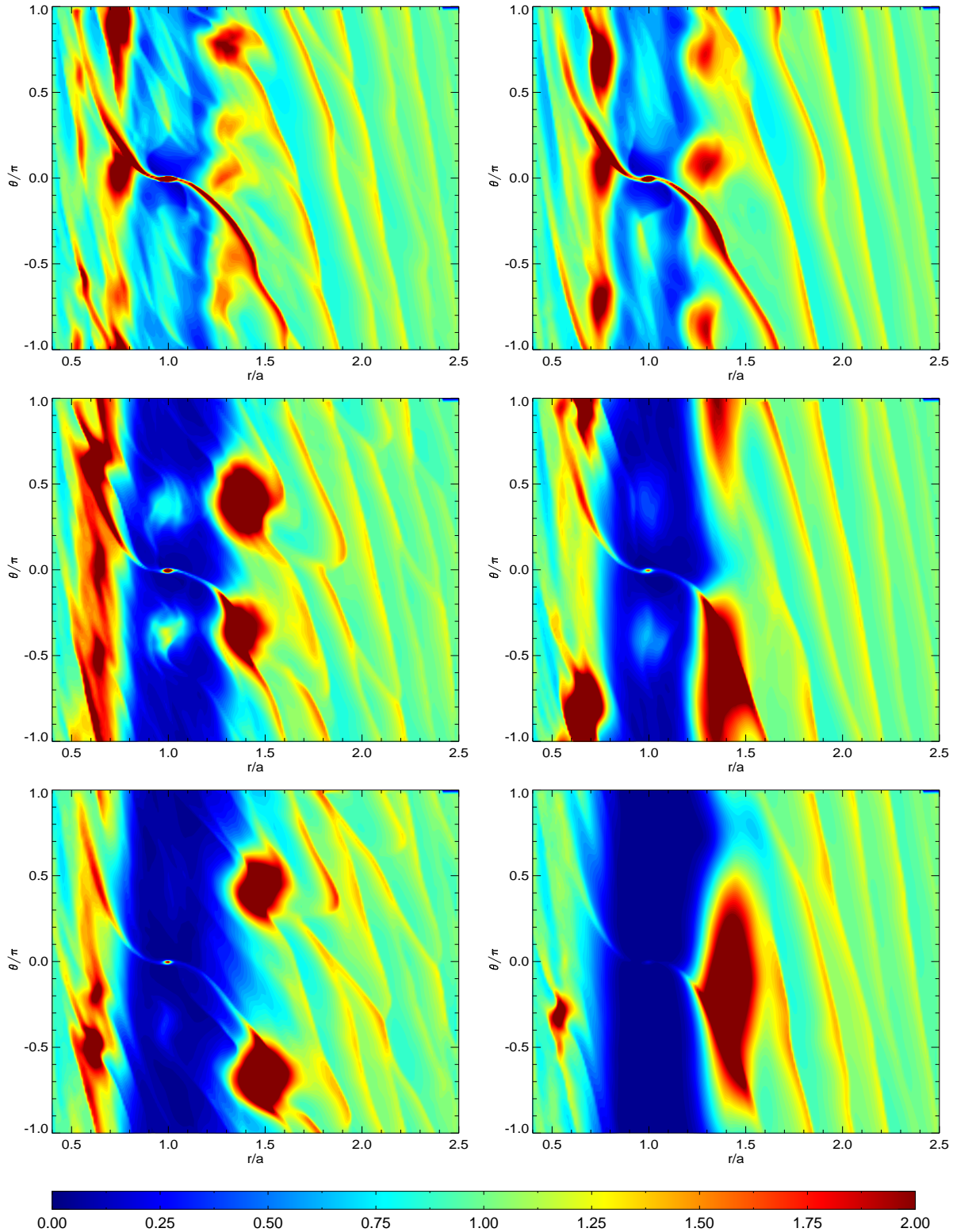
Figure 4.5 Snapshots of the density after 20, 100 and 300 orbits (top, middle and bottom rows). Results from the standard run are on the right; results from the pure TVD run are on the left. The two vortices still present at 300 orbits in the pure TVD run have merged into a single vortex by 400 orbits as in the standard run.
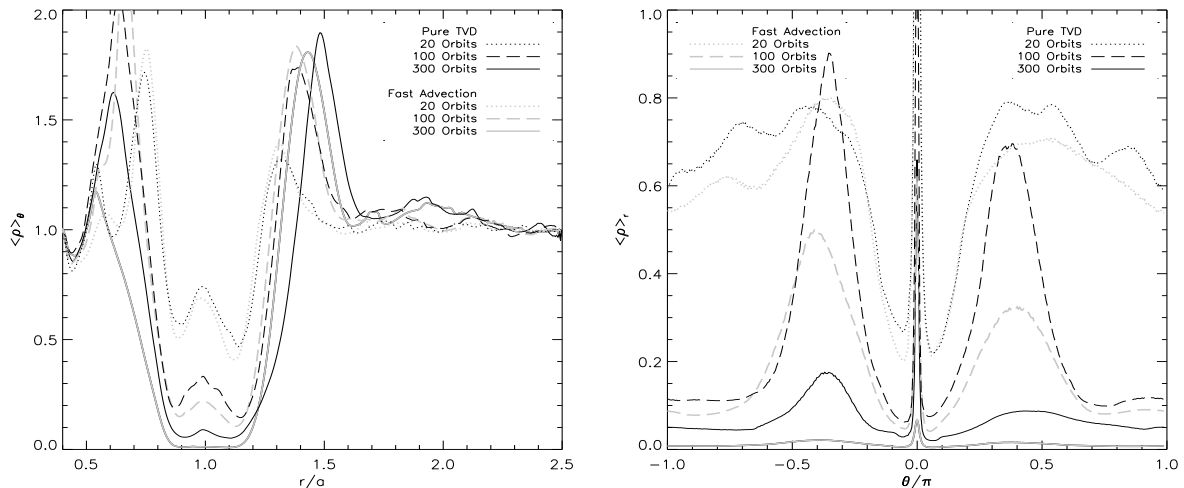
Figure 4.6 Comparisons of the azimuthally averaged density and the radially averaged density in the gap region. Results drawn in dark lines are from the pure TVD run; those in the pale lines are from the standard run. As expected the use of the algorithm increases the diffusion in the simulation.

might expect, the latter yields poorer results, since the global angular momentum is not formally conserved in such a case.

In Figure 4.7 I compare the density of the solution sets after 10, 50 and 100 orbits. While both properly capture the locations of the spiral arms, the formulation which conserves angular momentum appears to have less diffusion than the one which does not. Figure 4.8 shows that while the angular velocity scheme captures the gross details of the gap such as its width and depth, other structures in the flow, such as the L_4 and L_5 libration islands, and the vortex lines are missing. In addition, the actual preservation of mass and angular momentum over the course of the simulation differ: the scheme that uses angular velocity as a solution variable conserves mass and angular momentum over 100 orbits to about 4% and 3%, respectively; the scheme that uses angular momentum as the solution variable conserves these quantities to within 2% and 1%, respectively.

4.6   Effects of viscosity

The presence of physical viscosity tends to smooth perturbations of physically conserved quantities. Parametrizing the viscosity for a flat, alpha-disk model (§2.3.2) results in a relationship between the viscosity coefficient, ν, and turbulent efficiency, α, given by $\nu = \alpha m_H^2 \sqrt{GM_* r}$ where m_H is the disk thickness. For an expected range of α-values taken to be $\alpha = 10^{-2} - 10^{-3}$ (Hartmann et al., 1998) and a disk thickness of $m_H = 0.05$, one finds $\nu \approx 10^{-4.5} - 10^{-5.5}$. Figure 4.9 shows the evolution of the density for a simulation with $\nu = 1 \times 10^{-5}$. The physical viscosity is implemented by direct differencing, accurate to order $(\Delta x)^2$, of the stress tensor terms as written out in Appendix A. While the evolution of the spiral arms occurs on the
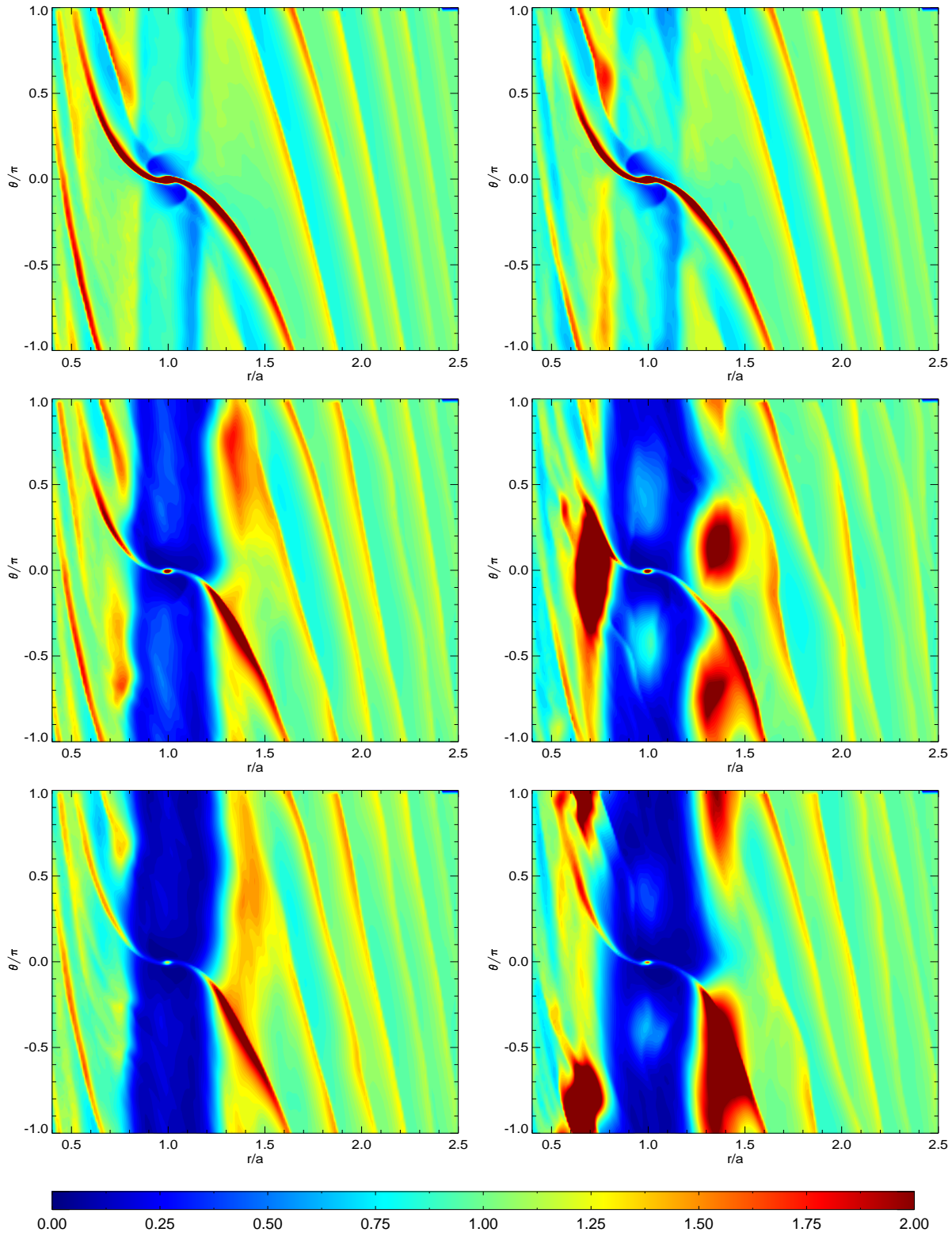
Figure 4.7 Snapshots of the density comparing results after 10, 50 and 100 orbits for two different sets of solution variables. Plots on the left are for the solution set, $(\rho, \rho u_r, \rho(u_\theta + r\Omega))$, while those on the right are for $(\rho, \rho u_r, \rho r(u_\theta + r\Omega))$.
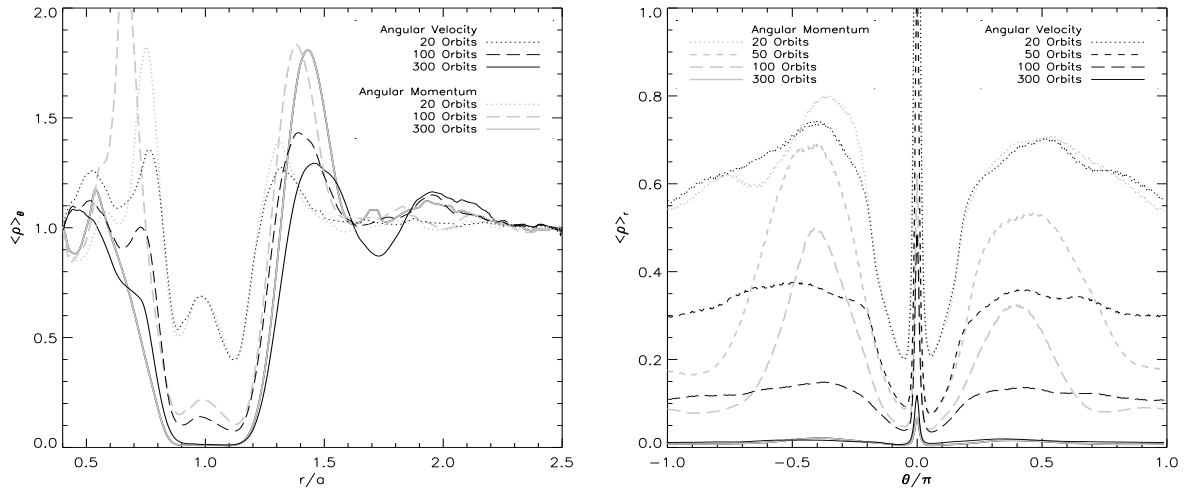
Figure 4.8 Comparisons of the azimuthally averaged density and the radially averaged density in the gap region. The runs drawn in dark lines use angular velocity as a solution variable, while those in the pale lines use angular momentum as a solution variable. While the gap profiles are close in the radial direction, they are substantially different in the azimuthal direction.

same timescale, many of the structures in the simulation are no longer present with this additional physical viscosity: the L_4 and L_5 libration islands are less marked and the vortex lines seen previously are absent. Figure 4.10 compares the radially and azimuthally averaged densities at several orbital times for runs with and without added physical viscosity. While it appears that there may be libration islands of fluid in the viscous run which begin to develop, their radial and azimuthal structure is smoothed out by the viscosity. Note that if the added physical viscosity is reduced in order by another half-magnitude, the libration islands are once again present (see below).

Any numerical algorithm exhibits numerical viscosity due to the combined results of diffusive and dispersive errors (discussed in §3.2). While physical viscosity is characterized by the form of the stress-strain tensor (in a Newtonian fluid there is a linear relationship between the two), an algorithm's numerical viscosity can differ from the physical viscosity, not only in the amplitude or spatial dependence of the viscosity coefficient, but also in the relationship between the stress and strain. Except for the most diffusive schemes, the numerical viscosity of an algorithm usually displays a nonlinear dependence on the velocity gradient. These higher order terms tend to introduce dispersion.

Despite this potential incongruity between physical and numerical viscosity, it is useful to have an estimate for the value of the viscosity coefficient at which the two viscosities may be considered approximately equal in their effects. In order to determine this value, I compare the results of several simulations with various levels of added physical viscosity (implemented as described at the beginning of this section). By reducing the value of the viscosity coefficient, ν, to a point where the results of the simulations are approximately the same irrespective of its addition, I obtain an estimate for the numerical viscosity present in the
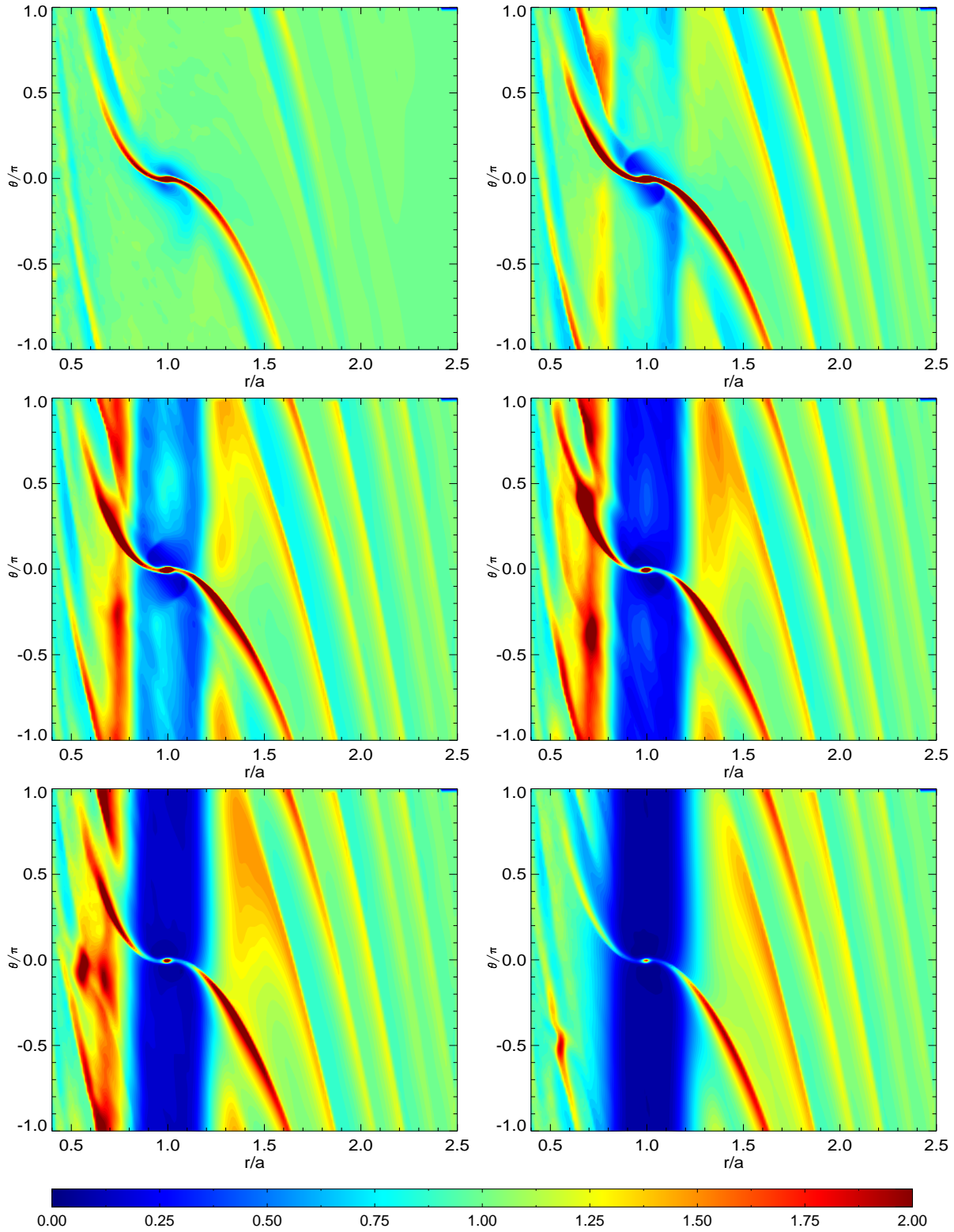
Figure 4.9   Snapshots of the density for a standard Jupiter mass run using the VL limiter scheme with a viscosity of $\nu = 1 \times 10^{-5}$ ($\alpha \approx 0.004$). Each plot in sequence shows a contour of the density taken after 5, 10, 20, 50, 100 and 300 orbits, respectively.
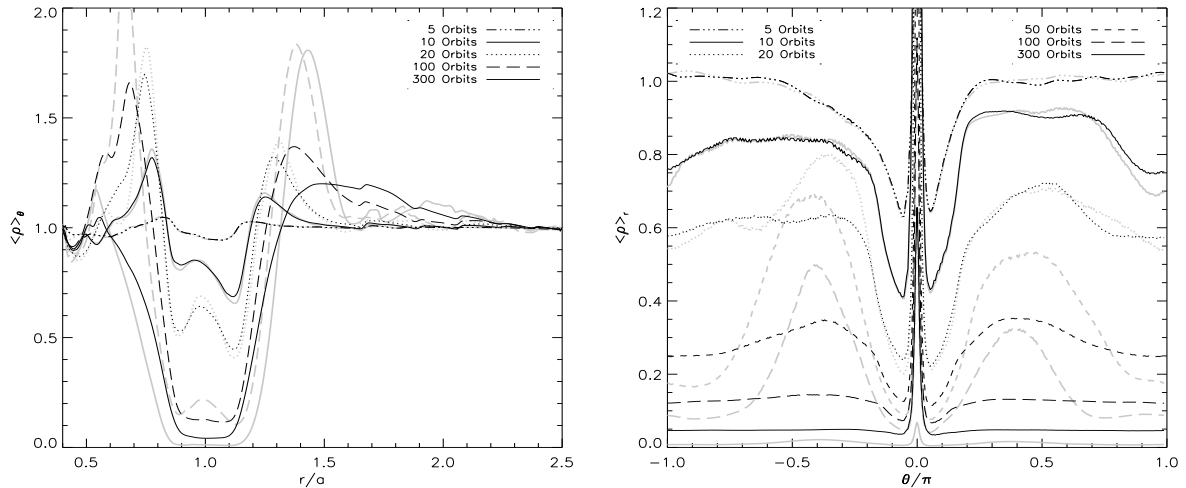
Figure 4.10 Comparisons of the azimuthally averaged density and the radially averaged density in the gap region. The dark lines have $\nu = 1 \times 10^{-5}$; the pale lines have no added viscosity. The viscosity makes the islands of fluid surrounding the L_4 and L_5 points unmaintainable.

simulation.

Figure 4.11 shows the results of decreasing the value of the physical viscosity coefficient from $\nu = 3 \times 10^{-4}$ to $\nu = 1 \times 10^{-6}$ in roughly half-magnitude increments. These simulations are performed using the VL limiter scheme run for 100 orbits. Increasing the level of viscosity present narrows the gap width, decreases its depth and softens the density gradient at its edges. Even when introduced at a level of 10^{-6} there is a difference in the averaged density profile, especially in the gap region. This suggests the numerical viscosity of the code is of approximately the same magnitude or less. The results using other limiter schemes are analogous and suggest a similar level of diffusion with more or less dispersion. They are discussed further in Section 4.8. Comparison with Figure 4.8 suggests that the numerical viscosity using the alternate solution variables, $(\rho, \rho u_r, \rho(u_\theta + r\Omega))$, is more than an order of magnitude higher.

In Figure 4.12 I show the variation of the torques with viscosity. All the torques are similar for the three lowest values of added viscosity. Only at values of $\nu = 1 \times 10^{-5}$ or larger are the differences discernable—the torques from the inner and outer parts of the disk both increase in magnitude, but the net torque decreases for large enough viscosities. In addition, the rapid oscillations damp beyond $\nu = 10^{-5}$ because the large outer vortex is no longer able to form. The increase of material within the gap region with larger viscosity could explain the increase in magnitude of the inner and outer torques. With a large enough viscosity, the asymmetry of the density profile causing the inner and outer torques on the planet are smoothed out, and the net torque decreases.
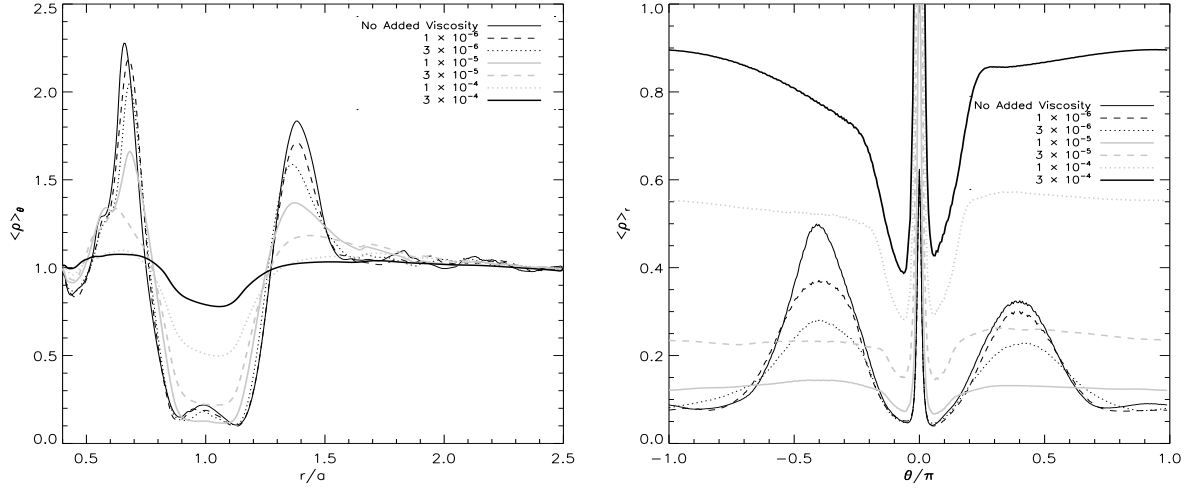
Figure 4.11 Comparisons of the azimuthally averaged density and the radially averaged density in the gap region at different values of added physical viscosity. All results use the VL limiter scheme and are taken at 100 orbits. The addition of physical viscosity, even at the level of $\sim 10^{-6}$, alters the results of the simulation, especially within the gap region, suggesting the numerical viscosity of the simulation is at the same level or lower. Comparison with Figure 4.8 suggests that the numerical viscosity using the alternate solution variables, $(\rho, \rho u_r, \rho(u_\theta + r\Omega))$, is more than an order of magnitude higher.

4.7   Effects of resolution and evidence of numerical convergence

The effects of numerical viscosity become more pronounced at lower resolutions. Figure 4.13 shows the azimuthally and radially averaged densities for several different resolutions above and below that of the standard run. The effects of increasing the resolution are most apparent inside the planet's orbit. Low radial resolution appears to make the slope of the gap shallower on the inside edge of the planet's orbit. For the two runs with the lowest radial resolution, $N_r = 128$, the libration islands of fluid do not exist—likely the numerical viscosity at these resolutions is too large for them to be maintained. Also, as the radial resolution increases, two distinct vortex lines and corresponding overdensities (at approximately $r = 0.55$ and $r = 0.70$) become apparent inside the planet's orbit, rather than just a single line, or none.

Note that increasing the azimuthal resolution relative to the radial resolution widens the gap profile. At resolutions where the results have not yet converged, it also effects some of the details of the structures presents within the disk, the vortex lines and the the libration islands, in a more complicated manner, because the number of iterations required to reach 100 orbits differs amongst differing resolutions by as much as a factor of three. Thus, the effects of numerical resolution are the result of a competition between an increased amount of diffusion from an increased number of iterations required, and a decreased amount of numerical viscosity due to the increased grid resolution.

Figure 4.14 shows the torques as the resolution is varied. The runs with the three or four highest resolutions are consistent with one another. Note that while the torques over the inner and outer parts are still
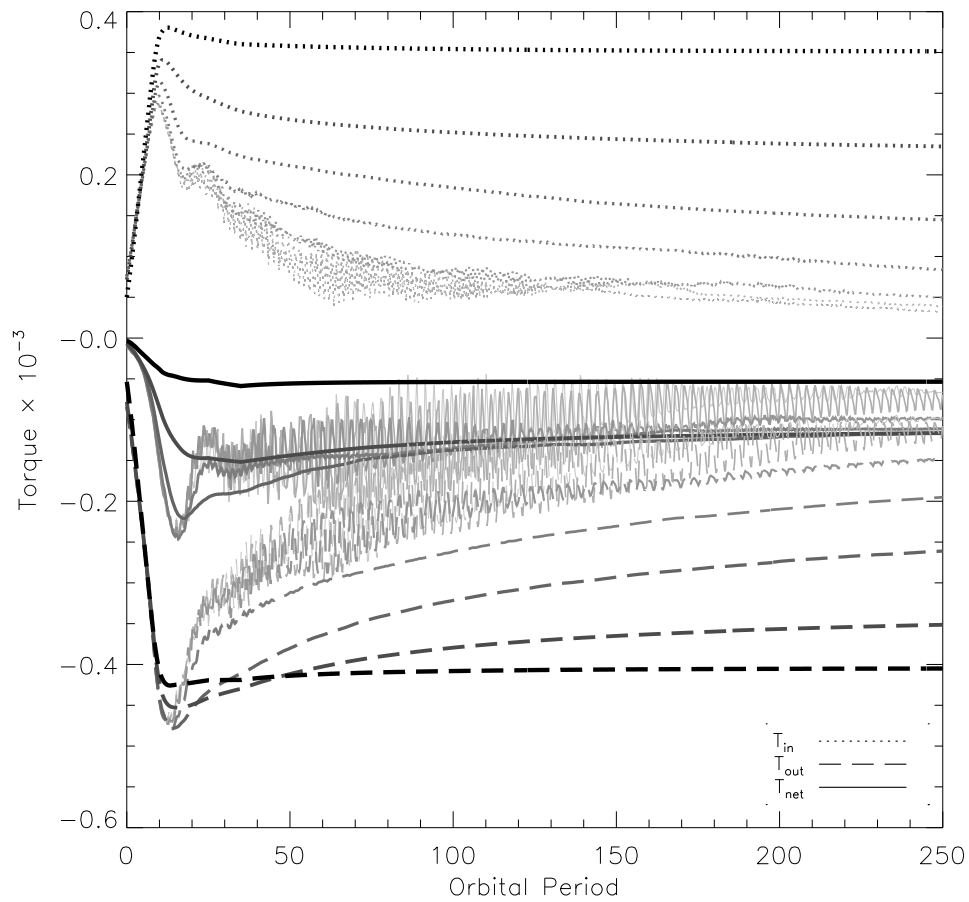
Figure 4.12 Time evolution of the torque on the planet's orbit due to the disk. The torque is broken up into components from the disk material inside (dotted) and outside the planet's orbits (dashed). Also plotted is the total net torque. The torque calculated is only that from material further out than one Hill radius from the planet. The palest lines have no added viscosity and from there the viscosity increases in half-magnitude increments from $\nu = 1 \times 10^{-6}$ to 3×10^{-4} for the heavy black lines.

Figure 4.13 Comparisons of the azimuthally averaged density and the radially averaged density in the gap region for different resolutions. All results use the VL limiter and are taken at 100 orbits. Results in the top row have no added viscosity, those in the bottom row have $\nu = 10^{-5}$.

increasing slightly in magnitude with higher resolution, the total torque remains the same, except for the two lowest resolution runs.

4.8   Effects of the limiter scheme

The results from the Sod shock tube test and the Kelvin-Helmholtz instability in Section 3.11 have already illustrated that the different limiter schemes show different levels of numerical viscosity—both differing diffusion and dispersion. Figure 4.15 shows the azimuthally and radially averaged densities for the minmod, VanLeer, MC, mixed (MB) and superbee schemes used in the standard run. As in Section 3.11, the superbee scheme shows the least diffusion, but the most dispersion. This is evident from the increased density of
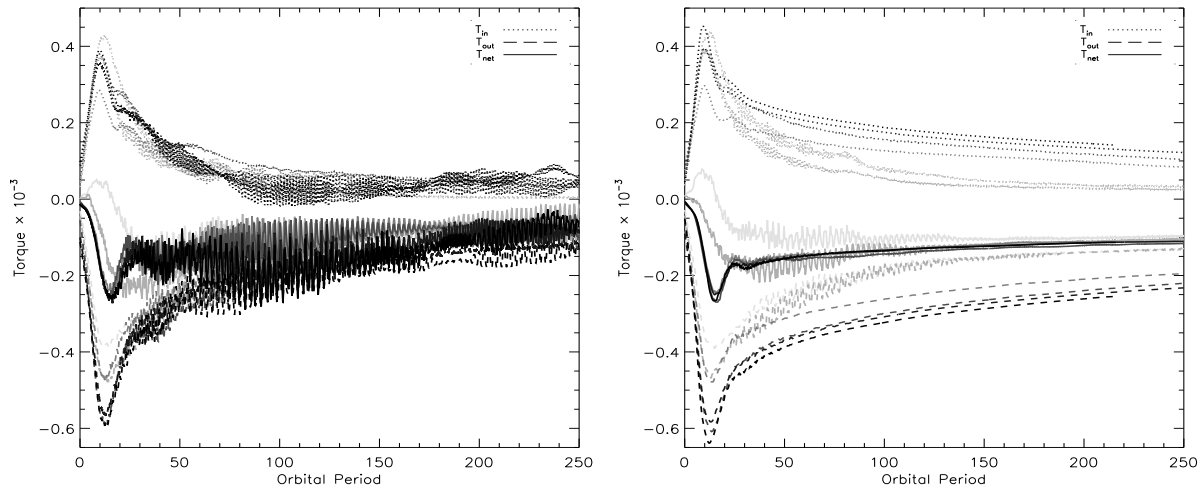
Figure 4.14 Time evolution of the torque on the planet's orbit due to the disk as a function of resolution. The graph on the left has no added physical viscosity; the graph on the right has $\nu = 1 \times 10^{-5}$. The torque is broken up into components from the disk material inside (dotted) and outside the planet's orbits (dashed). Only material further than one Hill radius from the planet is included in the calculation. The palest lines show the lowest resolution of the 5 runs. The darkest lines show the highest resolution.

the fluid in the gap region, as well as its oscillatory structure. The mixed and MC schemes also show a fair amount of dispersion. The VanLeer scheme seems to show diffusion comparable to that of the MC and mixed schemes but substantially less dispersion. As before the minmod scheme show the most diffusion.

Figure 4.16 shows the resulting torques for each limiter scheme. These are consistent with the previous results.

4.9   Summary

The above comparisons shown that the modified TVD code produces results consistent with a wide variety of other codes examined in the protoplanet comparison project (de Val-Borro et al., 2006). In particular I confirm the existence of libration islands at the L_4 and L_5 points, the asymmetry in the density of those islands, the sign and magnitude of the torque exerted on the planet by the disk and the growth and merging of vortices outside the planet's orbit at low viscosity.

I find that the large vortex which forms outside the planet's orbit causes substantial torque oscillations on the planet. These oscillations correspond to repeated passes of the vortex by the planet. Increasing the viscosity beyond $\nu \geq 10^{-5}$ damps the formation of the vortex thereby removing the oscillatory signature from the torque. The existence of additional *vortex lines* to the inside of the planet's orbit are demonstrated.

Through use of the *FARGO*-like fast-advection algorithm, I achieve a reduction in the required simulation time by a factor of 6.5. I also find that use of the absolute angular momentum rather than angular velocity as a solution variable decreases the simulations numerical diffusion by over an order of magnitude.
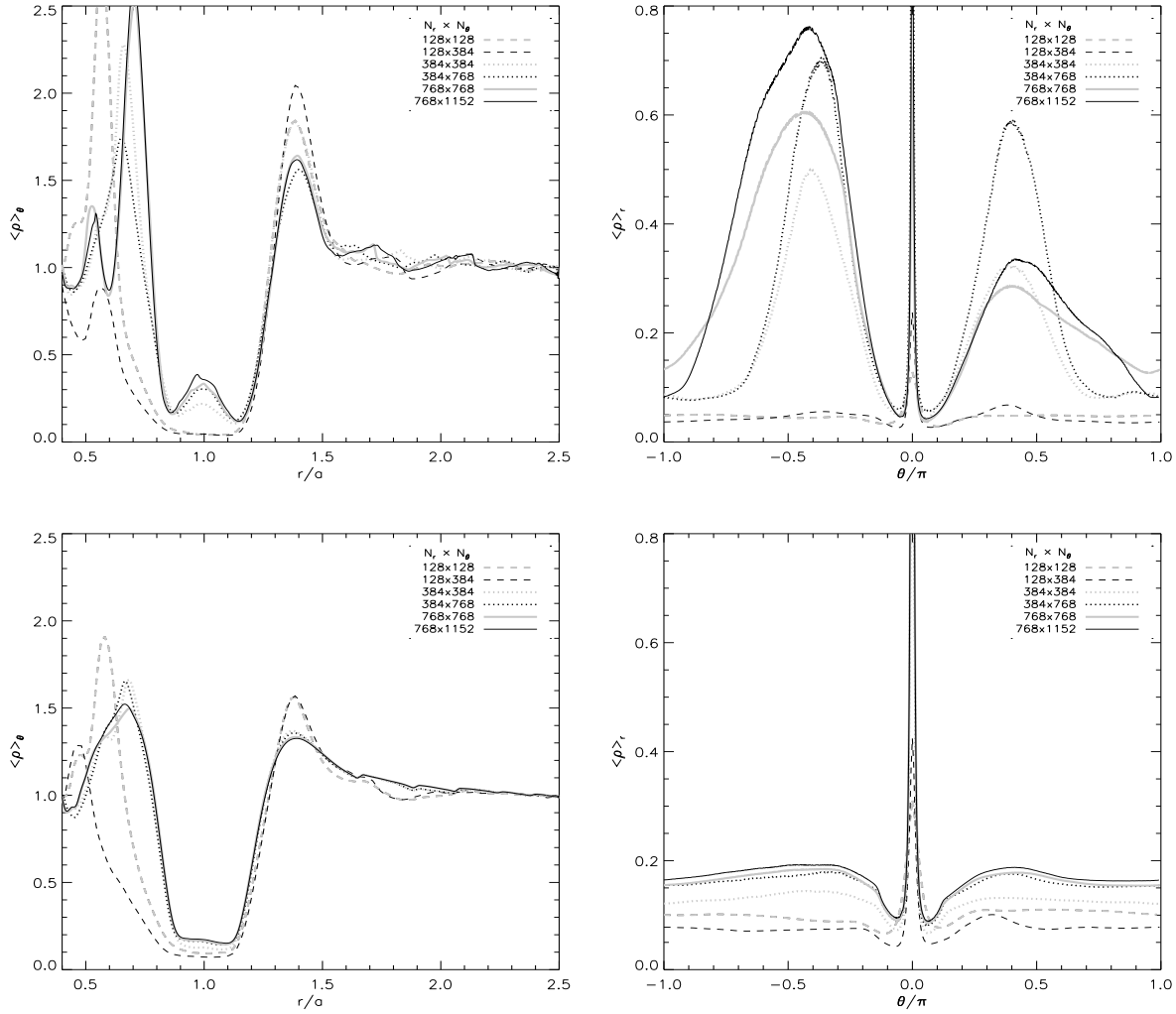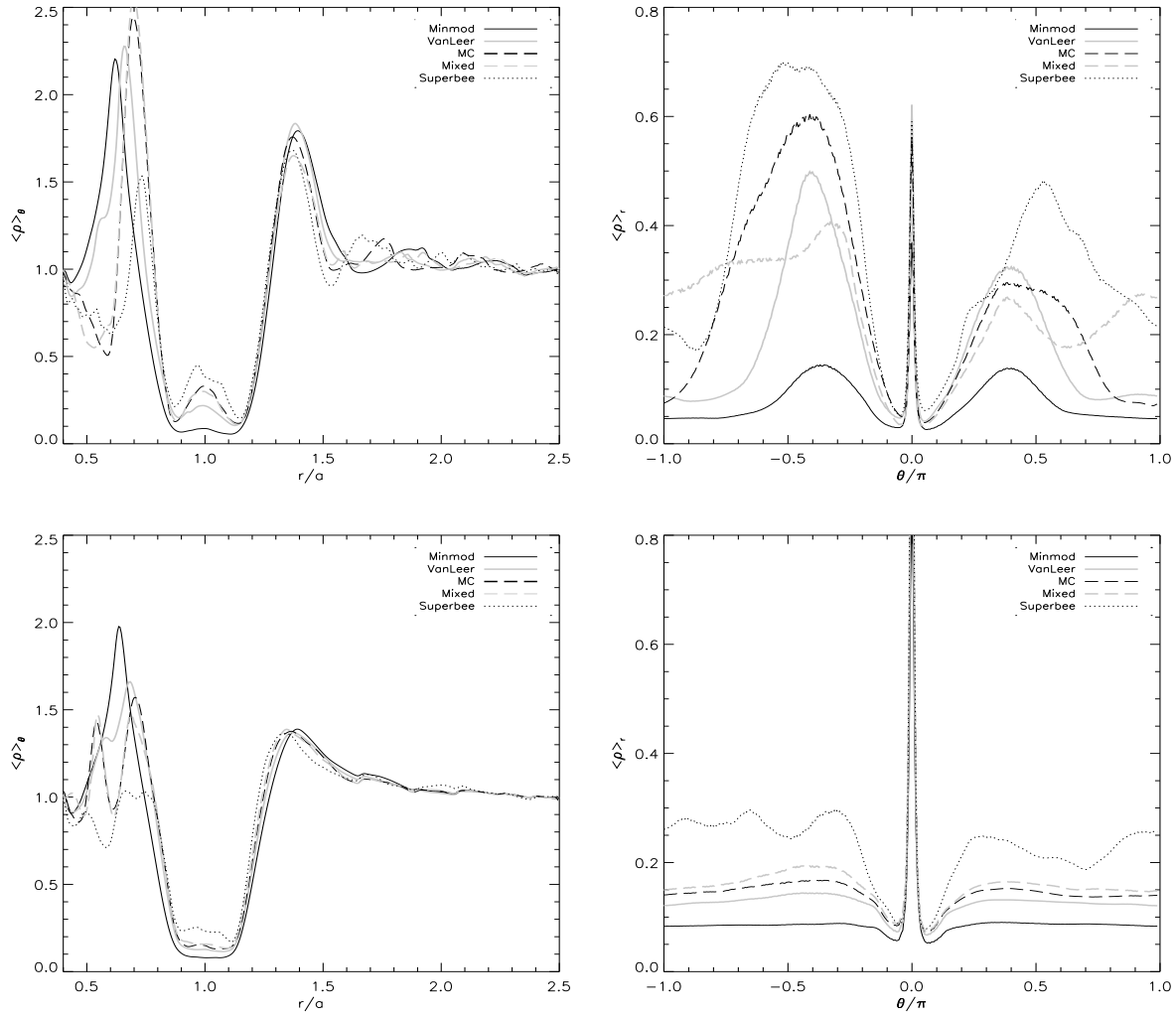
Figure 4.15 Comparisons of the azimuthally averaged density and the radially averaged density in the gap region for different limiter schemes. All results are taken at 100 orbits. The top two graphs have no added physical viscosity; the bottom two graphs have $\nu = 1 \times 10^{-5}$.
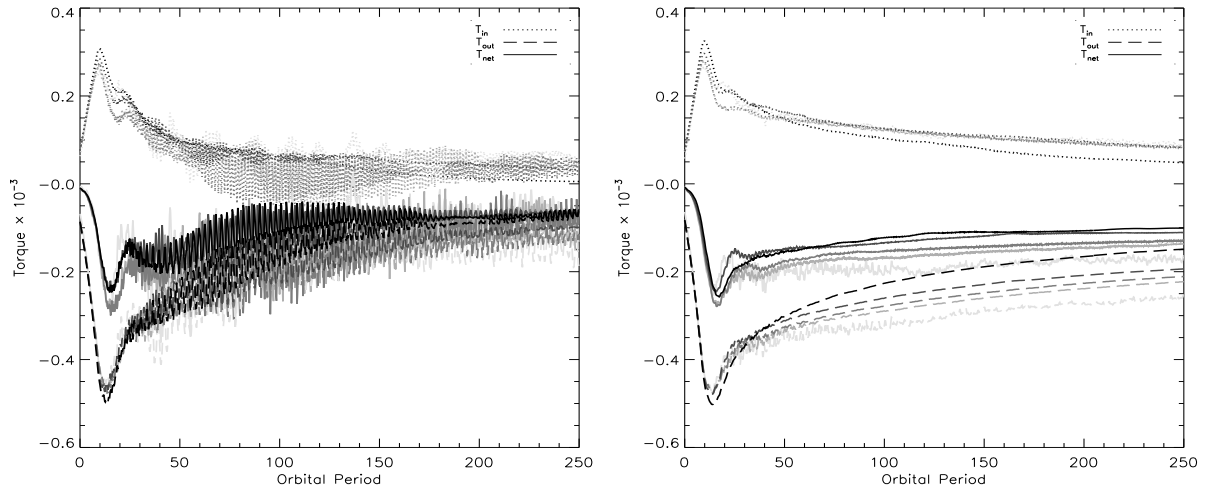
Figure 4.16 Time evolution of the torque on the planet's orbit due to the disk for different limiters. The graph on the left has no added physical viscosity; the graph on the right has $\nu = 1 \times 10^{-5}$. The torque is broken up into components from the disk material inside (dotted) and outside the planet's orbits (dashed). Only material further than one Hill radius from the planet is included in the calculation. The lines from darkest to palest correspond to the minmod, VanLeer, MC, mixed (MB) and superbee limiter schemes respectively.

I determine the level of numerical viscosity present within the code to be $\nu < 10^{-6}$ ($\alpha < 10^{-3.5}$), enabling the simulation of situation with Reynolds numbers on the order of $Re = UL \times 10^6$.

Results from the single protoplanet problem concerning the relative diffusion and dispersion due to the various limiter schemes are consistent with results presented in Section 3.11.

Chapter 5

The Role of Potential Vorticity, Vortices and Jets in Early Solar System Development

5.1 Introduction

In the previous chapter, I provided a careful examination of the developed code and the physically relevant structures produced in planet-disk simulations. In this chapter, I further examine these structures and the influence they have on the disk's development. I first examine the evolution of the potential vorticity distribution, and show its relation to corresponding jet and vortex/eddy structures in disks with a single protoplanet (§5.2) and with two protoplanets (§5.3). In the latter case, I demonstrate the failure of two widely separated planets to readily evacuate the inter-orbital gas between them. This conclusion modifies previously published results and holds even in the presence of viscosity and in runs with low resolution. In Section 5.4, I examine the connection between the eddy and jet structures and the averaged angular momentum transport. I then clarify how the above results for two-planet disk depend on the orbital separation of the planets (§5.5). Finally in Section 5.6, I relate a possible explanation for these results, discuss consequences of the observed structures for the evolution of such systems and comment on the applicability of such an analysis to three-dimensional disks.

5.2 Potential vorticity and structure evolution for a single protoplanet

In Figures 5.1 and 5.2, I show the evolution of PV, $Q \equiv \rho^{-1}(\nabla \times u + 2\Omega) \cdot \hat{z}$ (see §2.5), and the azimuthal velocity field over 300 orbits, for a Jupiter-mass planet using the Van Leer limiter and a resolution of $N_r \times N_\theta = 768 \times 1152$. Vortices develop, appearing in the PV map as closed contours which have negative PV relative to the background value. Initially several small vortices form at roughly the same radii, but spaced in azimuth. After 20 orbits, two vortex lines are visible, and by 50 orbits a third vortex line has appeared. As the simulation progresses, the vortices grow and begin to merge, until a single large vortex develops. These vortices appear as overdensities in density maps discussed in Chapter 4. They are likely the result of

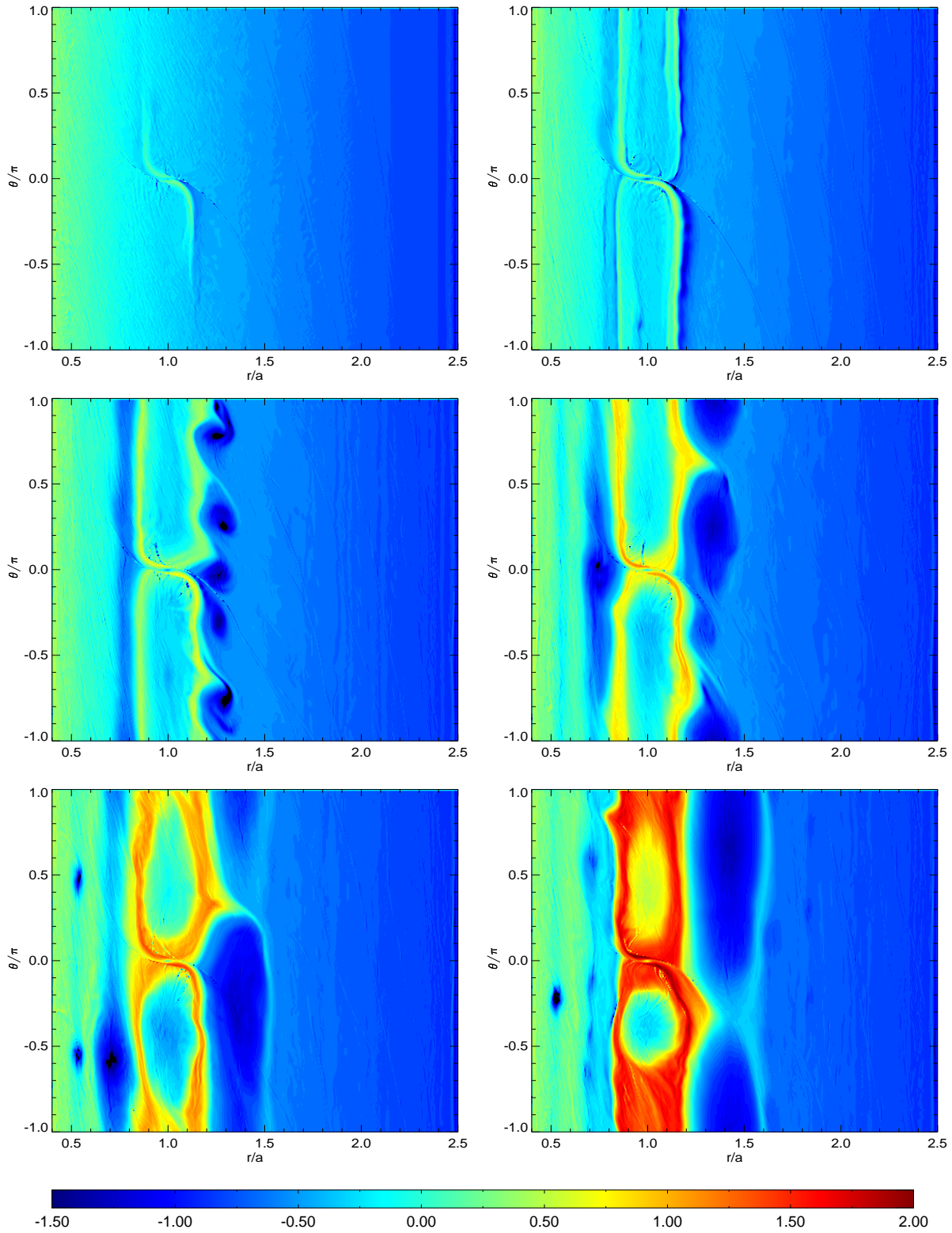Figure 5.1 Potential vorticity contours for a Jupiter-mass run using the VL limiter at resolution, $N_r \times N_\theta = 768 \times 1152$. Each plot in sequence shows a contour of log Q taken after 5, 10, 20, 50, 100, and 300 orbits, respectively.
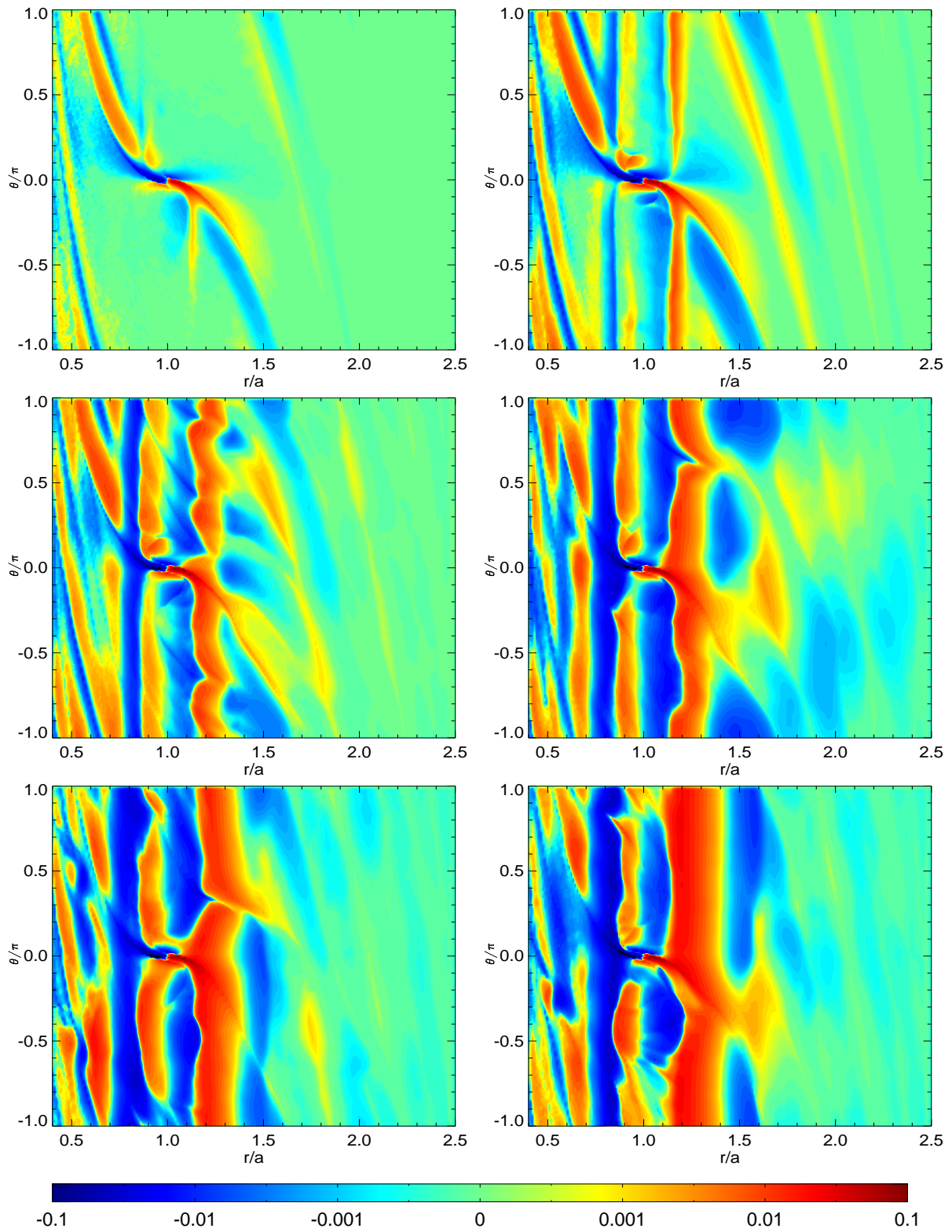
Figure 5.2 Contours of jet velocity (see text) corresponding to the PV shown in Figure 5.1. Note that strong jets develop to either side of the gap.

Rossby-wave instabilities as discussed by Papaloizou & Lin (1989); Lovelace et al. (1999); Li et al. (2000, 2001) and references therein.

The velocity shown in Figure 5.2 (hereafter referred to as the *jet velocity*), has the Keplerian velocity subtracted off so that $u_{jet} = u_\theta - v_K$. The jet velocity is the "headwind" or "tailwind" that a test particle moving at the Keplerian speed would feel inside the flow. Two large jets develop to either side of the gap (centered at approximately $r = 0.80$ and $r = 1.25$). These jets oppose the local Keplerian value of the shear velocity which would otherwise be positive inside the planet's orbit and negative outside its orbit. At the centers of the two main jets, the jet speed is more than 3% of the background Keplerian flow—60% of the sound speed. As such, these structures are substantial local deviations from the overall balanced flow. They are a result of the density gradient at the edge of the gap. This gradient and the flow it induces is strong enough that rotation causes an azimuthal jet to form, analagous to the formation of mid-latitude jet streams on the Earth from the latitudinal temperature gradient. The second set of jets which develop within the gap $(0.9 \lesssim r \lesssim 1.1)$) correspond to particles on horseshoe orbits of the coorbital region. These jets and the corresponding horseshoe-orbit regions shrink as the gas azimuthally opposite the planet is evacuated and eventually become the libration islands/eddies discussed in Chapter 4. Note that with increased resolution the time required for the horseshoe regions to shrink increases.

I display azimuthally averaged values of the PV and the jet velocity in Figure 5.3. Note that extrema in the PV profile correspond to a vanishing jet velocity and the centers of the jets straddle the gradients of PV between successive extrema, with negative slopes of PV corresponding to positive jet velocities and positive slopes of PV corresponding to a negative jet velocities. The observed jet velocity is induced by deviations in PV from the balanced flow (PV anomalies) via PV inversion. The initial value of the PV shown in Figure 5.3 is the background value due to the Keplerian velocity gradient. This background value acts as a mean-flow reservoir from which PV disturbances can extract energy. An analysis for two-dimensional, incompressible, inviscid fluids shows that while negative PV perturbations are able to extract momentum and energy from the mean flow and thus grow in magnitude, positive perturbations are not (Lithwick, 2007). This analysis may explain why the radii in Figure 5.3 which have negative PV relative to the background flow are the locations at which vortices are able to grow and merge—they correspond to the previously discussed vortex lines. Note that a gradient of PV supports the propagation of Rossby waves, and breaking Rossby waves are associated with many of the jets seen in the Earth's atmosphere (see §2.5.2).

As demonstrated in the previous chapter, the presence of too much viscosity destroys the growth and merging of vortices, possibly because it suppresses the Rossby-wave instability mechanism. de Val-Borro et al. (2007) arrive at the same conclusion in a forthcoming work. Viscosity does not, however, change the relative sign of the PV with respect to the background in these regions. In Figure 5.4 I show the PV distribution as in Figure 5.1 above, but for a simulation run with $\nu = 1 \times 10^{-5}$. Although the vortices are absent, Figure 5.5 indicates that jets still form. I compare the averaged values of the PV and jet velocity from the viscous and inviscid runs in Figure 5.6. Both the main jets and the pair within the gap are comparatively weaker, and the latter pair are eventually sheared away as the distribution of PV in the gap homogenizes.
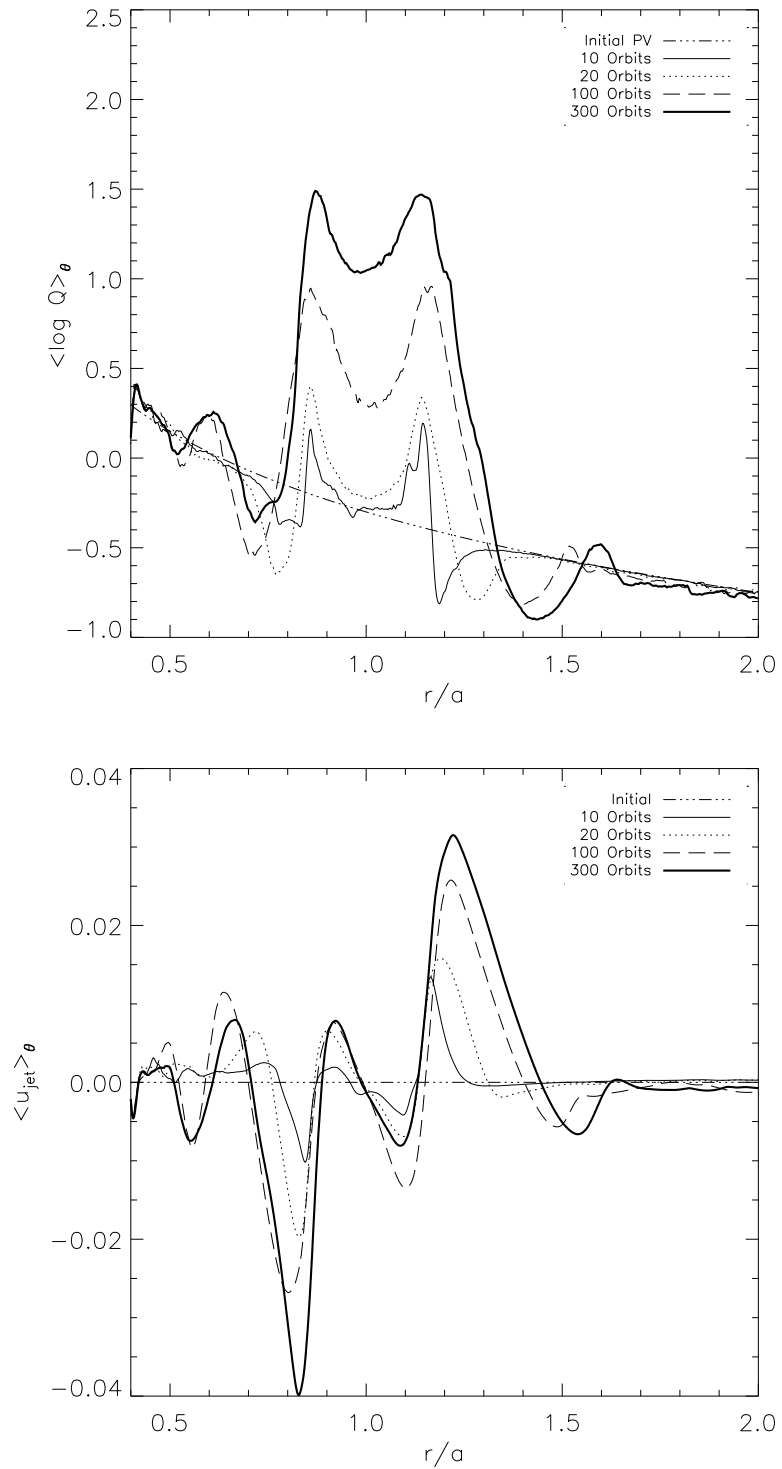
Figure 5.3 Azimuthally-averaged logarithmic PV and jet velocity corresponding to the results shown Figures 5.1 and 5.2.
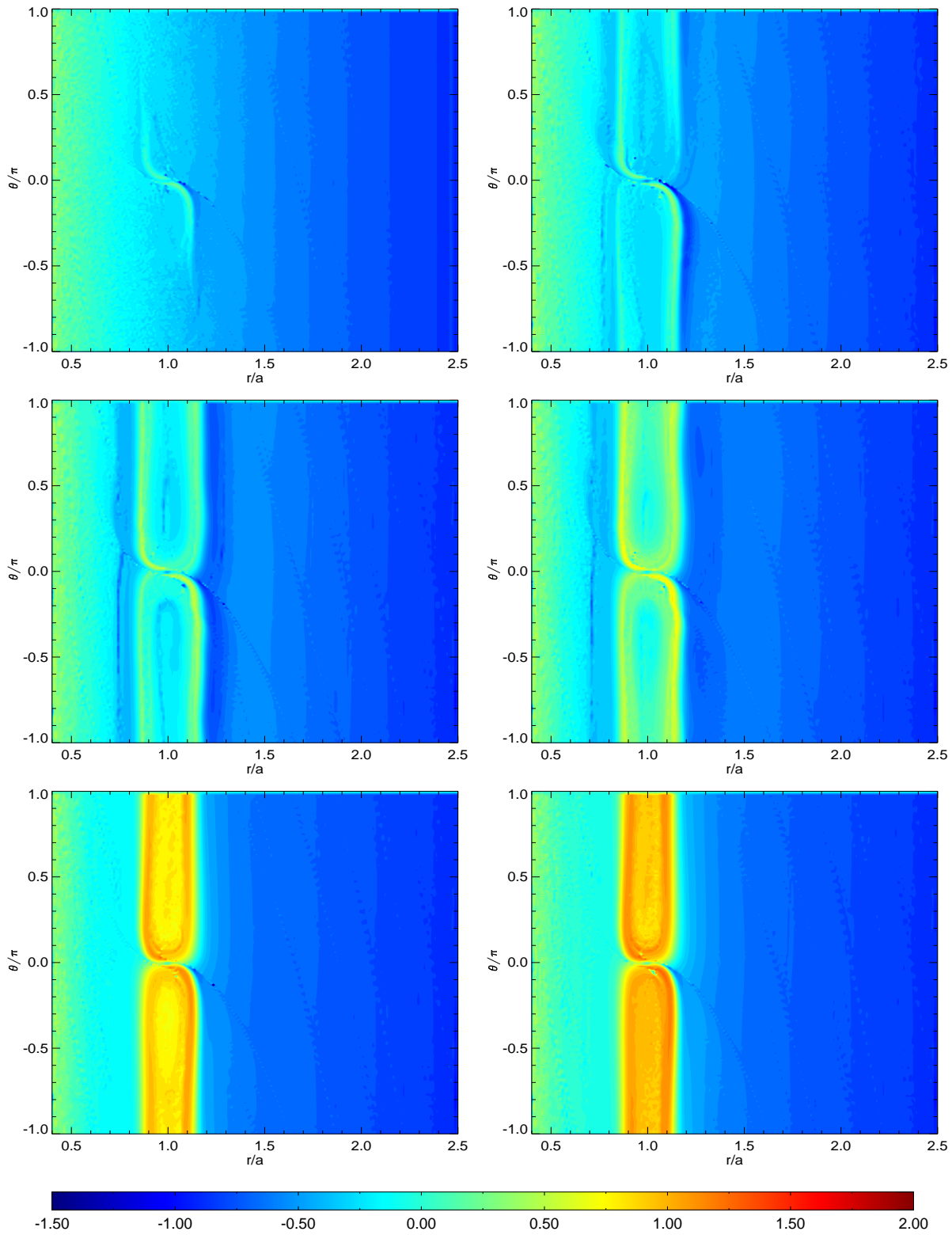
Figure 5.4 Potential vorticity contours for a Jupiter-mass run using the VL limiter at resolution, $N_r \times N_\theta = 768 \times 1152$ with $\nu = 10^{-5}$. Each plot in sequence shows a contour of $\log Q$ taken after 5, 10, 20, 50, 100, and 300 orbits, respectively.

Figure 5.5 Contours of jet velocity (see text) corresponding to the PV shown in Figure 5.4. Note that jets still develop to either side of the gap.

Figure 5.6 Azimuthally-averaged logarithmic PV profile and jet velocity corresponding to Figures 5.4 and 5.5, respectively shown in heavy lines; those corresponding to Figures 5.1 and 5.2 (without viscosity) are shown in pale lines.
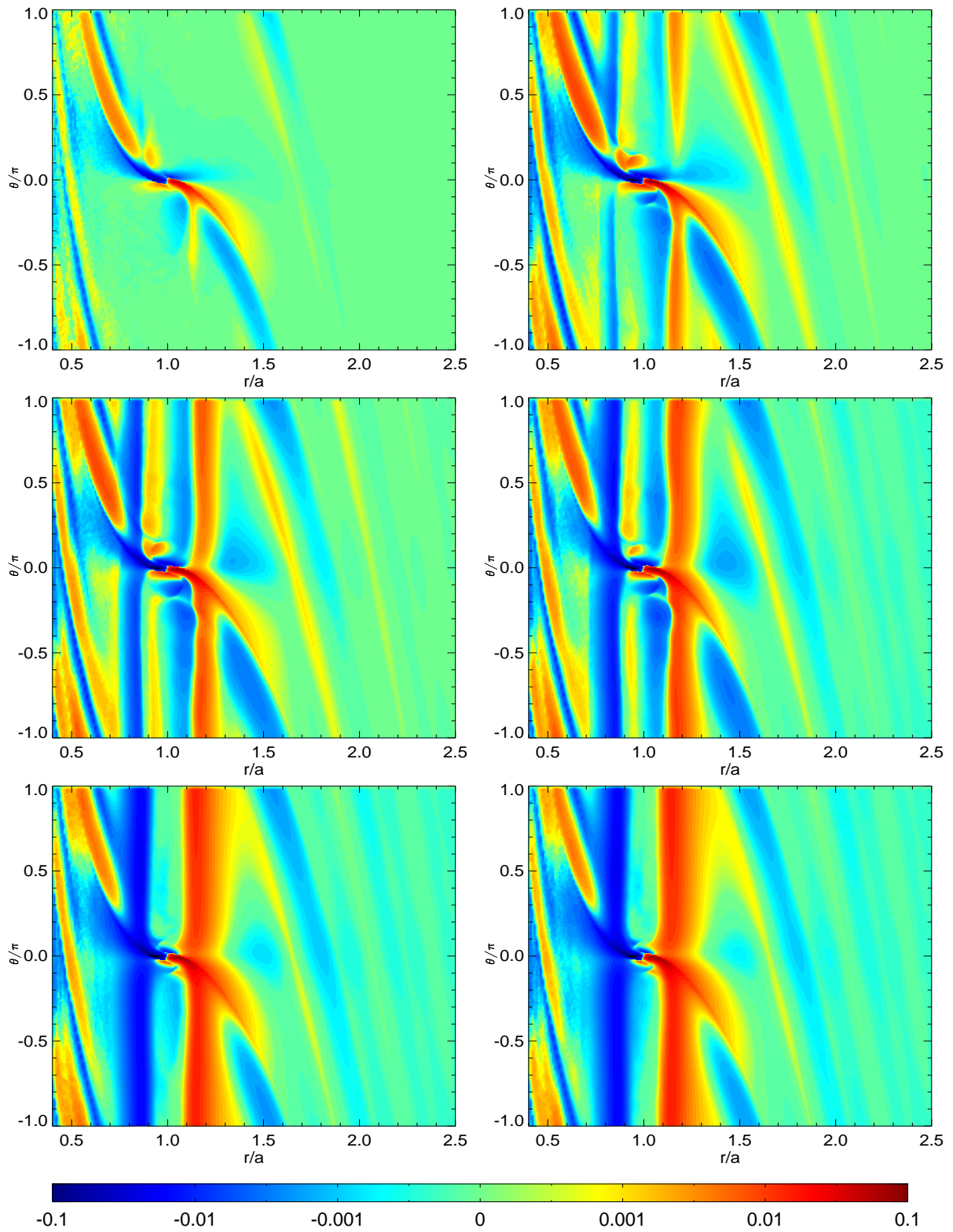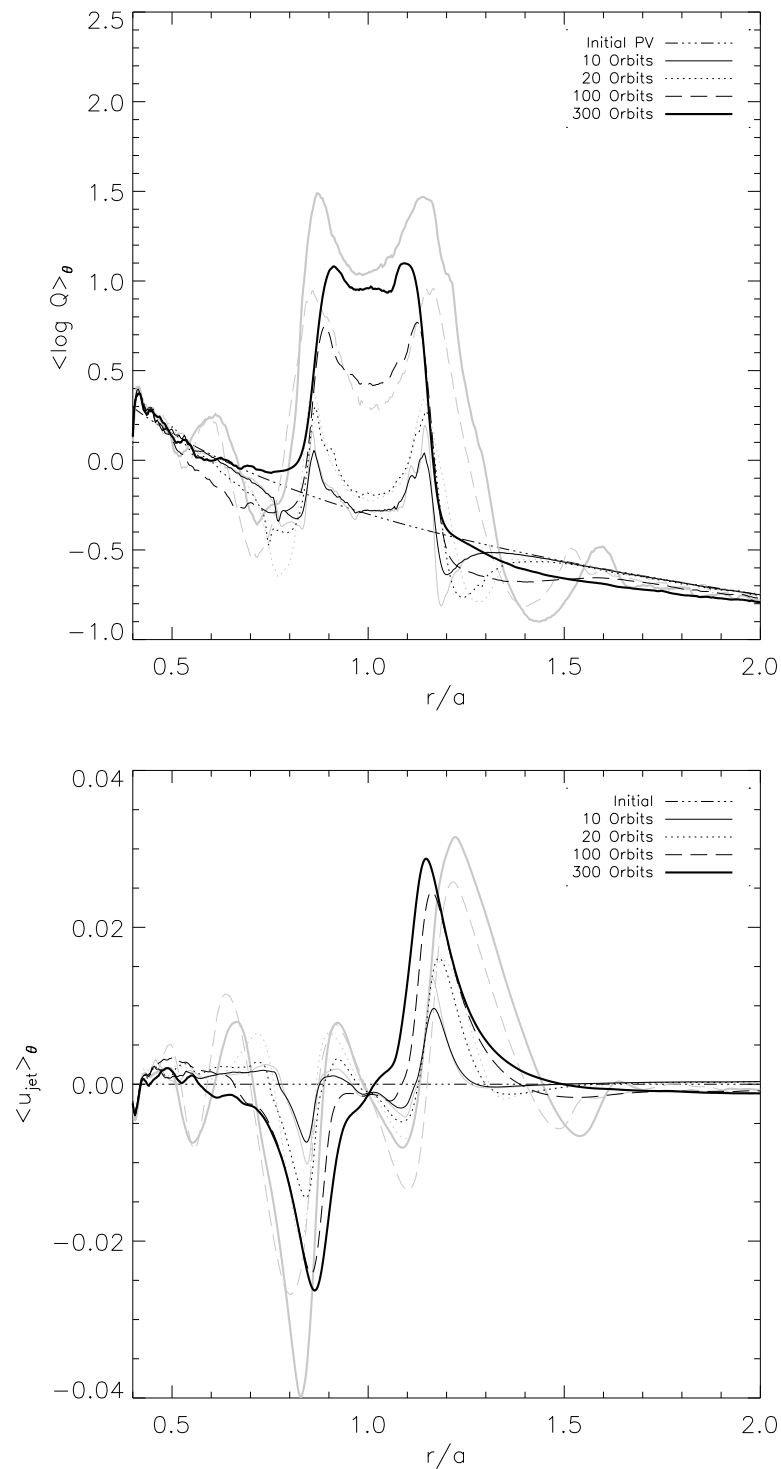
5.3   Potential vorticity evolution with two planets

With two protoplanets present, the gas surrounding each planet evolves in a manner similar to a single pro-
toplanet, as long as the planets are widely separated enough. I examine the effect of altering the separation
distance in Section 5.5. For now, I focus on the case where the second planet is at twice the orbital radius
of the first planet. For the remainder of this chapter, units are specified relative to the inner planet (which is
fixed with semi-major axis, $a_1 = 1$) unless otherwise stated.

The problem setup for two-planet simulations is nearly the same as that for a single planet. I still perform
the simulations in the corotating frame of the first planet; however I place the system's origin at the central
star, rather than the system's center of mass. To account for the new origin I redefine the planets' potentials
as

$$\phi_i = \frac{-Gm_i}{\sqrt{|r - r_i|^2 + \epsilon^2}} + \frac{Gm_i}{r_i^3} r \cdot r_i, \tag{5.1}$$

where r_i is the position of the planet. The second term accounts for the indirect acceleration of the frame due
to the planet's gravitational tug on the star. I extend the radial range to $[0.4a_1, 4a_1]$, increase the standard
resolution to $N_r \times N_\theta = 768 \times 1152$ and set both planet's masses to be $10^{-3} M_*$. The potential of the inner
planet increases from zero to its full value for the first 10 orbits as before; the potential of the outer planet
likewise increases over 10 orbits, beginning once the inner planet has completed 20 orbits.

In Figures 5.7, 5.8 and 5.9 I show the density, PV and jet velocity distributions for a simulation with two
planets after 50, 100, 200, 400, 600, and 800 orbits of the inner planet. As before vortex lines form to either
side of both planets. Because the two planets are close enough, the vortex lines between them merge together
after 100 orbits (see §5.5). As the simulation progresses, the gas between the two planets accumulates into
a single vortex and is very slow to clear from between the two planets. Figure 5.10 shows the azimuthally
averaged quantities from the above simulations at a selection of times. The mass of gas within the vortex
at 800 orbits is approximately $4m_p$. Such a concentration of material would have dynamical effects on
the system, however, following its evolution in a self-consistent manner would require accounting for the
self-gravity of the gas in the region.

These results hold even for simulations run with added viscosity or at low resolution. I show in Figure
5.11 the azimuthally averaged density and PV distributions after 50, 100, 400, and 800 orbits for two such
simulations. The first is a high resolution run as before but with $\nu = 1 \times 10^{-5}$. As in the single-planet case,
viscosity prevents the formation of vortices—the large vortex which previously formed between the planets
is absent. Nonetheless, the sharp gradients of PV and corresponding jets between the planets remain as in
the viscous single-planet simulation. While, the inter-orbital mass clears somewhat faster than the inviscid
run, roughly half the mass initially present in the region still remains after 1000 orbits (see Fig. 5.12 below).
Averaged quantities from a low resolution run ($N_r \times N_\theta = 128 \times 128$) are also shown in Figure 5.11. Vortices
form to either side of the outer planet just as in the high-resolution run despite the coarse resolution. Again,
sharp PV gradients and a pair of jets are present between the planets.

In Figure 5.12 I compare the change in mass of different regions of the disk for the standard two-planet
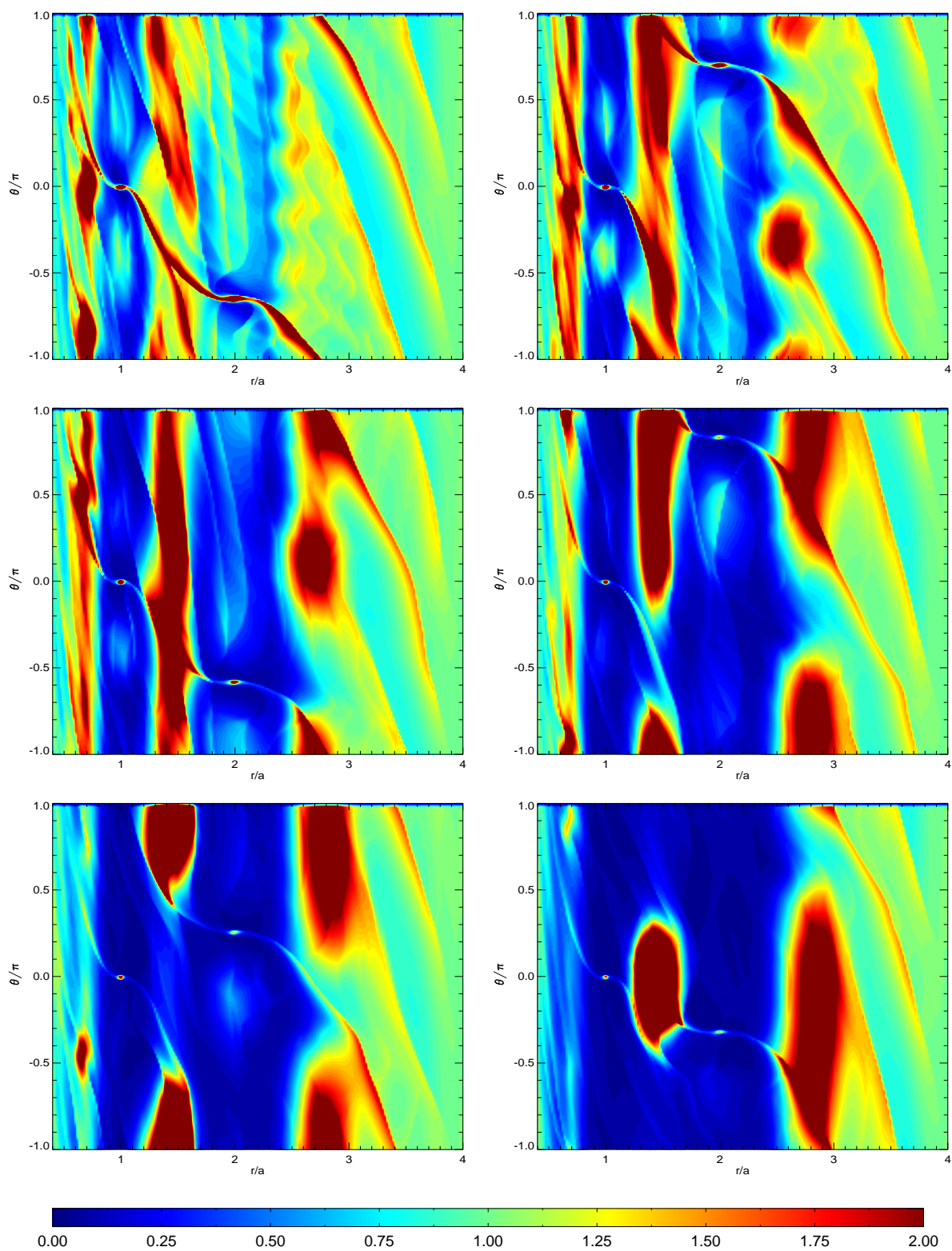
Figure 5.7 Density contours from a simulation with two protoplanets with orbital radii of $(a_1, a_2) = (1, 2)$ taken after 50, 100, 200, 400, 600, and 800 orbits of the inner planet.
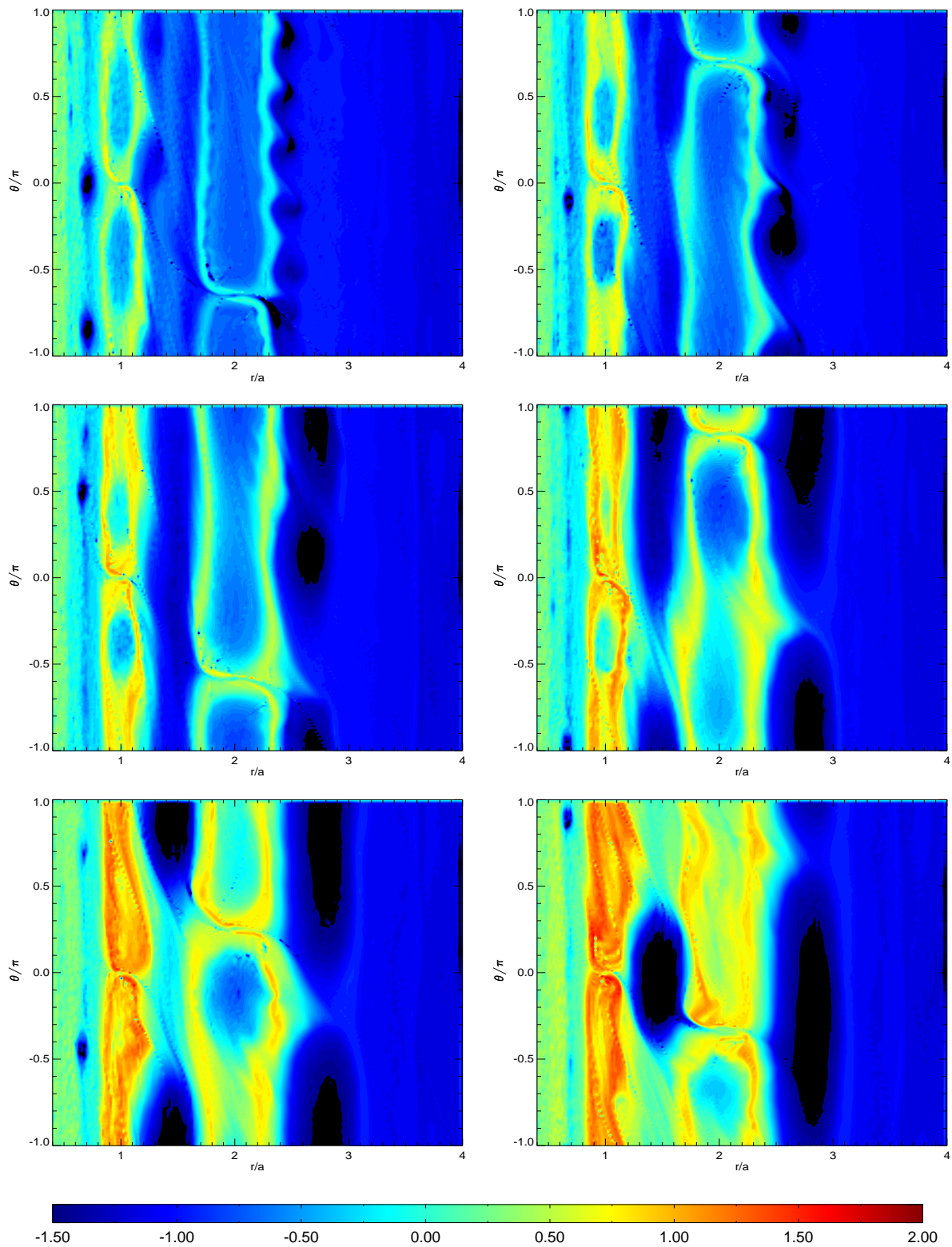
Figure 5.8 Contours of logarithmic PV from a simulation with two protoplanets with orbital radii of $(a_1, a_2) = (1, 2)$ taken after 50, 100, 200, 400, 600, and 800 orbits of the inner planet.
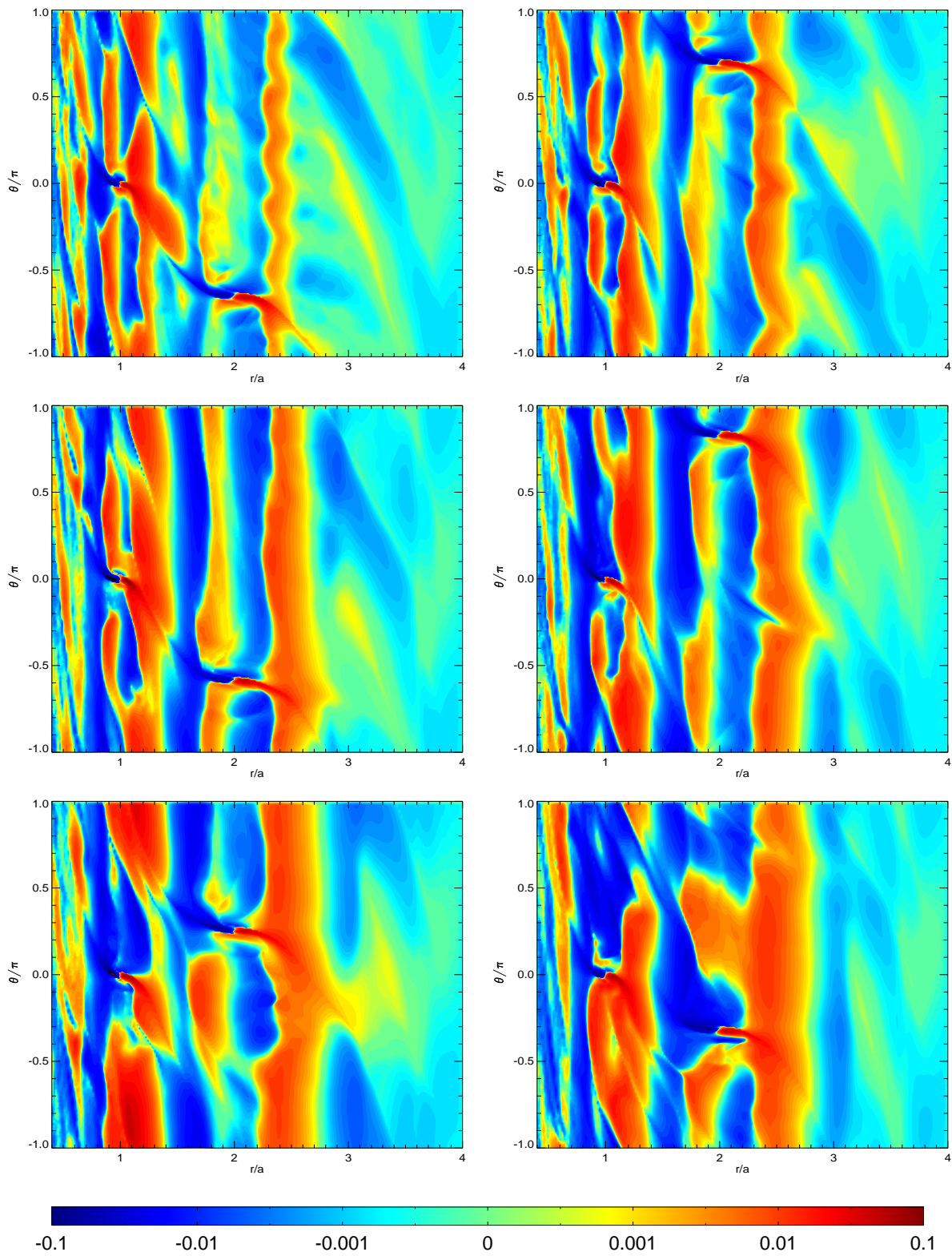
Figure 5.9 Jet velocity from a simulation with two protoplanets with orbital radii of $(a_1, a_2) = (1, 2)$ taken after 50, 100, 200, 400, 600, and 800 orbits of the inner planet.
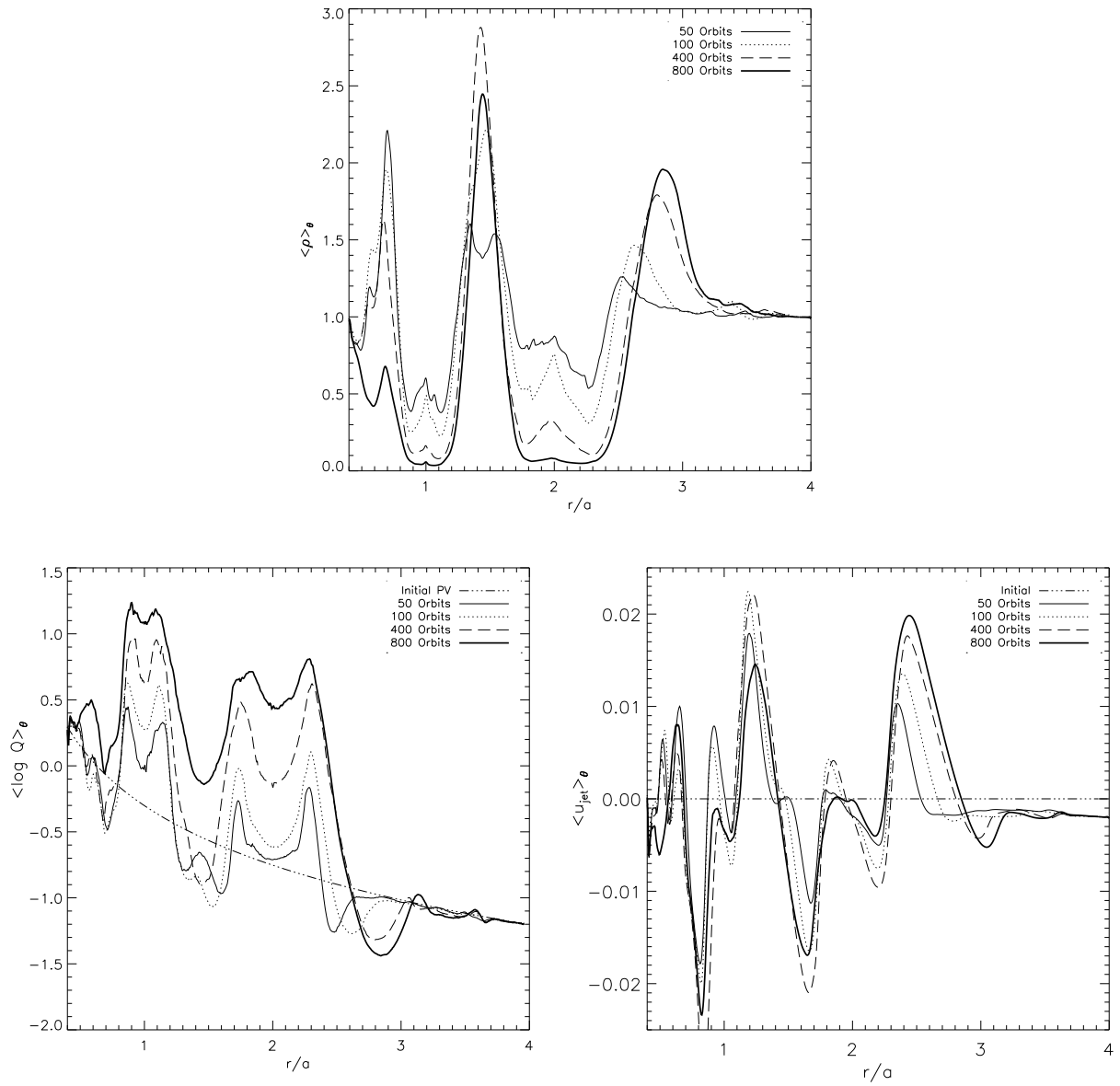
Figure 5.10 Azimuthally averaged density, logarithmic PV profile and jet velocity corresponding to Figures 5.7, 5.8 and 5.9.
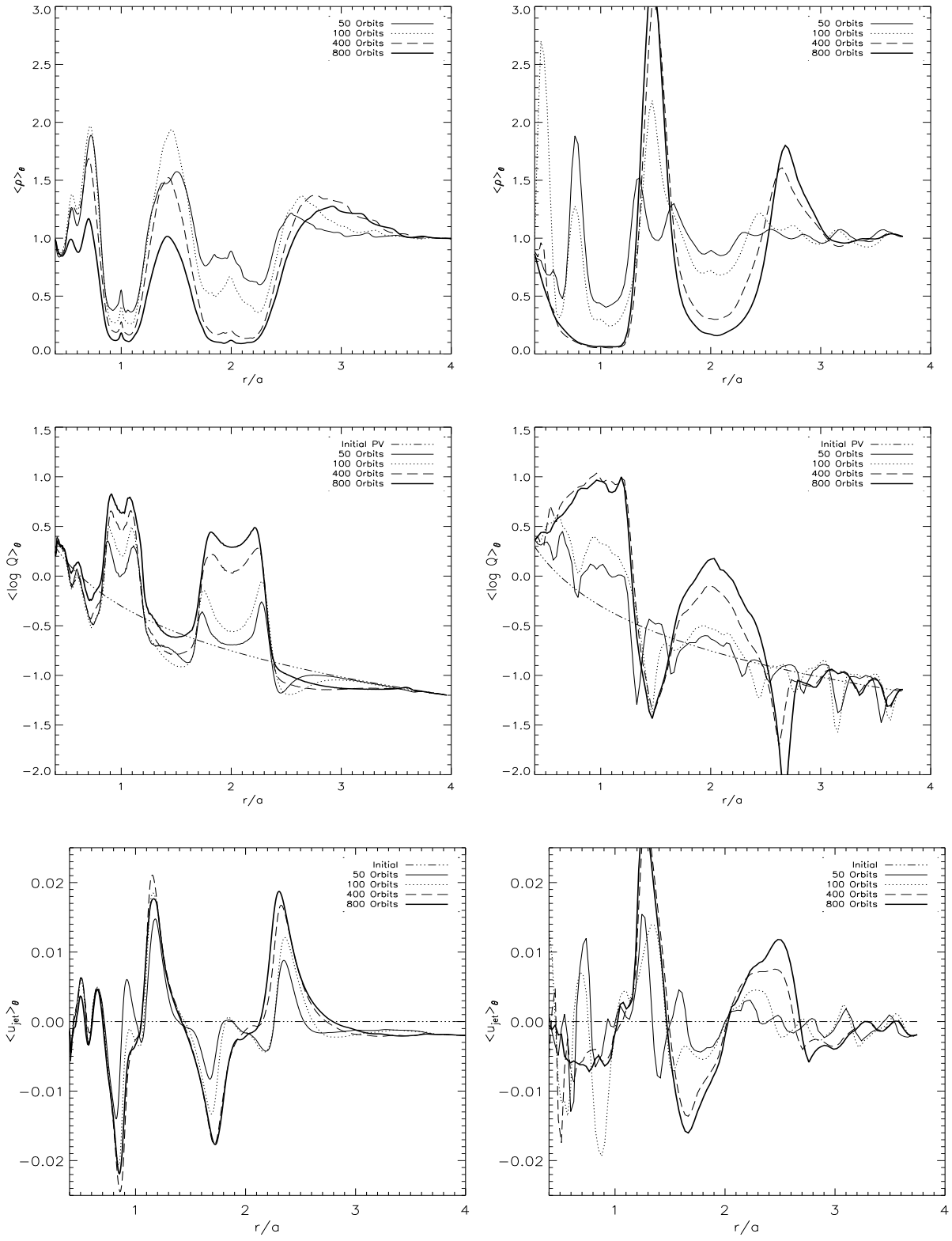
Figure 5.11 Azimuthally averaged density logarithmic PV and jet velocity (top. middle and bottom rows, respectively). Plots on the left are from the standard two-planet resolution run with $\nu = 10^{-5}$; those on the right are from a low-resolution run ($N_r \times N_\theta = 128 \times 128$).
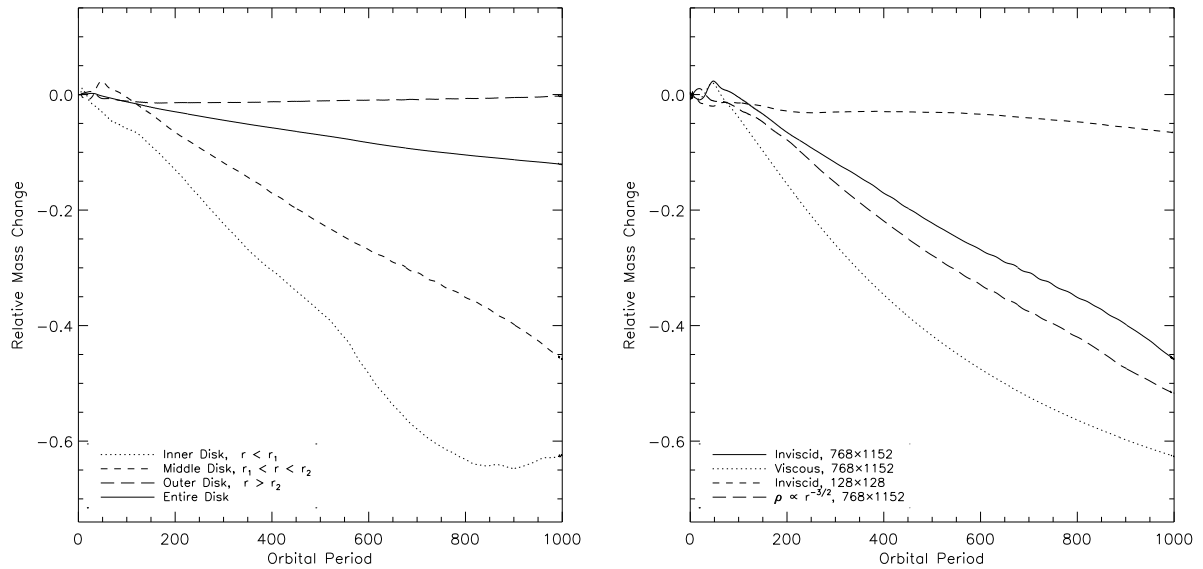
Figure 5.12 Change in mass of various regions of the disk, normalized in each region with respect to the initial mass within that region. The plot on the left shows evolution of different disk regions for the inviscid, $N_r \times N_\theta = 768 \times 1152$ run. The plot on the right shows the evolution of the middle disk only for four different combinations of viscosity, resolution and density profile.

simulation, with the viscous and low resolution runs. After 1000 orbits in the standard run, more than half of the original gas mass still remains in the region between the planets. This slow clearing of the gas from the inter-orbit region contradicts previous low-resolution results by Kley (2000). In that work, the region between two planets with the same relative spacing was seen to clear within 300 orbits to better than 95%. Although the planets were allowed to migrate with respect to one another, the mass evacuation primarily occurred while the planets were still far apart. The simulation by Kley also used a $r^{-3/2}$ profile for the density distribution. The evolution of the mass from a run with this density profile is also plotted in Figure 5.12 and only affects the results in the inter-orbital region by 5%.

5.4   The connection with averaged transport

The connection observed in atmospheric systems between jet structure and reduced transport offers a possible explanation for the slow clearing observed in the above simulations. According to such studies, the evolution of strong jets in the flow is synonymous with weakened transport in the region. Strictly speaking, the transport is lowered for materially conserved quantities, rather than globally conserved quantities. Thus while the evolution of the total mass within the region is not constrained, the evolution of the total mass between a given number of tracked fluid elements should be. Because it is not possible to track material elements in an Eulerian code, observing the density within the region provides an estimate of how the material mass elements are evolving only in an spatially and temporally averaged sense.

In order to estimate the level of momentum transport within the disk, I calculate an effective viscosity coefficient based on the radial movement of angular momentum. Following work by Balbus & Hawley (1998); Li et al. (2001); Klahr & Bodenheimer (2003), I define the total averaged angular momentum flux as

$$\mathcal{F}_m = r^2(< \rho u_r u'_\theta > + < \rho u_r \bar{u}_\theta >), \tag{5.2}$$

where the angled brackets represent a spatial and azimuthal average. In writing equation (5.2), I assume that the azimuthal velocity can be decomposed into mean and varying components as $u_\theta = \bar{u}_\theta + u'_\theta$. The first term in the equation represents transport due to correlations in the velocity components while the second indicates the direct radial flow of matter. It is possible to define an instantaneous, local version of an effective alpha-parameter based on equation (5.2) as

$$\alpha(r, \theta, t) = \frac{\rho u_r(u_\theta - \bar{u}_\theta)}{p}, \tag{5.3}$$

where $p = \rho c_s^2$ is the pressure. All variables are functions of r, θ and t except for the azimuthally averaged velocity, $\bar{u}_\theta(r, t)$, and the sounds speed, $c_s(r)$. I also define an averaged effective alpha-parameter, averaged in azimuth and time,

$$< \alpha >= \frac{< \rho u_r u_\theta > - < \rho u_r \bar{u}_\theta >}{< \rho > c_s^2}. \tag{5.4}$$

Note that the above equations can be written in terms of the jet velocities since $u_\theta - \bar{u}_\theta = u_{jet} - \bar{u}_{jet}$ by the definition of the jet velocity.

In the upper half of Figure 5.13 I show the instantaneous, local alpha-parameter values for a single planet simulation after 200 orbits. The simulation was performed with resolution, $N_r \times N_\theta = 768 \times 1152$, and over a radial range of $(0.4a_1, 4a_1)$. The calculated value of the alpha-parameter varies considerably both spatially and temporally. Averaging in azimuth and time produces a more coherent picture. In the lower half of Figure 5.13 I show the temporally averaged local alpha-parameter. Instantaneous values were sampled 11 times an orbit and averaged over 10 orbits. In the averaged results, eddies at the L_4 and L_5 points display their characteristic quadrupole transport pattern (Li et al., 2001). Also visible beyond $r = 2$ is transport
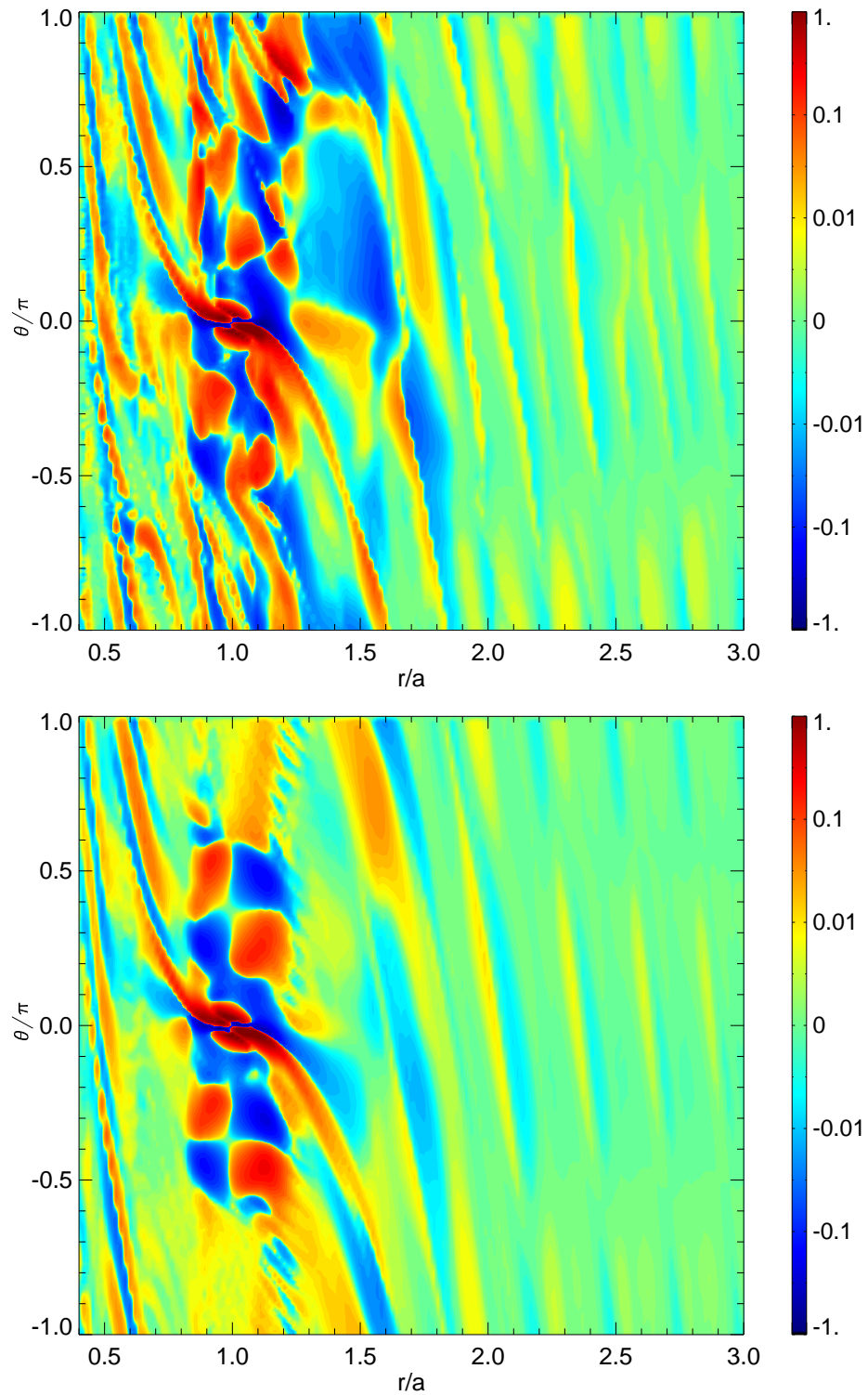
Figure 5.13 Local transport calculated as an α-parameter for a single planet simulation. The top plot shows the instantaneous transport; the bottom plot is averaged over 10 orbits.
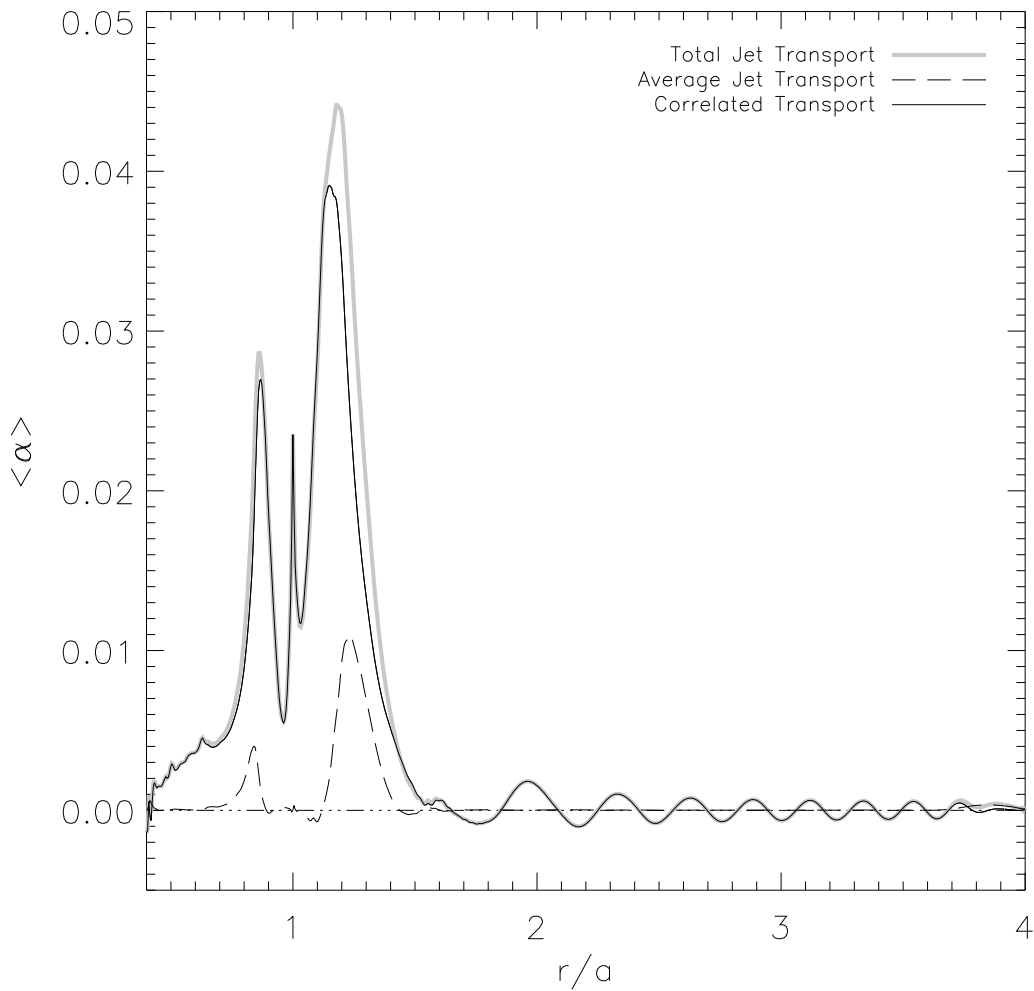
Figure 5.14 Individual transport components: the thick grey line shows the total averaged transport, the solid black line show the component due to velocity correlations, and the dashed line show the component due to the averaged jet velocity.

associated with the spiral arms.

I present calculations of the fully averaged alpha-parameter (eq. [5.4]) in Figure 5.14. Results shown were averaged in azimuth and in time over 10 orbits as described above. Averaged over 5 orbits, they are almost identical and only when averaged over 2 or 1 orbits do larger fluctuations occur. The averaged jet transport as calculated is only significant near the centers of the large jets which bracket the planet's orbit. Most of the transport is due to correlations between velocities. Note that the spiral density wave transport is clearly shown with a strength given by $\alpha \sim 10^{-3}$.

In the top graph, bottom panel of Figure 5.15, I show the averaged alpha-parameter calculated from a simulation with two planets. Results from two separate simulations with just one of the two planets are also shown. The averaged value of the alpha-parameter correlates with the relative vorticity distribution, q, and
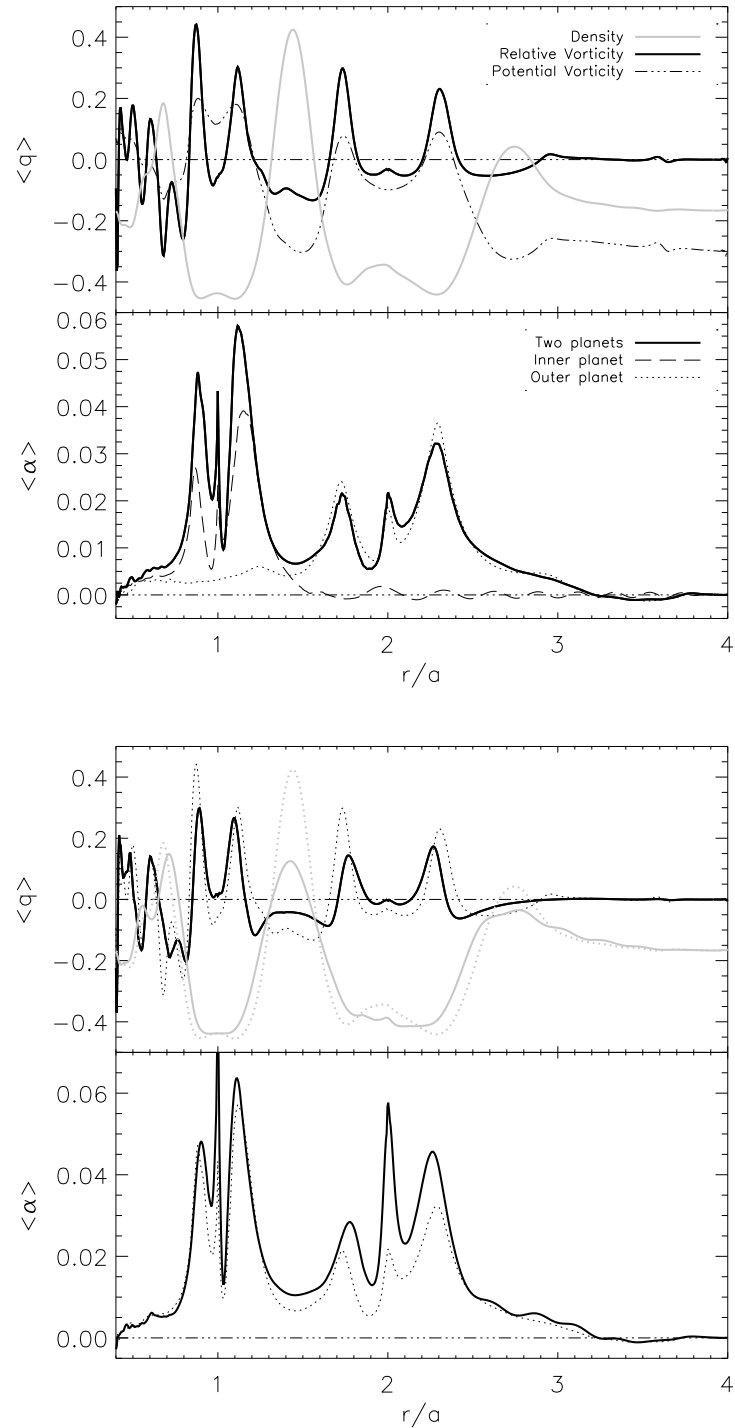
Figure 5.15 Top graph: effective alpha (bottom panel) from a two-planet simulation ($a_1 = 1, a_2 = 2$) after 200 orbits of the inner planet and from separate simulations with only one of the planets present. The top panel shows that relative vorticity, PV and density all correlate with the effective alpha-parameter. In the top panel the ordinate marks only the relative vorticity; the logarithmic PV and density have been appropriately scaled. Bottom graph: relative vorticity, density and effective alpha from a run with $\nu = 1 \times 10^{-5}$ (solid lines) along with the same results from the inviscid run (dotted lines) shown in the top graph.

by extension with the PV distribution as well, as can be seen by comparison with the top panel. Also shown is the averaged density which correlates negatively with alpha. When interpreted as a measure of the angular momentum transport occurring in the disk, the above result supports the conjecture made in the previous section, that the jet structures between the planet are associated with lowered transport. Such an association could explain the slow clearing of gas between the planets. In the bottom graph of Figure 5.15 I compare the averaged value of the alpha-parameter from an inviscid run and a run with $\nu = 1 \times 10^{-5}$ ($\alpha = 0.004$). As expected, viscosity increases the measured level of transport and other correlated values adjust accordingly. Over much of the range between the planets, the amount of added viscosity has resulted in an equivalent increase in the calculated value of $< \alpha >$.

5.5   The dependence of mass clearing on the planets' relative separation

In order to further investigate the relationship between disk structures, and the amount of mass and radial momentum transport, I performed a series of simulations with the outer planet at different relative orbital separations from the inner planet. Figures 5.16 and 5.17 show density and PV contours for six simulations all with $a_1 = 1.$, but with $a_2 = 1.6, 1.7, 1.8, 1.9, 2.0, 2.5$. Results shown are after 200 orbits of the outer planet, in order to consistently compare the inter-orbit region. As the separation between the planet decreases, the mass between the planets clears in a shorter and shorter amount of time. Correspondingly, as the separation increases, larger and more massive vortices are apparent in the inter-orbital region. For the simulation with $a_2 = 2.5$, the two vortex lines which form outside the inner planet's orbit and inside the outer planet's orbit are spaced widely enough that they do not merge over the course of the simulation (1000 orbits of the inner planet).

In Figure 5.18 I show the relative change in mass of the inter-orbit region for each of the separations. A clear trend exists: as the separation increases the rate at which mass is removed from the region decreases. Calculations of the averaged alpha-value are shown in Figure 5.19 along with the relative vorticity and density for the six simulations after 200 orbits of the inner planet. If the planets' orbits are close enough, the vorticity profiles begin to overlap, increasing the transport in the inter-orbital region and evacuating the gas more quickly.

5.6   Discussion and future work

I have demonstrated that jets evolve spontaneously in disks with one or two protoplanets, forming to either side of a planet's orbit as the planets affect the flow in the disk. The jets are related to the sharp radial gradients in the potential vorticity through PV inversion and to the density (pressure) gradients that occur in the disk (most prominently at the gap edges). The density gradients and the flows they produce are strong enough that rotation causes an azimuthal jet to form. This process is analagous to the mid-latitude zonal jet streams in the Earth's atmosphere which result from the latitudinal pressure gradient.
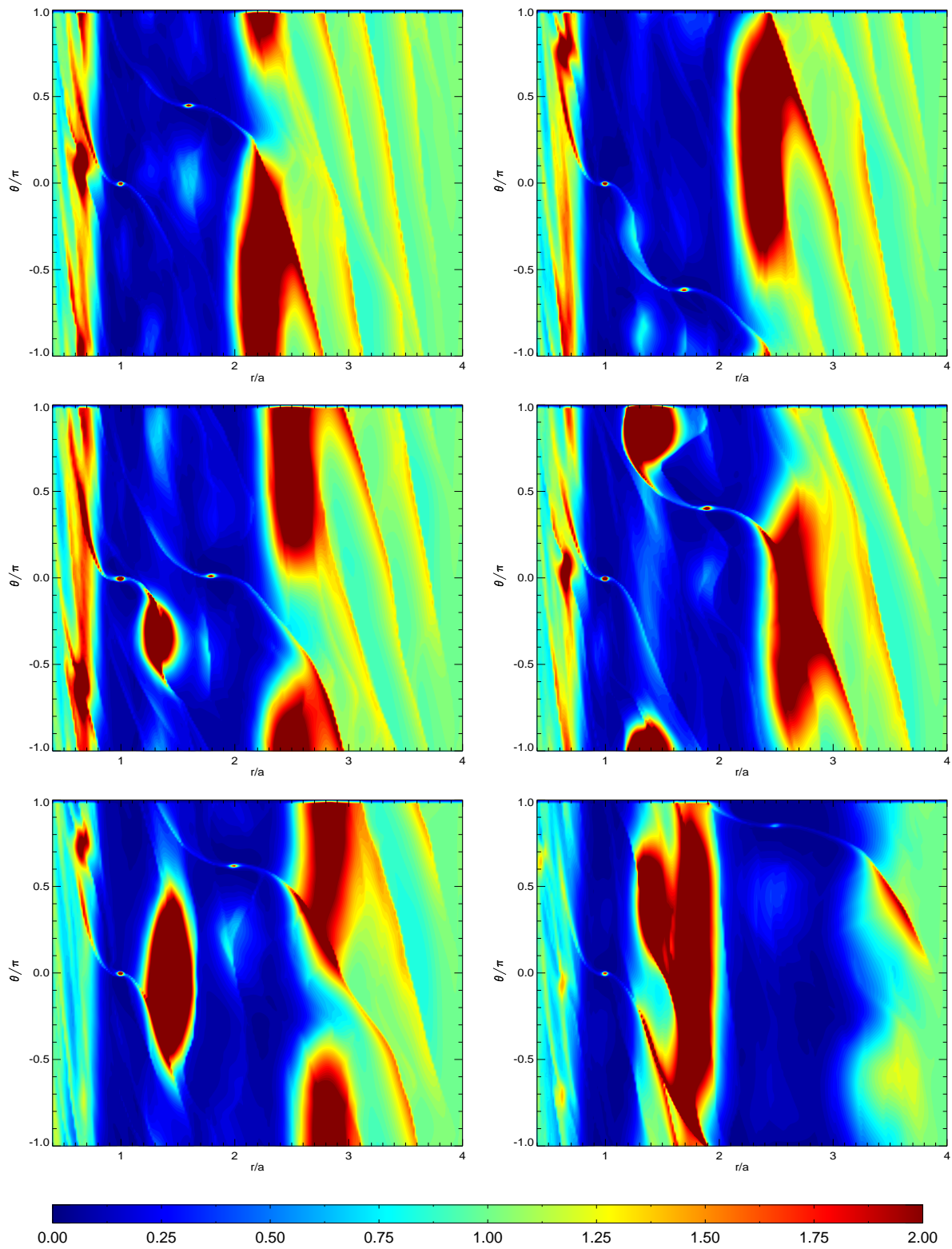
Figure 5.16 Density contours after 200 orbits of the outer planet for simulations with $a_1 = 1$ and $a_2 = 1.6$, 1.7, 1.8, 1.9, 2.0, and 2.5, left to right and top to bottom.
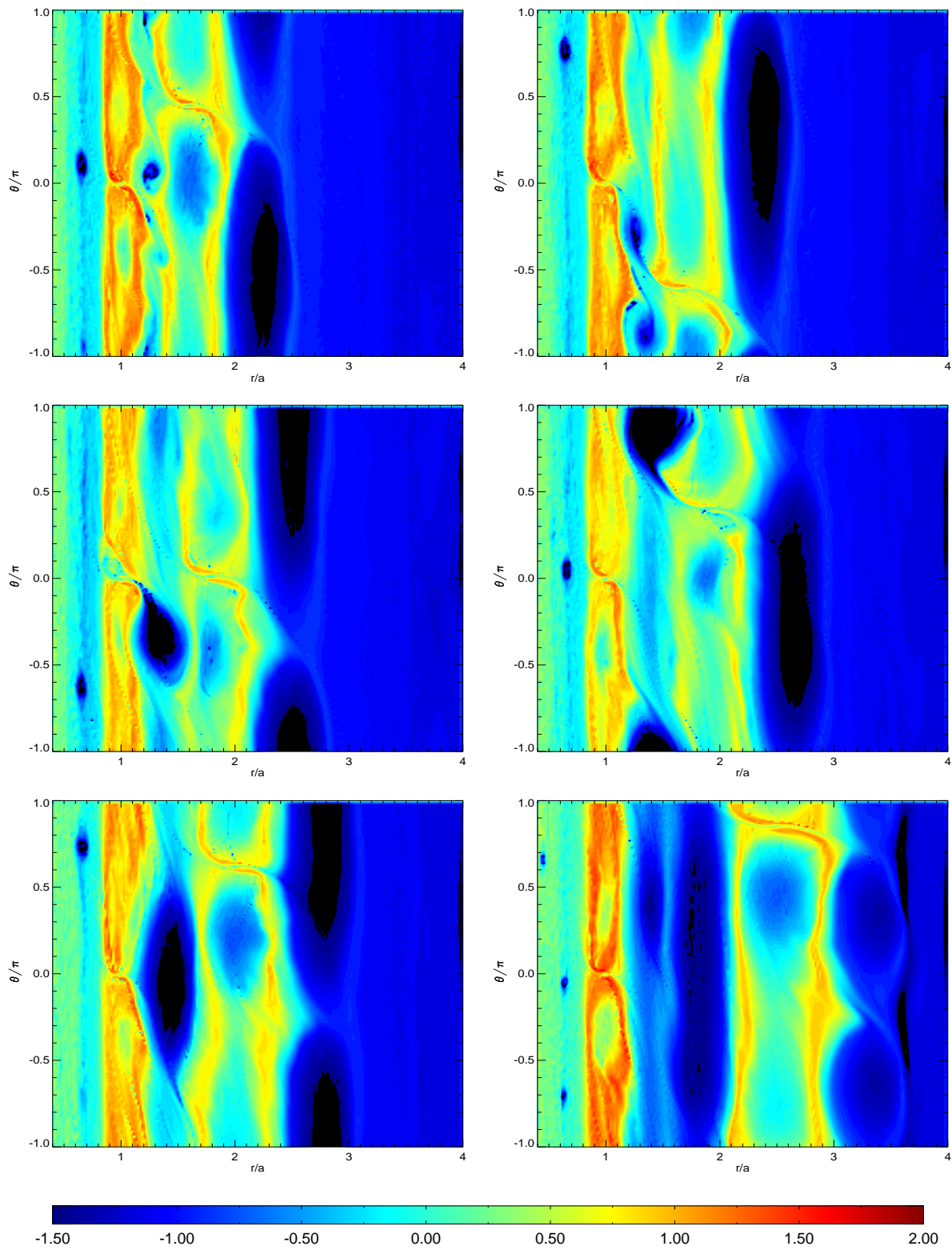
Figure 5.17 PV density contours after 200 orbits of the outer planet for simulation with $a_2 = 1.6$, 1.7, 1.8, 1.9, 2.0, and 2.5, left to right and top to bottom.
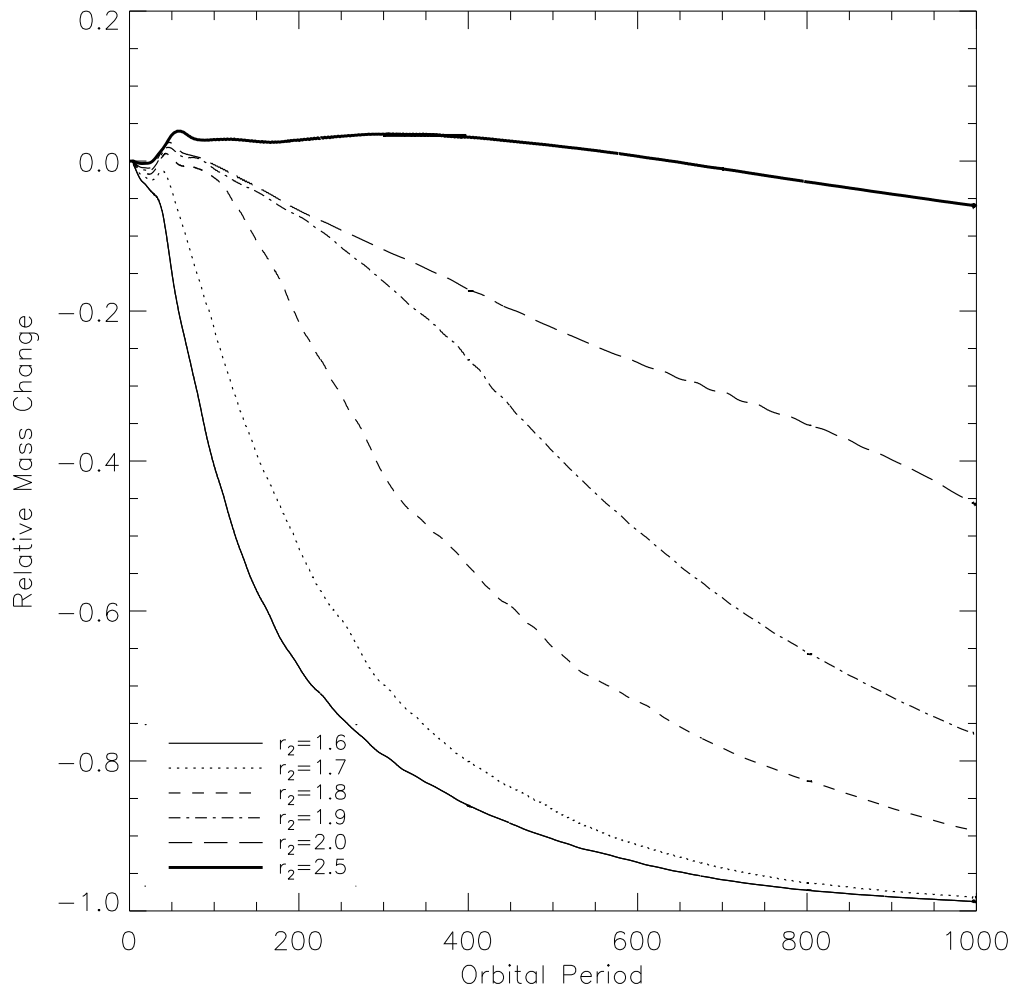
Figure 5.18 Change in mass of the region between the planets with radius separation between the planets. Each curve is normalized relative to the initial mass between the planets.
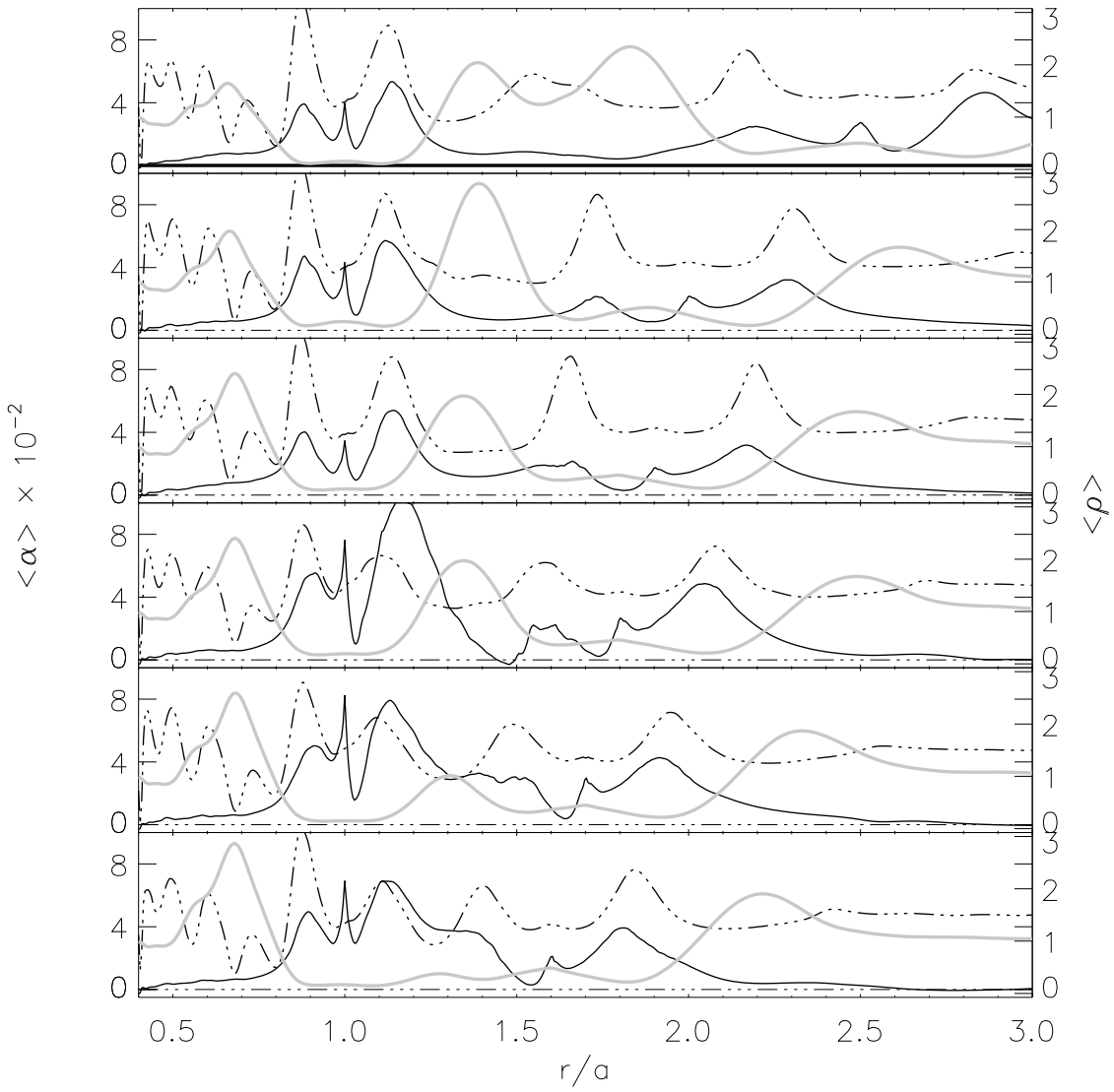
Figure 5.19 Each panel shows the estimated level of transport calculated as an effective alpha-parameter (dark line), relative vorticity (dot-dashed line, scale not shown) and density (pale line) after 200 orbits of the inner planet.

Studies in atmospheric physics show that jets are able to form spontaneously from forced turbulence via baroclinic instabilities (Rhines, 1975; Panetta, 1993; Cho & Polvani, 1996), and that this process may explain jet patterns seen in the Earth's atmosphere and in the atmospheres of the Jovian planets (Williams, 1978). (Klahr & Bodenheimer, 2003) have demonstrated that baroclinic turbulence in planetary disks produces outward angular momentum transport with a strength characterized by $\alpha = 10^{-4}$–10^{-2}. The turbulence in their simulations produced both pressure and Rossby waves and generated large vortices with significant overdensities. I have presented results where similar vortices develop, possibly through a similar baroclinic instability process. In the two-planet simulations a large vortex was formed between the planets, with a mass of approximately $4m_p$. Such a concentration of gas will likely affect the subsequent evolution of the system, and a self-consistent examination of its evolution would require accounting for the gas' self-gravity. I have also shown that the vortex formation process is damped once $\nu \gtrsim 10^{-5}$, however the jet formation process is more robust in this sense, and jet structures form even in viscous flows.

These jets are important. Atmospheric studies show weakened transport along the central axes of jets, or equivalently along gradients of PV (McIntyre, 1989; Sommeria et al., 1989; Haynes & Shuckburgh, 2000; Marshall et al., 2006). Prompted by these studies, I have estimated the averaged radial angular momentum transport present in the my simulations as an effective alpha-parameter. The measured transport calculated as such is outwards as expected and varies considerably over the extent of the disk, from as large as $<\alpha> = 0.01$–0.05 in the vicinity of the planets to $<\alpha> \sim 10^{-3}$ away from the planets, in regions where the spiral arms dominate. Near the planets I have shown that the calculated alpha-parameter correlates with the PV itself, rather than with its gradient. Alpha-parameters greater than 10^{-2} are quite large and could reflect the strictly two-dimensional nature of the simulations rather than layerwise two-dimensional turbulence embedded in a rotating three-dimensional disk. I also caution that while interpreting the average alpha-parameter as a measure of the radial angular momentum transport appears reasonably consistent, their actual relationship should be further investigated.

I have demonstrated a clear connection between the rate of mass evacuation and the inter-orbital separation of two protoplanets embedded in as gas disk. These results are also consistent with measurements of the averaged transport, calculated as an alpha-coefficient. As the orbital separation of the planets increases, the measured alpha-parameter decreases and the mass evacuates more slowly. In particular, I find different results than those published by Kley (2000) when the outer planet is at twice the orbital radius of the inner planet. Kley showed the mass in the inter-orbital region was evacuated to 95% over 300 orbits, while my simulations only showed evacuation to 40% over 1000 orbits. These conclusions stand even when performed at low resolution and with added viscosity at the level, $\nu = 1 \times 10^{-5}$. The relation between the jets and weakened transport provides a possible explanation for this discrepancy. Simulations with larger numerical viscosity would produce weaker jet structures. If the jet structures become too weak and too broad, the transport in the jet region increases.

It is still unclear how this gas is able to cross the planet's orbit. Masset & Ogilvie (2004) have shown how fluid streamlines pass across a single potential component in the presence of viscosity, but an actual planet

potential creates a mess of overlapping resonances in the coorbital region. The overlap of these resonances suggests the region should be chaotic. Whether the gas is able to "tunnel" across the region anywhere, or only at preferred azimuths (such as at the planet's azimuth, or opposite its azimuth) is unknown.

Furthermore, I have not yet addressed the source of the vorticity generated in the simulations that I presented. While baroclinic generation of PV is certainly possible, Ou et al. (2007) have noted that the distribution of baroclinically generated vorticity expected in such a disk does not match the actually observed PV generation. The authors suggest mixing of PV may be responsible for this discrepancy. This explanation is especially interesting in light of atmospheric studies which suggest inhomogeneous mixing may be responsible for creating PV gradients in the first place. Yet what is the source of this mixing? The generation of vortex lines at $r \approx 0.55$ and $r \approx 0.7$ are locations of the $m = 3$ and $m = 5$ Lindblad resonances. While refraining from a descent into numerology, it is nonetheless tempting to suggest that spiral density waves launched in such regions could provide an alternative source of wave activity leading to the observed jets and possibly provide a source of mixing in the coorbital region. Simulations with an adiabatic code may help address this question since Rossby waves are not supported in such a disk. If the formation of jets still occurs in such simulations, it would indicate that the spiral density wave activity causes their formation.

Finally, I consider possible differences that three-dimensional effects could have on these results. Because the system under consideration is rotating, even flow in a two-dimensional disk should still be layerwise two-dimensional. This can be estimated by the Ekman number for the flow, defined as $E = \nu/2\Omega H^2$. For an alpha-disk this expression becomes $E = \alpha m_H^2 \sqrt{(GM_* r)}/(2\Omega H^2) = \alpha/2$. Since observations suggest $\alpha = 10^{-3}$–10^{-2}, the flows in such disks should be well-approximated as two-dimensional except near planets where substantial departures from vertical hydrostatic equilibrium exist. This result suggests the sort of analysis I have presented here could prove useful in three dimensions.

Chapter 6

Summary

A major aim of this thesis was to develop a hydrodynamic code based on the cosmological TVD code by Trac & Pen (2003) suitable to perform studies of protoplanetary disks.

I have presented the details of the original code and the TVD algorithm in Chapter 3. I have modified the original code to work on cylindrical grids and have implemented a *FARGO*-like algorithm (Masset, 2000) which reduces the require computation time for near-Keplerian flows. I have demonstrated a reduction in the required simulation time by a factor of 6.6 when applied to a disk simulation of interest (§4.4). I have also tested the accuracy of the code on four different hydrodynamic problems: a Sedov-Taylor blast wave, an oblique Sod shock-tube, a Kelvin-Helmholtz instability and a cylindrical bow shock. The code performs well on these tests, where comparison with theoretical solutions is possible and shows similar performance to the piece-wise parabolic code, *VH-1*, on all four tests.

In Chapter 4 I performed a detailed study of the code's performance on the now-standard protoplanet-disk interaction problem. This problem examines the interaction of a single massive protoplanet embedded in a gas disk. The results from this comparison were consistent with the ensemble results published in de Val-Borro et al. (2006), which compared results a large variety of codes on this problem. In particular, both the latter project and my own simulations display libration islands in the coorbital region of the planet (at the L_4 and L_5 points) and an asymmetry in the density of those islands. They also correctly predict the sign the torque exerted on the planet by the disk and show comparable torque magnitudes. The growth and merging of vortices outside the planet's orbit in low-viscosity simulations are also confirmed. Furthering results presented in de Val-Borro et al. (2006), I have demonstrated the existence of additional *vortex lines* to the inside of the planet's orbit, and confirmed that the large vortex which forms exterior to the planet's orbit, substantially affects the measured torque on the planet. The large oscillations present in the torque correspond to repeated passes of the outer vortex by the planet and increasing the viscosity beyond $\nu = 10^{-5}$ damps the formation of the vortex and removes the oscillatory signature from the torque.

I have determined the level of numerical viscosity present within the code to be $\nu \lesssim 10^{-6}$ ($\alpha \lesssim 10^{-3.5}$), enabling the simulation of flows with Reynolds numbers on the order of $Re = UL \times 10^6$ for characteristic speeds, U and length-scales, L. The results of the viscosity tests indicate that using angular velocity rather

than angular momentum as a solution variable increases numerical diffusion by over a magnitude. This result complements previous work by Kley (1998).

In Chapter 5 I continued to explore the flow structures observed in the previous chapter, including the consideration of multiple protoplanets. I demonstrated spontaneous evolution of jets in disks with either one or two protoplanets. Because the flow is balanced, these jets are related to potential vorticity anomalies via PV inversion, $Q = \rho^{-1}\nabla \times u$. Formation of large vortices with non-negligible concentrations of gas ($\sim 4m_p$) are consistent with results by Klahr & Bodenheimer (2003) which show similar vortex formation in simulations of forced baroclinic turbulence. In addition I showed that the vortex formation process is damped once $\nu \gtrsim 10^{-5}$, however the jet formation process is more robust and jet structures are able to form even in viscous flows.

I found different results than those published by Kley (2000) which demonstrated efficient evacuation of inter-orbital gas between two protoplanets. Instead, simulations of a variety of planet orbital separations indicate that the efficiency of gas evacuation depends strongly on the separation of the planets' orbits and simulation of the parameters used by Kley show *inefficient* gas removal. These conclusions stand even when performed at low resolution and with viscosity. I proposed a possible explanation for the inefficient clearing, suggesting weakened transport associated with the jet structures which form between the protoplanets may be responsible. I tested this conjecture by measuring an effective alpha-parameter associated with the radial angular momentum transport. The measured parameter indicated outward transport of angular momentum as expected and showed that the amount of measured transport is indeed lower lower in the jet region, and is correlated with the potential vorticity.

Part II

Ejection of Planets from Binary Systems

Chapter 7

Introduction

The majority of solar-type stars in our neighborhood ($\sim$ 60%) are in binary or higher multiple systems (Duquennoy & Mayor, 1991). Despite this majority, there are still questions as to how many of these multiple-star systems host planets, and whether or not the planet formation process inside these systems differs markedly from that around single stars. Radial-velocity surveys have shown that $\sim$ 20% of the extrasolar planets reside in binaries (Eggenberger et al., 2004), but the true fraction is likely higher as these surveys select against observing known binaries.

While it is clear that much theoretical and observational effort is still needed to fully answer the above questions, significant progress has been made in one sub-area of this issue—the dynamical stability of planets in binary systems. The body of literature on this topic is extensive, with most studies focusing on numerical techniques. Hénon & Guyot (1970) numerically studied periodic planet orbits in circular binaries (circular restricted problem) as a function of the binary mass ratio. Benest (1993) included binary eccentricity in his study but only focused on a few astronomical systems. Rabl & Dvorak (1988) also considered eccentric binaries but limited their studies to equal-mass stars. Holman & Wiegert (1999, hereafter HW99) is the most comprehensive and homogeneous study to date. They numerically integrated (initially circular) planet orbits for 10^4 binary periods, and charted out the stability region as a function of binary separation, eccentricity, and mass ratio, for both the S-type (circum-stellar) and P-type (circum-binary) planetary orbits. Pilat-Lohinger & Dvorak (2002) have since included the effect of planet eccentricity but found it to be less important than the binary eccentricity. With the intention to quantify the confines of habitable zones around binary stars, Musielak et al. (2005) also investigated the stability of both S-type and P-type planetary orbits in circular binaries. To this end, they adopted a criterion for stability that differs slightly from that used in other works. However, they found results that largely agree with those from previous works, including those of HW99. Marzari et al. (2005) examined the stability of multiple planets in binary systems, including the effects of mutual planetary perturbations. In this case, interactions among the planets themselves appear to be the leading cause for instability. David et al. (2003) concentrated on studying ejection timescales for planets within the unstable region. They established an empirical formula for the ejection timescale that is a steep function of the periastron distance from the binary companion. Beyond a certain distance, however,

this trend is expected to break down and the ejection time become infinite (the system becomes stable). The location of this break is the boundary for which we are interested in searching. Since the aforementioned papers have confirmed the HW99 results, we focus on comparing our analytical results against HW99 exclusively. The numerical results of HW99 and Rabl & Dvorak (1988) uphold general expectations that the stability space (comprising the binary's eccentricity and the ratio of the planet's semimajor axis to that of the binary) shrinks with decreasing stellar separation, with increasing orbital eccentricity, and with increasing companion mass. However, the underlying physical mechanism for planet ejection has yet to be demonstrated. Moreover, current computational capabilities limit the integration time (up to 10^4 binary orbits in HW99) and permit only coarse-grid parameter searches. The former limitation may allow longer term instabilities to be missed while the latter blurs the transition from stability to instability, hiding the existence of possible '(in-)stability islands.'

In this thesis we study individual orbital resonances and the conditions for which they overlap, focusing on the orbits most relevant for radial-velocity searches—the so-called S-type orbits (Dvorak, 1984), where the planet orbits around one of the stars. The second star is considered to be an external perturber. Through this examination we expose the instability mechanism to be chaotic diffusion of the planet's orbital elements, caused by overlap of mean-motion subresonances. These subresonances typically lay atop one another, but this degeneracy is lifted by the secular forcing of the companion star, thereby increasing the extent of the instability phase space. We are also able to delineate the topology of transition between stability and instability and exclude the existence of longer term instabilities. This study is limited to coplanar systems. In Chapter 8 we discuss the necessary background and present the analytical arguments that allow us to determine the boundaries of stability. We also compare our calculations against numerical results from HW99 (§9.1) and results of ejection timescales from David et al. (2003) (§9.2). Section 9.3 includes an argument for scaling the instability boundary analytically, which allows us to exclude the existence of long-term instabilities, and section 9.4 finishes the chapter with a discussion of the limiting case of the circular restricted problem. Final comments and future directions are produced in Chapter 10.

Chapter 8

Resonance Overlap Formalism

8.1 The two-body problem

The two-body problem is discussed in a wide variety of textbooks (for example, see Murray & Dermott, 1999). Here we will merely quote the required results. Consider two masses, m_1 and m_2, located at positions r_1 and r_2 with respect to an inertial origin, and subject to their mutual gravitational forces, $F_1 = G\frac{m_1 m_2}{r^3}r$, and $F_2 = -G\frac{m_1 m_2}{r^3}r$, where $r = r_2 - r_1$. Their relative motion may be described by the equation

$$\frac{d^2 r}{dt^2} + \mu \frac{r}{r^3} = 0, \tag{8.1}$$

where $\mu = G(m_1 + m_2)$ and G is Newton's gravitational constant. The two objects' center of mass, $(m_1 r_1 + m_r r_2)/(m_1 + m_2)$, is constrained to move at constant velocity in a straight line with respect to the inertial origin. In terms of unit vectors, $\hat{r}$ and $\hat{\theta}$, which are respectively along and perpendicular to r we can write

$$\ddot{r} - r\dot{\theta}^2 = -\frac{\mu}{r^2}, \tag{8.2}$$

which may be integrated to yield

$$r = \frac{a(1 - e^2)}{1 + e\cos(\theta - \varpi)}. \tag{8.3}$$

This equation describes the motion of m_2 in terms of its separation, r, and true longitude, θ, for an elliptical orbit about m_1 with eccentricity, e, and semimajor axis, a. The angle, ϖ, is the longitude of pericenter. The length of time required to complete a single orbit is the period, T, given as $T = 4\pi^2 a^3/\mu$, which allows us to define a *mean motion*, $n = 2\pi/T$. The angular velocity, $\dot{\theta}$, is a nonlinear function of time and the longitude, θ. It is therefore convenient to introduce the mean longitude, λ, which is defined to be a linear function of time such that $\lambda = nt + \epsilon$ where ϵ is the initial mean longitude (called the *mean longitude at epoch*). Note that all longitudes are defined relative to an common, arbitrary reference direction.

8.2   The three-body problem via the disturbing function

Unlike the two-body problem, the general case of three-body motion is non-integrable. It may, however, be simplified by limiting oneself to the restricted problem where two of the bodies' masses are negligible in comparison with that of the third. One then considers each of the lesser objects to be independently executing two-body motion relative to the principal body with the inclusion of perturbing effects due to the other lesser mass body. The perturbing potential of the second lesser body is called the disturbing function. We apply this type of analysis to planet orbits within an S-type binary star system, where the planet orbits one of the stars and the orbit of the second, companion star encircles both the planet and the primary (Dvorak, 1984). We treat the external companion star and the planet as the lesser objects, and the central star as the primary. While the mass of the companion star is not small in comparison to the mass of the primary, m_c, we shall assume that its perturbation of the planet's orbit via the disturbing function does remain small due to its extended distance from the planet. We may express the disturbing function, $\mathcal{R}$, using Legendre polynomials in the form (eq. 6.24 Murray & Dermott, 1999, hereafter MD99)

$$\mathcal{R} = \frac{Gm'}{a'} \sum S_j \cos \varphi_j. \tag{8.4}$$

We use primed variables to refer to orbital elements of the companion star and unprimed variables for those of the planet. S_j is a strength coefficient that depends on the eccentricities of the planet and the companion, and on the ratio of their semimajor axes, $\alpha = a/a'$. The planet's mean motion, n, is expressed as $n^2 = G(m_c + m)/a^3$. The angle argument is

$$\varphi_j = j_1 \lambda' + j_2 \lambda + j_3 \varpi' + j_4 \varpi. \tag{8.5}$$

The summation in equation (8.4) is formally over all integer combinations, (j_1, j_2, j_3, j_4), that satisfy the *d'Alembert relation*: $j_1 + j_2 + j_3 + j_4 = 0$. Explicit expansions of S_j (in terms of the companion and planet's orbital elements) to low orders may be found in a variety of sources including Table B.3 of MD99. MD99 also provide formulas either to fully expand the disturbing function to arbitrary order or to expand those terms associated with a particular value of the argument, φ_j, to arbitrary order. Because the full expansion of the disturbing function to higher and higher orders quickly becomes unwieldy, it is useful to consider the expansion for limited values of the argument. Only those which are of longest period (most slowly varying) are most relevant to the dynamics. It is assumed that the individual contributions of shorter period angles will cancel out one another.

   To understand which combinations of angles one expects to have the longest period, consider the unperturbed (2-body) problem. In this problem the mean longitudes, (λ, λ'), vary linearly with time at a rate given by their mean motions while the longitudes of pericenter, (ϖ, ϖ'), are constant. In a similar manner, we expect that the mean longitudes will vary rapidly in comparison to the longitudes of pericenter in the perturbed system. Then any φ_j comprising a combinations of angles that do not contain the mean

longitudes will vary slowly in comparison to other combinations. These combinations in which the mean longitudes are absent are termed secular. Combinations which do contain the mean longitudes may also be slowly varying but only in a spatially localized region. To see this effect, consider the general argument, $\varphi_j = j_1\lambda' + j_2\lambda + j_3\varpi' + j_4\varpi$, where $\lambda' = n't + \epsilon'$ and $\lambda = nt + \epsilon$. Assuming that ϵ' and ϵ are small in comparison to $n't$ and nt, allows us to write $(j_1\lambda' + j_2\lambda) \approx (j_1n' + j_2n)t$. Then near locations where the orbital periods are approximately commensurate and satisfy, $(j_1n' + j_2n) \approx 0$, the total argument, $(j_1\lambda' + j_2\lambda)$ is of longer period than either orbital period individually. Arguments for which this condition occurs are termed resonant. In averaging over the possible combinations of (j_1, j_2, j_3, j_4) for a particular value of α we only consider secular and resonant terms to be relevant to the orbital dynamics and assume that the values of φ_j for other combinations are of short-enough period for their contributions to cancel out one another.

Removing these types of fast oscillating terms in equation (8.4) by integrating over an appropriately long time and keeping terms to the lowest order in eccentricities, we obtain the averaged disturbing function for the planet due to the perturbations of the companion star:

$$
\begin{aligned}
\mathcal{R} \;=\; & \frac{Gm'}{a'}\Big[f_{s1}(e^2 + e'^2) + f_{s2}ee'\cos(\varpi' - \varpi) \\
& + f_r e'^{|j_3|}e^{|j_4|}\cos\varphi_j \Big].
\end{aligned}
\tag{8.6}
$$

The first two terms in the brackets arise from the two lowest order secular interactions, ($j_1 = j_2 = 0$, with either $j_3 = j_4 = 0$ or $j_3 = -j_4 = 1$) while the last term accounts for MMRs situated at $j_1n' + j_2n \approx 0$. In particular, these could include resonances that share the same j_1 and j_2 values but have different j_3 (and therefore j_4) values. We call these 'subresonances' of a given MMR (j_1, j_2). Their importance will be expounded later. The coefficients f_{s1} and f_{s2} are functions of α alone. Explicit expressions for them are listed in Table B.3 of MD99. The f_r coefficient depends on α as well as the particular resonance under consideration. Note that the strength of the f_r term depends on the product of the planet's and companion's eccentricities to the exponent of the subresonance order. Explicit values for f_r are listed in MD99 only up to fourth order. To obtain values of f_r for higher order terms there are two expansion formulas provided (eqs. [6.36] and [6.113] in MD99). Appendix B details the expansion of equation (6.113) for the types of resonances we consider. We find that computer expansions of both formulas yield similar results and agree with the explicit formula at low order. As a caveat we note that both of these formulas make use of the series solution to Kepler's equation (eq. [2.52] in MD99) in order to determine the expansion in terms of an object's orbital elements. Since the series solution to Kepler's equation diverges for $e, e' > 0.6627434$, the expansion formulas are not valid above that value, and so we limit our studies to $e, e' < 0.6$.

8.3 Lagrange's equations

Lagrange's equations describe the variation in the orbital elements given an expansion of the disturbing function in terms of the orbital elements. Derivations of these equations may be found in Brouwer &

Clemence (1961) or Roy (1988). We state to lowest order in eccentricity the equations describing variations in n, e, ϖ and ϵ:

$$\dot{n} = -\frac{3}{a^2}\frac{\partial \mathcal{R}}{\partial \lambda} \tag{8.7}$$

$$\dot{e} = -\frac{1}{na^2 e}\frac{\partial \mathcal{R}}{\partial \varpi} \tag{8.8}$$

$$\dot{\varpi} = -\frac{1}{na^2 e}\frac{\partial \mathcal{R}}{\partial e} \tag{8.9}$$

$$\dot{\epsilon} = -\frac{e}{2na^2}\frac{\partial \mathcal{R}}{\partial e}. \tag{8.10}$$

Variations of the planet's orbital elements are then found using the disturbing function for the planet (eq. 8.6) as

$$\dot{n} = 3j_2 C_r n e'^{|j_3|} e^{|j_4|} \sin \varphi_j \tag{8.11}$$

$$\dot{e} = -C_{s2}\, e'\sin(\varpi' - \varpi) + j_4 C_r e'^{|j_3|} e^{|j_4|-1} \sin \varphi_j \tag{8.12}$$

$$\dot{\varpi} = 2C_{s1} + C_{s2}\frac{e'}{e}\cos(\varpi' - \varpi) + |j_4| C_r e'^{|j_3|} e^{|j_4|-2} \cos \varphi_j \tag{8.13}$$

$$\dot{\epsilon} = C_{s1}e^2 + \frac{C_{s2}}{2}e'e\cos(\varpi' - \varpi) + \frac{|j_4|}{2}C_r e'^{|j_3|} e^{|j_4|} \cos \varphi_j. \tag{8.14}$$

where the C-coefficients are related to the f-coefficients in equation (8.6) by $C_x = [Gm'/(na^2 a')]f_x \approx (m'/m_c)n\alpha f_x(\alpha)$. The variation in ϵ is smaller than that in ϖ by a factor of e^2 and can be neglected. Perturbations of the companion's orbital elements due to the planet are also ignored.

8.4   Exact resonance and resonance overlap

Exact resonance is defined to occur when both $\sin \varphi_j = 0$ and $\dot{\varphi}_j = 0$, *viz.* $j_1 n' + j_2 n + j_4 \dot{\varpi} = 0$. Near such a location, φ_j librates about its resonant value. Moving away from the resonance location, there exists a boundary beyond which φ_j changes from libration to circulation. This boundary defines the *width* of the resonance, namely, the range of space over which the resonance dominates the planet's dynamics. When the widths of two resonances become comparable to their separation, the planet can be affected simultaneously by these overlapping resonances. Mathematically, the overlap of two or more resonances causes neighboring trajectories to diverge exponentially with time (Chirikov, 1979; Wisdom, 1980). This exponential divergence occurs on the Lyapunov timescale (T_L), which, as argued by Holman & Murray (1996), is comparable to the libration timescale for the resonances in question. Except in the case of 'bounded chaos,' orbital parameters for the planet undergo unbounded random walks leading to ejection on a timescale called the event timescale (T_e). Though T_e fluctuates depending on the system, studies have shown that it roughly correlates with T_L (see, e.g. Lecar et al., 1992).

In our problem, the large companion mass produces strong secular forcing, making it different from

typical solar system dynamics problems. Firstly, even if the planet initially has zero eccentricity, it is forced to oscillate with an eccentricity amplitude (eqs. [8.12] and [8.13]),

$$e_{\text{sec}} = \frac{C_{s2}}{2C_{s1}} e',$$
(8.15)

on the short secular timescale $\Delta t \sim 2\pi/(2C_{s1})$. The magnitude of e_{sec} decreases with decreasing α but can be as large as $e'/2$ near the 3:1 resonance. We performed numerical integrations confirming that in our problem, the secular eccentricity scales with the binary eccentricity according to $e_{\text{sec}} \sim \alpha e'$. This scaling is expected based on the low order expansion for C_{s1} and C_{s2} in equation (8.15). For equal-mass binaries, the secular timescale ranges from ~ 20 planet orbital periods at the 3:1 resonance to ~ 1000 periods at the 20:1 resonance. While the secular timescale is likely too long when compared with the mean motion of the companion to allow for the 'evection resonance' (Touma & Wisdom, 1998) (when the companion mass dominates so that $\mu \rightarrow 1$, the evection resonance may become important—see §9.4), it is typically much shorter than the resonant timescale. Considering also that the resonant strength is at maximum when the value of e is at its largest, we can assume that $e = e_{\text{sec}}$ for the resonant interactions. Planets possessing a free eccentricity in addition to the forced value can reach higher overall eccentricity and will therefore be more unstable.

The second effect of the companion's strong secular forcing is to displace the centroid of different subresonances away from each other. In Appendix C, we derive expressions for the centroid position and the width of a resonance when secular forcing is important. While the width remains unchanged from the non-secular case, the centroid of a MMR is shifted from its nominal position, $(j_1n' + j_2n = 0)$, by an amount, $|\delta n| \sim |2(j_4/j_2)C_{s1}|$. When there is no secular forcing, the subresonances are degenerate and lie atop one another. Secular forcing displaces the subresonances from one another, greatly expanding the region over which overlap occurs. This effect is illustrated in Figure 8.1 for two groups of MMRs.

8.5   Chaotic diffusion and planet instability

Resonance overlap generates chaotic diffusion, but as pointed out by Murray (1992), under some circumstances resonance overlap will lead only to 'bounded chaos'—unpredictable but limited variations in the orbital elements. One such example is provided by Gladman (1993). Results from our numerical experiments (Fig. 9.1) as well as discussions in §9.4 suggest that this is not a major concern for determining the instability boundary. In the remaining discussion we, therefore, do not distinguish between the concepts of resonance overlap and planet instability.

Another question relates to whether the overlap between subresonances is as potent as that between distinct MMRs, thereby leading to planet instability on an astronomically interesting timescale (see the review by Malhotra, 1998). Our calculations in §9.2 suggest that this is indeed so—the chaotic diffusion caused by subresonance overlap leads to planet ejection within 1 Gyr (often much faster) for most systems.
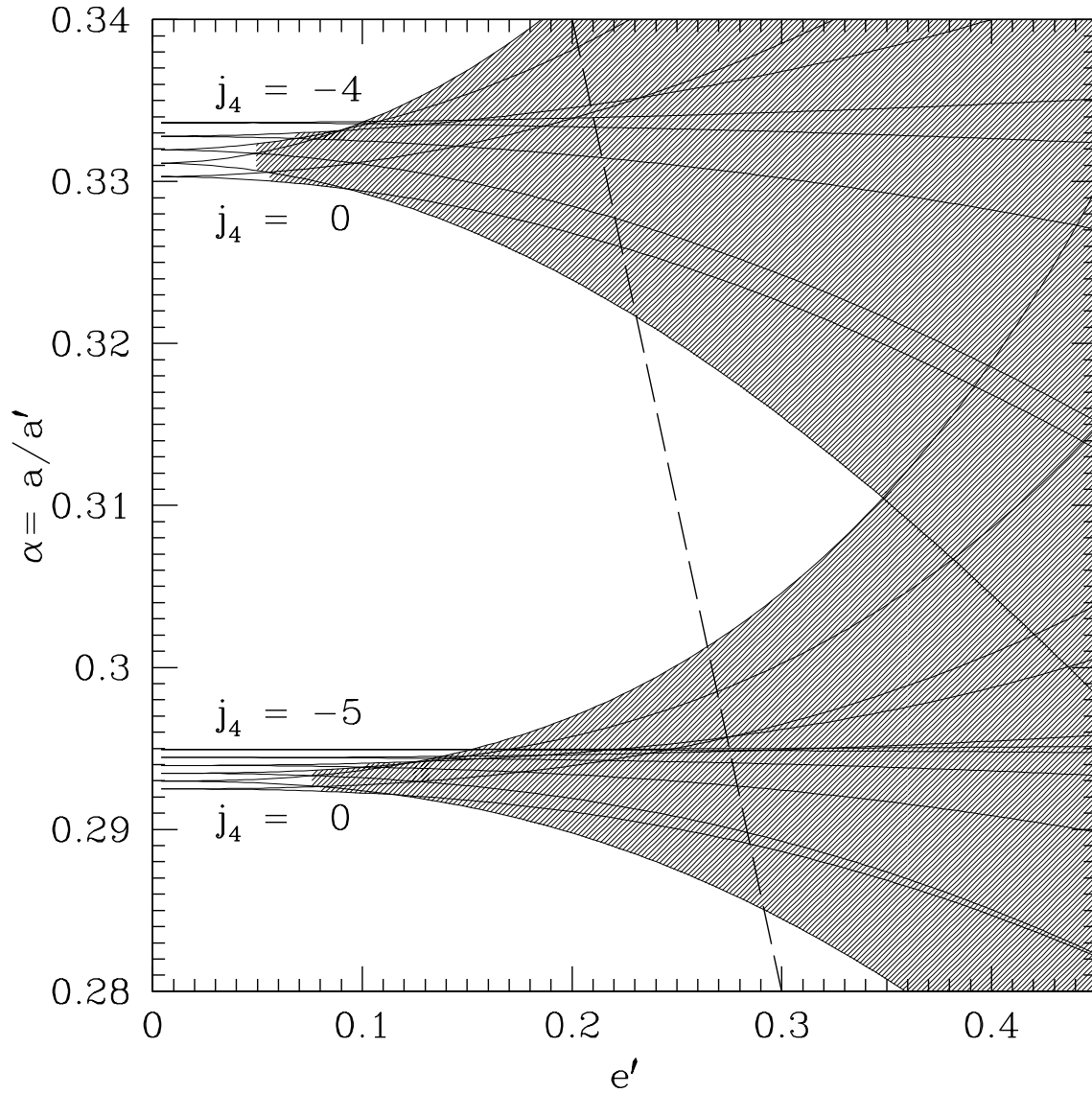
Figure 8.1 Location and width of various subresonances as a function of e', obtained for a $\mu = 0.1$ binary system. The top group are the subresonances of the 5:1 MMR (identified by their respective j_4 values) and the lower group, of the 6:1 MMR. We take the planet eccentricity to be the secularly forced value. The centroids of different subresonances within a distinct MMR are displaced from each other due to both secular and resonant forcing though the secular effect dominates at low values of e'. Shaded regions are regions of instability, as defined in the text. Overlap between subresonances of the same MMR covers a much larger region than overlap between distinct MMRs.

Chapter 9

Results

9.1 Comparison with numerical results

In our determination of regions of resonance overlap, we include resonances with $j_1 \geq 3$, $-4 \leq j_2 \leq -1$, and $-|j_1 + j_2| \leq j_3 \leq 0$. We restrict the value of j_2 since the strength of a resonance scales as $e'^{|j_3|} e^{|j_4|} \propto e'^{|j_1+j_2|}$ (eq. [8.6]). For a given orbital separation, α, the ratio of j_2/j_1 is determined by Kepler's third law: $\alpha^3 = (j_2/j_1)^2(1 - \mu)$; hence, the strongest resonances have $j_2 = -1$. In fact, we show that even the $j_2 = -4, -3$, and -2 resonances do not affect the instability boundary much. Moreover, while $j_3 = 0$ is the strongest subresonance in solar system dynamics (in light of Jupiter's small eccentricity), we find here that all subresonances are essential to determine the overlap region.

Coupling strengths are calculated using the aforementioned series expansion formulas in MD99 (Appendix B) while the location and width of each resonance are obtained using the formulas derived in Appendix C. The planet eccentricity is taken to be the secularly forced value (eq. [8.15]). All coefficients are evaluated at exact resonance, assuming the resonance width is small. In the (α, e') phase space, a region is designated as unstable if more than one resonance (or subresonance) spans it. We further assign a similar status, at the same value of e', to the entire extent in α of the subresonance in which this region is situated (see Fig. 8.1). Depending on its orbital phase, a planet situated within a single resonance (elsewhere spanned by additional resonances), but which is still outside of the region of overlap proper, may (or may not) librate into the latter. This definition of unstable regions ensures that all potentially unstable orbits are included. Again, our analytical study is limited to $e' < 0.6$ to ensure a convergent disturbing function.

Our full results are shown in Figure 9.1 for mass-ratio $\mu = 0.1$, and in Figure 9.2 for $\mu = 0.5$. As has been indicated in Figure 8.1, the instability boundary is jagged, with jutting peninsulas and narrow inlets. This is different from the smooth lines presented by HW99. However, their curves largely trace the outline of our results. The two sets of results can be considered consistent since HW99 carried out their investigation over a crude grid in $\alpha - e'$ space. To confirm this, we perform similar numerical integrations, with a much finer grid in a selected region of $\alpha - e'$ space. We adopt the Hierarchical Jacobi Symplectic integrator by Beust (2003), an add-on to the SWIFT package (Duncan et al., 1998) for studying dynamics in

multiple-star stellar systems. Planets are initialized to have random orbital phases and an eccentricity given by $e = e_{\rm sec}$ (initializing planets with zero eccentricity produces similar results). We integrate their orbits for 3000 binary periods. The stability of these orbits is indicated in the inset of Figure 9.1. The detailed topology agrees well with that obtained from our perturbation analysis, and in many cases, one can even identify the (sub)-resonances responsible for the instability. This suggests that resonance overlap and the consequential chaotic diffusion is the mechanism responsible for the planet instability observed in HW99's numerical investigation. Moreover, there is little evidence for 'bounded chaos' near the instability boundary, so that one can adopt the boundary of resonance overlap as the boundary for planet instability.

9.2 Sensitivity to initial conditions and relevant timescales

In an effort to delineate the differences between the chaotic dynamics existing within regions of resonance overlap and the regular dynamics existing just outside such regions, we numerically integrate two sets of two initially 'close' planets. Both sets of planets are situated near the 5:1 MMR in a binary system with mass ratio, $\mu = 0.1$, and eccentricity, $e' = 0.2$. The first set of two planets are situated directly within the region of overlap at $\alpha = 0.33$, while the second set are situated just outside the region of overlap at $\alpha = 0.32$. Figure 9.3 presents the results of integrating the first set of two planets initialized with identical orbital parameters except for a tenth of a degree difference in their orbital phase positions. The Lyapunov timescale is defined as the timescale for exponential divergence between two infinitesimally close orbits. We roughly estimate this timescale for the trajectories presented in Figure 9.3 and obtain $T_L \approx 10$ binary orbits. De-correlation in the semimajor axis and eccentricity becomes apparent to the eye after approximately 50 binary orbits. The libration time within this resonance, which one expects to be of the same order as the Lyapunov time, is ~ 54 binary orbits. The planets are ejected in turn after about 1500 and 3800 binary orbits. Further integrations at the same location with different initial orbital phases show a wide spread in the ejection times ranging from 50–4000 binary orbits, corresponding to $\sim 10^4 - 10^6$ years for a solar-mass binary at 50 AU. Murray & Holman (1997) and David et al. (2003) presented two different empirical expressions that relate the (widely scattered) ejection time (T_e) to the Lyapunov timescale. The former found a relationship between these two timescales (applicable to overlapping subresonances) given by $T_e/T' = 10^a(T_L/T')^b$ where $a = 1.45$ and $b = 1.68$, and where T' denotes the binary orbital period. Applying this formula to our case yields an ejection time of $T_e/T' \sim 2000$. The expression by David et al. (2003) yields a comparable value of $T_e/T' \sim 7400$. Within the scatter, both values agree with our experiments.

Adopting the expression by Murray & Holman (1997), we obtain ejection times for various resonances. Lower order resonances lead to faster ejection, while at the higher end, for example, the 30:1 resonance, we find $T_e/T' \leq 10^6$ for a system with mass-ratio, $\mu = 0.1$. This resonance (corresponding to $\alpha = 0.10$) defines the maximum stable orbit for the most eccentric binary orbit we consider ($e' = 0.66$). The ejection time corresponds to ~ 400 Myrs for a 50 AU binary and ~ 1 Gyrs for a 100 AU binary (all assuming a total system mass equal to one solar mass). We conclude that overlap of subresonances leads to planet ejection
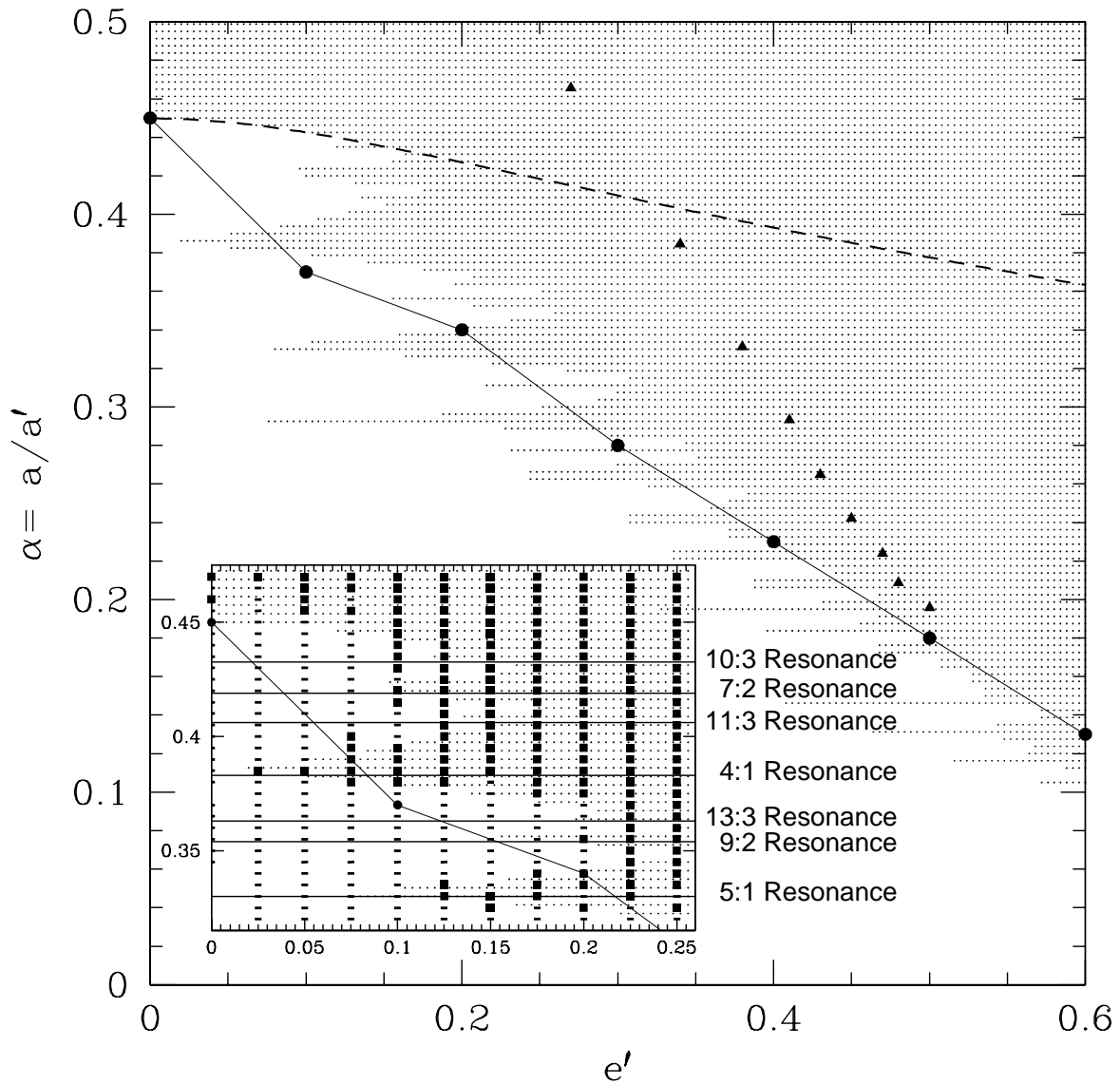
Figure 9.1 Stability diagram for planets in a $\mu = 0.1$ binary system. The solid curve connecting filled circles locates the maximum stable value of $\alpha = a/a'$ as obtained by HW99 while dots map regions of instability caused by resonance overlap. Resonances included in this calculation are described in the text. The instability boundary as it exists when considering only the distinct MMRs (keeping $j_3 = 0$) is denoted by filled triangles. Over the eccentricity range of interest, overlap between subresonances is the most significant source of planet instability. As $e' \to 0$, widths of most resonances approach 0 except for the 2:1 and 3:1 resonances. The dashed curve shows the lower confine of a 3:1 resonance—overlap between subresonances within the 3:1 MMR can explain instability in circular binaries. **Inset:** results of numerical integration over a selected region of $\alpha - e'$ space. Dashes represent planet orbits that remain stable for more than 3000 binary periods; filled squares represent unstable orbits. Horizontal lines indicate centroid locations of certain MMRs that are responsible for the jutting peninsulas. At each e' value, $j_2 = -1$ MMRs yield the shortest period unstable orbits. Stability for points near the instability boundary are sensitive to the initial conditions. Regions of resonance overlap coincide with that for planet instability and there is little evidence for 'bounded chaos.'
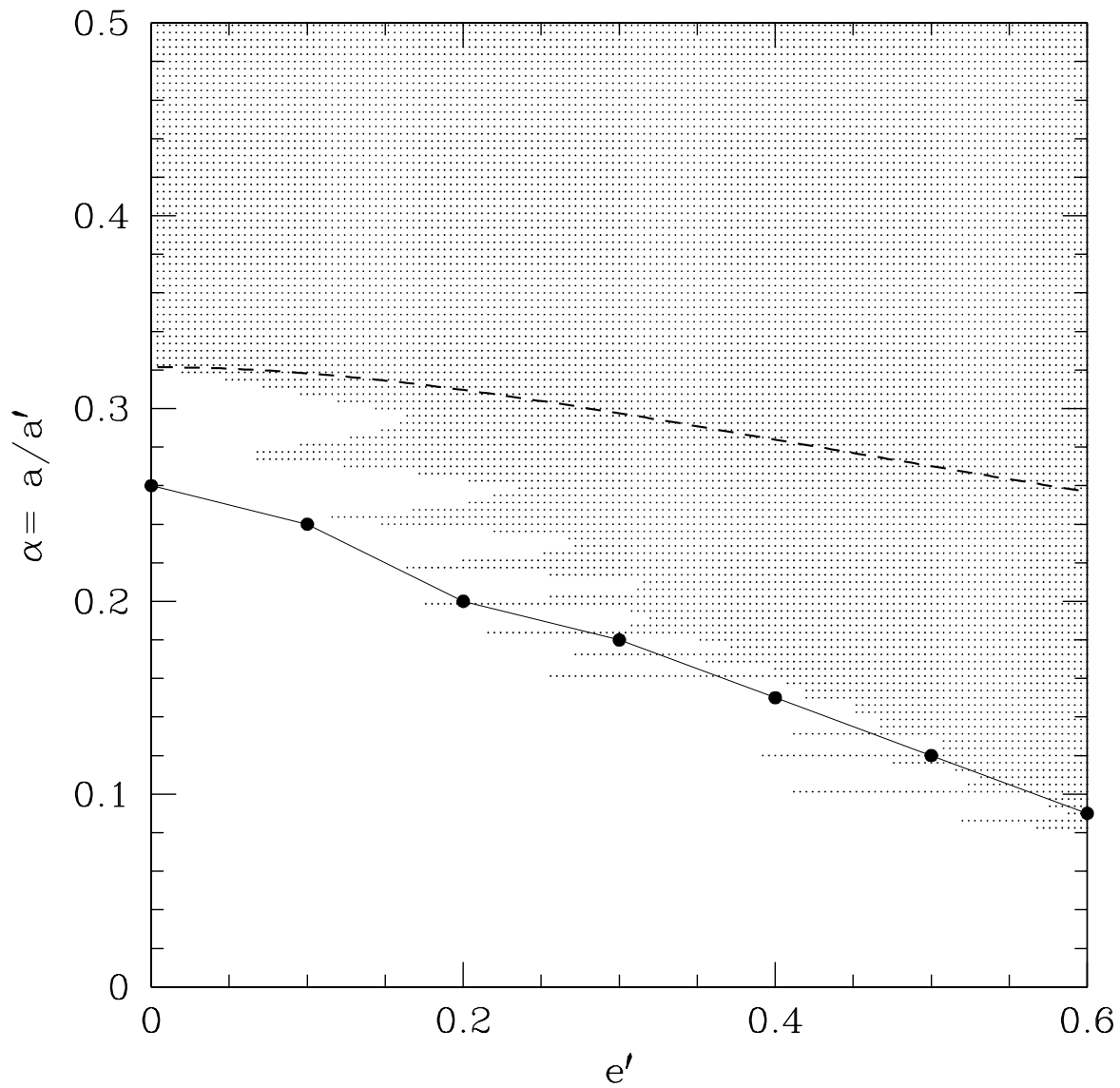
Figure 9.2 Stability diagram for planets in an equal-mass binary system ($\mu = 0.5$). Symbols are the same as those in Fig. 9.1. We obtain these results using a perturbation formula that is strictly valid only for $\mu \ll 1$—this may account for some of the discrepancy between our results (dots) and those of HW99 (solid curve).
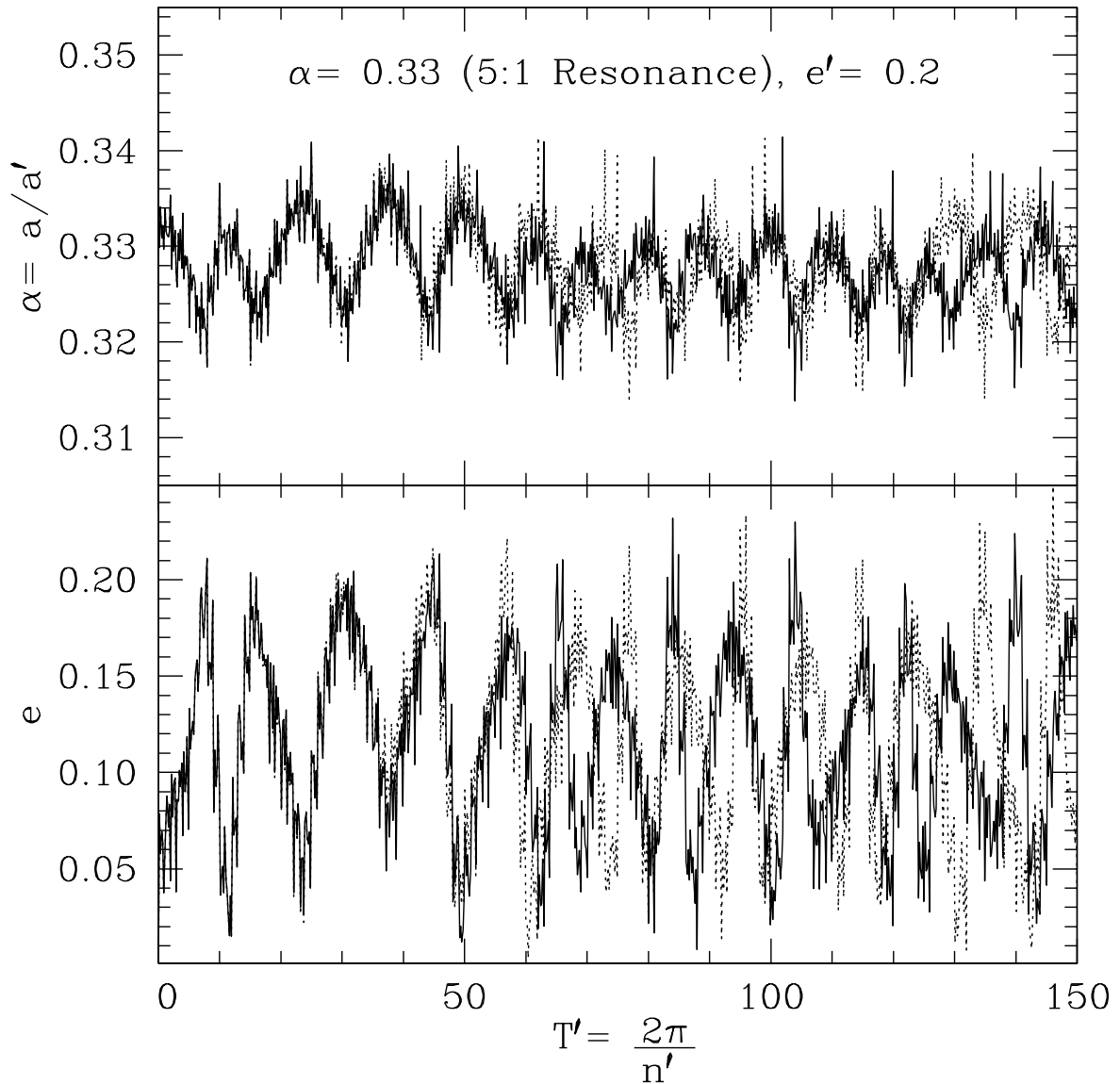
Figure 9.3 Numerical integration of two planets initialized with identical orbital parameters except for a tenth of a percent difference in the initial orbital phase. The resulting semimajor axis and eccentricity are plotted as functions of time (measured in binary orbital periods, T') in solid (or dotted) curves for each planet. The Lyapunov timescale is estimated to be $\sim$ 10 binary orbits, and de-correlation in the semimajor axis and eccentricity becomes apparent to the eye after approximately 50 binary orbits. The planets are ejected in turn after about 1500 and 3800 orbits.

on astronomically interesting timescales, for the parameters we have considered.

By contrast, the results of the second set of integrations where we instead situate the two planets just outside the region of resonance overlap at a location given by $\alpha = 0.32$, do not exhibit sensitivity on the initial conditions and no planet is ejected within our integration time ($10^4 T'$). The transition to instability occurs over a narrow region.

One major discrepancy between our results and those of HW99 can be observed in Figure 9.2 for equal-mass binaries: at low binary eccentricity, the HW99 curve falls below that obtained from our perturbation analysis. This discrepancy likely reflects failure of the expansion formula when μ is large.

A key question of interest asks, for a given binary (with fixed μ and e'), what is the longest period stable planet orbit? HW99 provided a fitting formula for the minimum unstable α as a function of μ and e'. Our results here indicate, however, that the minimum unstable α should be reduced by up to $\sim 20\%$ from their values. This reduction is related to the thin instability peninsulas evident in our Figures 9.1 and 9.2.

9.3   Analytical scaling of the instability boundary

In order to understand how the outline of the instability boundary depends on various parameters, we propose the following rough scaling argument. Let neighboring subresonances be spaced by Δn, where due to secular forcing, $\Delta n \approx 2|C_{s1}/j_2|$ (eq. [C.10]). Resonant interactions also modify the centroid of a resonance, but they are less important than the secular effect for small e'. The width of an individual subresonance, k, is expressed in equation (C.11), which in most cases can be simplified as $k \approx [4j_4^2/(3j_2)]\,|C_r|e'^{|j_3|}e^{|j_4|-2}$. Adopting $e = e_{\text{sec}} \approx \alpha e'$, and requiring resonance overlap ($\Delta n < 2k$), we find that instability occurs when

$$\alpha \geq \alpha_{\text{crit}} = \left(\frac{3}{4|j_4|^2 e'^{|j_1+j_2|-2}} \left| \frac{C_{s1}}{C_r} \right| \right)^{1/(|j_4|-2)}. \tag{9.1}$$

Defining $f_4 = |j_4/j_1|$, and relating j_1 to α by Kepler's Law, $|j_1/j_2|^2\alpha^3 = 1 - \mu$, we recast equation (9.1) as

$$\alpha \geq \alpha_{\text{crit}} \approx \left(\frac{3}{4f_4^2|j_2|^2(1-\mu)e'^{|j_1+j_2|-2}} \left| \frac{C_{s1}}{C_r} \right| \right)^{1/(f_4|j_1|-5)}, \tag{9.2}$$

where j_1 is also a function of α. Numerically, we observe that, regardless of the mass ratio and the resonance involved, $|C_{s1}/C_r|$ rises monotonically with f_4 and clusters around 0.02 when $f_4 \approx 0.5$. For simplicity, we solve for α_{crit} considering only $f_4 \approx 0.5$. The results are plotted in Figure 9.4 for three mass ratios. When only $j_2 = -1$ MMRs are considered, the $\mu = 0.1$ and $\mu = 0.2$ results sit atop each other falling somewhat below the respective HW99 curves at small values of e' and above them at large values. Besides errors resulting from our crude approximation in taking $f_4 = 0.5$, two other factors may contribute to this discrepancy. The first is that we are searching for the very minimum value of α at each value of e' that allows resonance overlap. As is shown in Fig. 9.1, this may lie up to 20% below the HW99 numerical result. The second factor is that we ignore overlap between distinct MMRs, which may be important at sufficiently
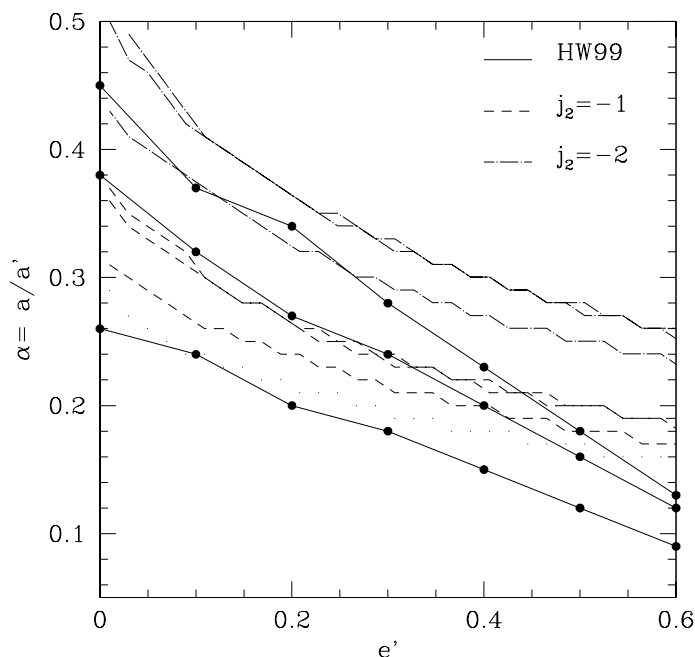
Figure 9.4 Comparison of instability boundaries obtained based on simple approximations of our analytical arguments (eq. [9.2]) and numerical results of HW99. The group of dashed curves represents the approximate overlap condition for $j_2 = -1$ MMRs while the dot-dashed group represents those for $j_2 = -2$ MMRs, both with $|j_4| = j_1/2$. Within each group, from top to bottom, the value of the mass-ratio is $\mu = 0.1, 0.2$ (these two curves almost coincide), and 0.5, respectively. If we recalculate the bottom-most dashed curve ($\mu = 0.5$) assuming the value of C_{s1}/C_r is 10 times smaller, we obtain the results shown in the dotted curve.

large e'. When $\mu = 0.5$, our curve consistently sits above the HW99 line, resembling the discrepancy shown in Fig. 9.2. This discrepancy, as we argue above, likely reflects failure of the expansion formula when μ is large (more below).

Despite these shortcomings, this simple analysis yields some useful insight. Comparing overlap conditions between those resonances with $j_2 = -1$ and those with $j_2 = -2$, reveals that the latter resonances always occur at a larger value of α for a given e' value. They are therefore not important for determining the instability boundary and we conclude that our neglect of $|j_2| > 4$ MMRs is valid. This point is further emphasized in the inset of Fig. 9.1. Based on this conclusion, we argue that instability boundaries obtained from finite-duration numerical integration are reliable, even though they may not detect long-term instabilities brought about by very high-order resonances (e.g., 50:3). A second point concerns the fact that we have ignored terms of order μ^2 in the expansion of the disturbing function, and that our C_{s1} and C_r coefficients are only correct to order-of-magnitude. We argue, however, that the instability boundary depends only on the ratio of C_{s1}/C_r (eq. [9.2]). Moreover, if, for instance, the true ratio of C_{s1}/C_r is smaller by a factor of 10 than our adopted value of 0.02, the instability boundary for $\mu = 0.5$ is moved downward in α by as much

as 10% (Fig. 9.4).

Our arguments here are based on very crude scaling relationships. They ought to be checked using more elaborate numerical experiments.

9.4   The circular, restricted limit

We focus on $e' = 0$ binaries to study the following two issues: the applicability of the Hill criterion for predicting planet instability, and the relevance of 'bounded chaos' that prevents us from using resonance overlap as a synonym for planet instability.

With the exception of those of first and second order, the widths of all other MMRs approach zero in circular binaries (see Appendix C). The resonance overlap condition in this limit is particularly easy to analyze. In the case where $\mu \rightarrow 0$ (the sun-asteroid-planet problem), Wisdom (1980) derived the overlap condition between first-order $(p + 1{:}p)$ resonances as described by $|1 - \alpha| = |a' - a|/a' \leq 1.307\mu^{2/7}$ (also see Duncan et al., 1989; Malhotra, 1998). For larger μ values, we argue that overlap between the $(3, -1, -1, -1)$ and $(3, -1, 0, -2)$ subresonances defines the lowest α value for which chaos can set in. Note that we calculate the resonance location and width using expansion formulas that are strictly valid only for $\mu \ll 1$. We suspect this approximation may lead to some uncertainty in the results in Fig. 9.5. Moreover, we have not considered the 'evection resonance', which becomes important as $\mu \rightarrow 1$ (Touma & Wisdom, 1998; Nesvorný et al., 2003).

The stability of planets in a circular binary can also be studied using the concept of Hill stability (e.g. see Murray & Dermott, 1999). In such systems, there exists an integral of motion, the Jacobi constant, which defines permitted regions of planetary movement. For a planet that begins with a circular orbit around one star (as in HW99), there is a critical value of α below which the zero-velocity curve with the same Jacobi constant is 'closed' and the planet cannot escape. For values of α greater than this critical value, the planet is allowed to escape by the Hill criterion but will *not* unless its orbit is chaotic due to overlapping resonances. In other words, the Hill criterion is a necessary but not a sufficient condition for planet *instability*.

To calculate α_{crit}, one needs to carefully consider the phrase 'begins with a circular orbit.' For $\mu \ll 1$ systems (analogous to the sun-asteroid-planet problem), it is more appropriate to actualize this condition by taking the sidereal velocity (velocity in the binary center-of-mass frame) to be Gm_c/a; while for $\mu \rightarrow 1$ systems (analogous to the planet-satellite-sun problem), the more reasonable approach is to set the synodic velocity (in the host star's rotating frame) to be Gm_c/a. For intermediate μ values, we adopt the approach that yields the higher value of α_{crit}. The resultant values of α_{crit} are plotted in Fig. 9.5 as a function of μ. In particular, in the limit where $\mu \rightarrow 1$, we find that $\alpha_{\text{crit}} = (1/3)R'_H = [(1 - \mu)/81]^{1/3}$ where R'_H is the Hill radius of the binary companion (Szebehely, 1978), and in the limit where $\mu \ll 1$, we find that $\alpha_{\text{crit}} = 1 - 2.1\mu^{1/3}$.

While the Hill criterion gives the energetic condition for planet instability, resonance overlap provides the dynamical cause. How do results from the Hill criterion compare with those from the resonance overlap
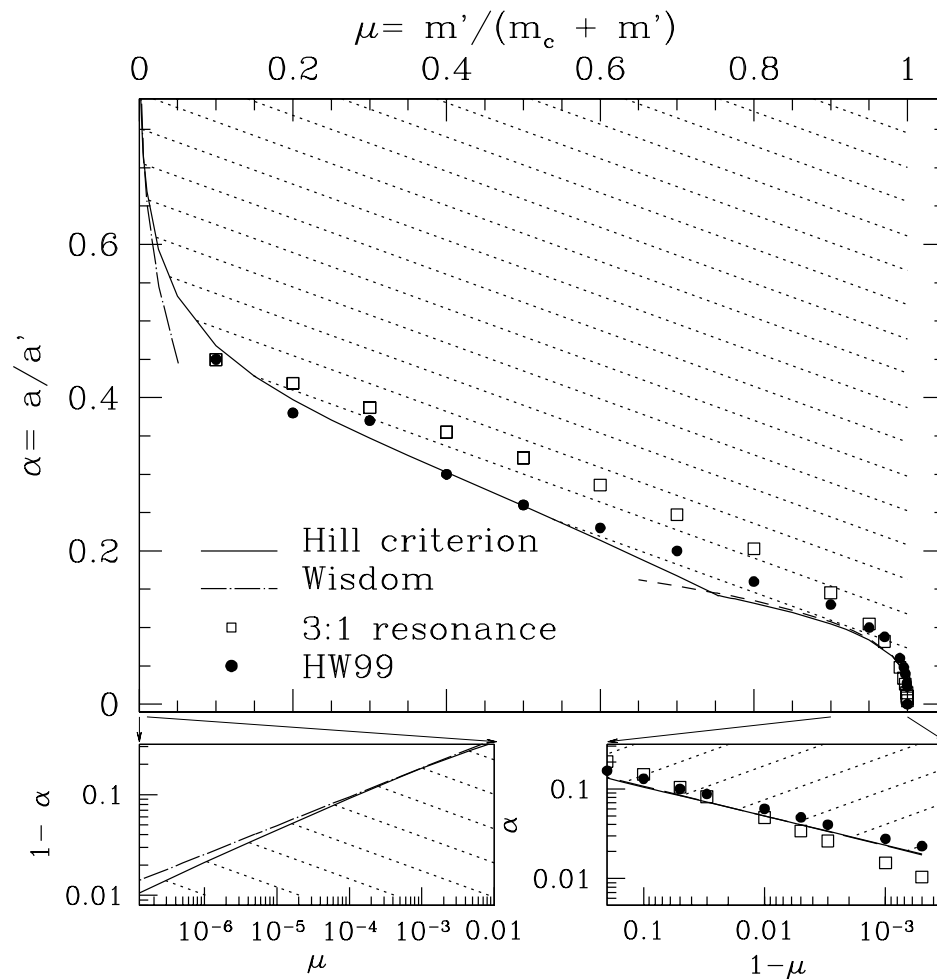
Figure 9.5 Instability boundary for circular binaries as a function of mass ratio μ. The solid curve depicts the result obtained based on the Hill criterion. Planets situated within shaded regions could potentially be ejected from the host star, though they will not be unless their orbits are chaotic. Overplotted are analytical results for the boundary of resonance overlap (and therefore chaos, but not necessarily ejection): as $\mu \to 0$, the overlap condition between $p + 1{:}p$ resonances yields $\alpha_{\text{crit}} \geq 1 - 1.307\mu^{2/7}$ (Wisdom, 1980); for larger μ values, overlap between the $(3, -1, -1, -1)$ and $(3, -1, 0, -2)$ subresonances occurs for α values above the open squares (this study). Locations of the open squares are calculated using expansion formula strictly valid only for $\mu \ll 1$. Also plotted (in filled circles) are the numerical results by HW99—planets situated above the filled circles are numerically shown to have unstable orbits. Left and right lower panels expand the view near $\mu = 0$ and $\mu = 1$, respectively. For $\mu \to 1$, the Hill criterion is well quantified as $(1/3)R_H = 0.23(1 - \mu)^{1/3}$ (dashed curve).

criterion? Intriguingly, they seem to closely trace each other over small μ, intermediate μ and large μ values (Fig. 9.5). The only exception is when $\mu \rightarrow 0$ (visible when $\mu \leq 10^{-4}$) for which the resonance overlap occurs over a larger range than does the Hill criterion. Gladman (1993) has studied this limit and concluded that 'bounded chaos,' producing unpredictable but limited variations in the orbital elements, reigns in the intervening region. This instance, however, is the only clear sign of bounded chaos in the circular, restricted problem.

Numerical results by HW99 (filled circles in Fig. 9.5) and our own simulations also confirm this seemingly coincidental agreement between the resonance overlap condition and the Hill criterion. It appears then, that in practice, the Hill criterion is not only a necessary, but also a sufficient condition for planet instability.

Chapter 10

Final Comments and Future Work

A planet in a binary system experiences both secular and resonant perturbations from the binary companion. It may be dislodged from its host star if it is simultaneously affected by two resonances. We find that overlap between subresonances of the same MMR accounts for the instability observed by HW99 and our own numerical integration. There is little evidence for bounded chaos, and the word "resonance overlap" can be interchanged with the word "orbital instability." Our instability boundaries largely agree with those obtained by HW99, albeit with many fine features. The jutting peninsulas and deep inlets in the instability boundary correspond to the instability (or stability) islands first observed by HW99. The presence of these islands suggests that the longest period stable orbit at each e'-value could be reduced by as much as 20% from the HW99 value. Moreover, our analysis suggests that overlap between very high-order resonances (e.g., 50:3) does not substantially modify the instability boundary: these weak resonances, while producing slow chaotic diffusion, which may be missed by finite-duration numerical integrations, do not contribute markedly to planet instability.

In detail, the centroids of different subresonances are displaced from each other by the strong secular forcing of the companion enlarging the phase space of resonance overlap. Chaotic diffusion caused by subresonance overlap is observed to be fast, unlike cases in the solar system. The longest ejection timescale in our study, corresponding to subresonance overlap within the 30:1 MMR, is $\sim 10^6$ binary orbits, or, 1 Gyr for a 100 AU solar-mass binary. For comparison, the 5:1 MMR overlap gives rise to an ejection time ~ 2000 binary orbits.

Compared with numerical integrations, our perturbation analysis has the following shortcomings: the perturbation strength is calculated accurate only to first order in the mass ratio between the companion and the host star, and the perturbation formula diverges for $e, e' > 0.66$.

As a final note, we raise the issue of stability in circular binary systems ($e' = 0$). While the Hill criterion (critical Jacobi constant) gives the energetic condition for planet instability, resonance overlap provides the dynamical cause. We observe that over almost the entire range of mass ratio, the Hill criterion and resonance overlap yield similar critical α-values, making the Hill criterion not only a necessary, but also a sufficient condition for planet instability.

Part III

Appendices

Appendix A

Flux form Euler equations in Cartesian and cylindrical coordinates

A.1 Cartesian coordinates

Expanding equations (2.10), (2.12), and (2.13) in Cartesian coordinates one obtains

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}[\rho u] + \frac{\partial}{\partial y}[\rho v] + \frac{\partial}{\partial z}[\rho w] = 0 \tag{A.1}$$

$$\frac{\partial \rho u}{\partial t} + \frac{\partial}{\partial x}[\rho u^2 + p] + \frac{\partial}{\partial y}[\rho uv] + \frac{\partial}{\partial z}[\rho uw] = -\rho\frac{\partial \phi}{\partial x} + \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} + \frac{\partial \sigma_{xz}}{\partial z} \tag{A.2}$$

$$\frac{\partial \rho v}{\partial t} + \frac{\partial}{\partial x}[\rho uv] + \frac{\partial}{\partial y}[\rho v^2 + p] + \frac{\partial}{\partial z}[\rho vw] = -\rho\frac{\partial \phi}{\partial y} + \frac{\partial \sigma_{yx}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} + \frac{\partial \sigma_{yz}}{\partial z} \tag{A.3}$$

$$\frac{\partial \rho w}{\partial t} + \frac{\partial}{\partial x}[\rho uw] + \frac{\partial}{\partial y}[\rho vw] + \frac{\partial}{\partial z}[\rho w^2 + p] = -\rho\frac{\partial \phi}{\partial z} + \frac{\partial \sigma_{zx}}{\partial x} + \frac{\partial \sigma_{zy}}{\partial y} + \frac{\partial \sigma_{zz}}{\partial z} \tag{A.4}$$

$$\frac{\partial e}{\partial t} + \frac{\partial}{\partial x}[u(e + p)] + \frac{\partial}{\partial y}[v(e + p)] + \frac{\partial}{\partial z}[w(e + p)] = -\rho\left[u\frac{\partial \phi}{\partial x} + v\frac{\partial \phi}{\partial y} + w\frac{\partial \phi}{\partial z}\right]$$

$$+ \frac{\partial}{\partial x}[(u\sigma_{xx} + v\sigma_{xy} + w\sigma_{xz})] + \frac{\partial}{\partial y}[(u\sigma_{yx} + v\sigma_{yy} + w\sigma_{yz})] + \frac{\partial}{\partial z}[(u\sigma_{zx} + v\sigma_{zy} + w\sigma_{zz})], \tag{A.5}$$

where $u_i \equiv (u, v, w)$ in Cartesian coordinates. Using the derived form for the stress tensor, equation (2.4) becomes

$$\sigma_{ij} = 2\mu\begin{bmatrix} \frac{\partial u}{\partial x} - \frac{1}{3}\boldsymbol{\nabla}\cdot\boldsymbol{u} & \frac{1}{2}(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}) & \frac{1}{2}(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}) \\ \frac{1}{2}(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}) & \frac{\partial v}{\partial y} - \frac{1}{3}\boldsymbol{\nabla}\cdot\boldsymbol{u} & \frac{1}{2}(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}) \\ \frac{1}{2}(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}) & \frac{1}{2}(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z}) & \frac{\partial w}{\partial z} - \frac{1}{3}\boldsymbol{\nabla}\cdot\boldsymbol{u} \end{bmatrix}, \tag{A.6}$$

where $\boldsymbol{\nabla}\cdot\boldsymbol{u} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}$.

A.2 Cylindrical Coordinates

To find the form of these equations in cylindrical coordinates one can perform the usual sequential applications of the chain rule. Or look up the forms in the back of a good Electromagnetism textbook. Either way, one obtains

$$\frac{\partial \rho}{\partial t} + \frac{1}{r}\frac{\partial}{\partial r}[\rho r u_r] + \frac{1}{r}\frac{\partial}{\partial \theta}[\rho u_\theta] + \frac{\partial}{\partial z}[\rho u_z] = 0 \tag{A.7}$$

$$\frac{\partial \rho u_r}{\partial t} + \frac{1}{r}\frac{\partial}{\partial r}[\rho r u_r^2 + pr] + \frac{1}{r}\frac{\partial}{\partial \theta}[\rho u_r u_\theta] + \frac{\partial}{\partial z}[\rho u_r u_z]$$

$$= -\rho\frac{\partial \phi}{\partial r} + \frac{1}{r}\frac{\partial r\sigma_{rr}}{\partial r} + \frac{1}{r}\frac{\partial \sigma_{\theta r}}{\partial \theta} + \frac{\partial \sigma_{zr}}{\partial z} - \frac{\sigma_{\theta\theta}}{r} + \frac{p}{r} + \frac{\rho u_\theta^2}{r} \tag{A.8}$$

$$\frac{\partial \rho u_\theta}{\partial t} + \frac{1}{r}\frac{\partial}{\partial r}[\rho r u_r u_\theta] + \frac{1}{r}\frac{\partial}{\partial \theta}[\rho u_\theta^2 + p] + \frac{\partial}{\partial z}[\rho u_\theta u_z]$$

$$= -\rho\frac{1}{r}\frac{\partial \phi}{\partial \theta} + \frac{1}{r}\frac{\partial r\sigma_{r\theta}}{\partial r} + \frac{1}{r}\frac{\partial \sigma_{\theta\theta}}{\partial \theta} + \frac{\partial \sigma_{z\theta}}{\partial z} + \frac{\sigma_{\theta r}}{r} - \frac{\rho u_r u_\theta}{r} \tag{A.9}$$

$$\frac{\partial \rho u_z}{\partial t} + \frac{1}{r}\frac{\partial}{\partial r}[\rho r u_r u_z] + \frac{1}{r}\frac{\partial}{\partial \theta}[\rho u_\theta u_z] + \frac{\partial}{\partial z}[\rho u_z^2 + p]$$

$$= -\rho\frac{\partial \phi}{\partial z} + \frac{1}{r}\frac{\partial r\sigma_{rz}}{\partial r} + \frac{1}{r}\frac{\partial \sigma_{\theta z}}{\partial \theta} + \frac{\partial \sigma_{zz}}{\partial z} \tag{A.10}$$

$$\frac{\partial e}{\partial t} + \frac{1}{r}\frac{\partial}{\partial r}[r u_r(e+p)] + \frac{1}{r}\frac{\partial}{\partial \theta}[u_\theta(e+p)] + \frac{\partial}{\partial z}[u_z(e+p)]$$

$$= -\rho\left[u_r\frac{\partial \phi}{\partial r} + u_\theta\frac{1}{r}\frac{\partial \phi}{\partial \theta} + u_z\frac{\partial \phi}{\partial z}\right] + \frac{1}{r}\frac{\partial}{\partial r}[r(u_r\sigma_{rr} + u_\theta\sigma_{\theta r} + u_z\sigma_{zr})]$$

$$+ \frac{1}{r}\frac{\partial}{\partial \theta}[u_r\sigma_{r\theta} + u_\theta\sigma_{\theta\theta} + u_z\sigma_{z\theta}] + \frac{\partial}{\partial z}[u_r\sigma_{rz} + u_\theta\sigma_{\theta z} + u_z\sigma_{zz}], \tag{A.11}$$

where $u_i \equiv (u_r, u_\theta, u_z)$ in Cylindrical coordinates. Similarly, the derived form for the stress tensor, equation (2.4), becomes

$$\sigma_{ij} = 2\mu\begin{bmatrix} \frac{\partial u_r}{\partial r} - \frac{1}{3}\nabla\cdot u & \frac{1}{2}(\frac{1}{r}\frac{\partial u_r}{\partial \theta} + \frac{\partial u_\theta}{\partial r} - \frac{u_\theta}{r}) & \frac{1}{2}(\frac{\partial u_r}{\partial z} + \frac{\partial u_z}{\partial r}) \\ \frac{1}{2}(\frac{\partial u_\theta}{\partial r} + \frac{1}{r}\frac{\partial u_r}{\partial \theta} - \frac{u_\theta}{r}) & \frac{1}{r}\frac{\partial u_\theta}{\partial \theta} + \frac{u_r}{r} - \frac{1}{3}\nabla\cdot u & \frac{1}{2}(\frac{\partial u_\theta}{\partial z} + \frac{1}{r}\frac{\partial u_z}{\partial \theta}) \\ \frac{1}{2}(\frac{\partial u_z}{\partial r} + \frac{\partial u_r}{\partial z}) & \frac{1}{2}(\frac{1}{r}\frac{\partial u_z}{\partial \theta} + \frac{\partial u_\theta}{\partial z}) & \frac{\partial u_z}{\partial z} - \frac{1}{3}\nabla\cdot u \end{bmatrix}, \tag{A.12}$$

where $\nabla\cdot u = \frac{1}{r}\frac{\partial r u_r}{\partial r} + \frac{1}{r}\frac{\partial u_\theta}{\partial \theta} + \frac{\partial u_z}{\partial z}$.

A.2.1 Angular momentum in cylindrical coordinates

Aligning the rotation vector along the z-axis, the vector $\boldsymbol{R}$ is simply the cylindrical radius, r. For $H = \rho r(u_\theta + r\Omega)$, equation (2.18) then becomes:

$$\frac{\partial H}{\partial t} + \frac{1}{r}\frac{\partial}{\partial r}[rHu_r] + \frac{1}{r}\frac{\partial}{\partial \theta}[Hu_\theta + pr] + \frac{\partial}{\partial z}[Hu_z]$$

$$= -\rho\frac{\partial \phi}{\partial \theta} + \frac{1}{r}\frac{\partial r^2\sigma_{r\theta}}{\partial r} + \frac{\partial \sigma_{\theta\theta}}{\partial \theta} + r\frac{\partial \sigma_{z\theta}}{\partial z} + \sigma_{\theta r}. \tag{A.13}$$

Appendix B

Expansion of the disturbing function for $j_1{:}j_2$ orbital resonances

Ellis & Murray (2000) derived an expansion of the disturbing function to maximum order N_{max} for a specific argument of the resonant angle, $\varphi_j = j_1\lambda' + j_3\lambda + j_3\varpi' + j_4\varpi + j_5\Omega' + j_6\Omega$, listed in MD99 as equations (6.113–6.126). For reference we list these equations here. The disturbing function due to an external perturber is given as

$$\mathcal{R} = \frac{\mu'}{a'}(\mathcal{R}_D + \alpha\mathcal{R}_E), \tag{B.1}$$

where

$$
\begin{aligned}
\mathcal{R}_D &= \sum_{i=0}^{i_{max}} \frac{(2i)!}{i!} \frac{(-1)^i}{2^{2i+1}} \alpha^i \\
&\times \sum_{s=s_{min}}^{i} \sum_{n=0}^{n_{max}} \frac{(2s-4n+1)(s-n)!}{2^{2n}n!(2s-2n+1)!} \sum_{m=0}^{s-2n} \kappa_m \frac{(s-2n-m)!}{(s-2n+m)!} \\
&\times (-1)^{s-2n-m} F_{s-2n,m,p}(I) F_{s-2n,m,p'}(I') \sum_{l=0}^{i-s} \frac{(-1)^s 2^{2s}}{(i-s-l)!l!} \\
&\times \sum_{\ell=0}^{\ell_{max}} \frac{(-1)^\ell}{\ell!} \sum_{k=0}^{\ell} \binom{\ell}{k} (-1)^k \alpha^\ell \frac{d^\ell}{d\alpha^\ell} b_{i+\frac{1}{2}}^{(j)}(\alpha) \\
&\times X_{-j_2}^{i+k,-j_2-j_4}(e) X_{j_1}^{-(i+k+1),j_1+j_3}(e') \\
&\times \cos[j_1\lambda' + j_3\lambda + j_3\varpi' + j_4\varpi + j_5\Omega' + j_6\Omega] \tag{B.2}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{R}_E &= -\kappa_m \frac{(1-m)!}{(1+m)!} F_{1,m,p}(I) F_{1,m,p'}(I') X_{-j_2}^{1,-j_2-j_4}(e) X_{j_1}^{-2,j_1+j_3}(e') \\
&\times \cos[j_1\lambda' + j_3\lambda + j_3\varpi' + j_4\varpi + j_5\Omega' + j_6\Omega], \tag{B.3}
\end{aligned}
$$

with $\kappa_0 = 1$ and $\kappa_m = 2$ for $m \neq 0$. We discuss the inclination functions, $F_{u,v,w}(I)$, the eccentricity functions, $X_c^{a,b}(e)$, and the Laplace functions, $b_{i+\frac{1}{2}}^{(j)}(\alpha)$, below. The following definitions apply for the direct part of the disturbing function:

$$
\begin{aligned}
q &= j_4 \\
q' &= -j_3 \\
\ell_{max} &= N_{max} - |j_5| - |j_6| \\
(p_{min}, p'_{min}) &= \begin{cases} (\frac{1}{2}|a|, 0) & \text{if } a = j_5 + j_6 < 0 \\ (0, \frac{1}{2}|a|) & \text{if } a = j_5 + j_6 \geq 0 \end{cases} \\
s_{min} &= \max(p_{min}, p'_{min}, j_6 + 2p_{min}, -j_5 + 2'_{min}) \\
i_{max} &= \text{Int}[(N_{max} - |j_3| - |j_4|)/2] \\
n_{max} &= \text{Int}[(s - s_{min})/2] \quad s_{min} \leq s \leq i \\
m_{min} &= \begin{cases} 0 & \text{if } s, j_5 \text{ are both even or both odd} \\ 1 & \text{if } s, j_5 \text{ are neither both even nor both odd} \end{cases} \\
(p, p') &= (-j_6 - m + s - 2n, j_5 - m + s - 2n) \quad p_{min}, p'_{min} \leq p, p' \leq s - 2n \\
j &= |j_2 + i - 2l - 2n - 2p + q|
\end{aligned}
$$

where $\text{Int}[x]$ takes the integer part of x.

We are interested in expanding arguments up to order $N_{max} = |j_1 + j_2|$ for coplanar systems where $j_5 = j_6 = 0$. We consider terms of the form $j_1 > 0$, $j_2 < 0$ and $-|j_1 + j_2| \leq j_3 \leq 0$. Along with the D'Alembert relation, we can then determine values for the above definitions to be:

$$
\begin{aligned}
m_{min} &= 0 \\
(p_{min}, p'_{min}) &= 0 \\
s_{min} &= 0 \\
\ell_{max} &= |j_1 + j_2| \\
i_{max} &= \text{Int}[(|j_1 + j_2| - |j_3| - |j_4|)/2] = 0 \\
n_{max} &= 0 \\
p &= p' = 0 \\
q &= j_4 \\
q' &= -j_3 \\
j &= |j_2 + j_4|
\end{aligned}
$$

With these values equation (B.2) reduces to

$$
\begin{aligned}
\mathcal{R}_D &= \frac{1}{2} \sum_{\ell=0}^{|j_1+j_2|} \frac{(-1)^\ell}{\ell!} \sum_{k=0}^{\ell} \binom{\ell}{k} (-1)^k \alpha^\ell \frac{d^\ell}{d\alpha^\ell} b_{\frac{1}{2}}^{(|j_2+j_4|)}(\alpha) \\
&\quad \times X_{-j_2}^{k,-j_2-j_4}(e) X_{j_1}^{-(k+1),j_1+j_3}(e') \\
&\quad \times \cos[j_1\lambda' + j_3\lambda + j_3\varpi' + j_4\varpi],
\end{aligned}
\tag{B.4}
$$

where we have use the fact that the inclination function, $F_{u,v,w}(I)$ (defined in MD99), is unity for $u = v = w = 0$.

The functions, $X_c^{a,b}(e)$, are Hansen coefficients defined for $\hat{\alpha} = \max(0, c - b)$ and $\hat{\beta} = \max(0, b - c)$ as

$$
X_c^{a,b}(e) = e^{|c-b|} \sum_{\sigma=0}^{\infty} X_{\sigma+\hat{\alpha},\sigma+\hat{\beta}}^{a,b} e^{2\sigma}
\tag{B.5}
$$

in terms of the Newcomb operators, $X_{c,d}^{a,b}$, which are themselves defined recursively as

$$
\begin{aligned}
X_{0,0}^{a,b} &= 1 \\
X_{1,0}^{a,b} &= b - a/2 \\
4c X_{c,0}^{a,b} &= 2(2b - a)X_{c-1,0}^{a,b+1} + (b - a)X_{c-2,0}^{a,b+2}.
\end{aligned}
\tag{B.6}
$$

We will see that cases where $d \neq 0$ are not relevant for the expansions we consider. The Newcomb operators also satisfy $X_{c,d}^{a,b} = 0$ if either c or d are negative, and have the symmetry property that $X_{c,d}^{a,b} = X_{d,c}^{a,-b}$ if $d > c$. MD99 state that the Hansen coefficients need only be calculated to order $N_{max} - |j_3| - |j_5| - |j_6|$ in e and to order $N_{max} - |j_4| - |j_5| - |j_6|$ in e'. For the cases we consider this condition requires Hansen coeffcients be calculated to orders $|j_4|$ and $|j_3|$ in e and e' respectively. This means $\sigma = 0$ is the only term required in both e and e' and we obtain

$$
\begin{aligned}
X_{-j_2}^{k,-j_2-j_4}(e) &= e^{|j_4|} X_{-j_4,0}^{k,j_2+j_4} \tag{B.7} \\
X_{j_1}^{-(k+1),j_1+j_3}(e') &= e'^{|j_3|} X_{-j_3,0}^{-(k+1),j_1+j_3}, \tag{B.8}
\end{aligned}
$$

where we have made use of the symmetry property in writing $X(e)$. We computationally solve for the Newcomb coefficients as required by implementing a recursive subroutine.

The Laplace coefficients, $b_s^{(j)}(\alpha)$ where $s = i + 1/2$ are defined by

$$
\frac{1}{2} b_s^{(j)}(\alpha) = \frac{1}{2\pi} \int_0^{2\pi} \frac{\cos j\psi \, d\psi}{(1 - 2\alpha \cos \psi + \alpha^2)^s}.
\tag{B.9}
$$

Because numerical integration of the above definition is slow to converge, we choose to solve for the Laplace

coefficients using the equivalent series form:

$$\frac{1}{2}b_s^{(j)}(\alpha) = \frac{s(s+1)...(s+j-1)}{1\cdot 2\cdot 3\cdot ...\, j}\alpha^j$$

$$\times \left[1 + \frac{s(s+j)}{1(j+1)}\alpha^2 + \frac{s(s+1)(s+j)(s+j+1)}{1\cdot 2(j+1)(j+2)}\alpha^4 + ...\right]. \tag{B.10}$$

The ℓ-th order derivatives of the Laplace function were also implemented as series solutions.

The indirect part of the disturbing function is only nonzero in cases where p, p' and m are integers equal to 0 or 1. An analysis of the integers involved shows that for cases in which we are interested, we have $p = p'$. Also when $p, p' = 1$ we have $m = -1$ (so that the indirect part does not appear) and when $p, p' = 0$, $m = 1$. This last possibility contributes and indirect term of the form

$$\mathcal{R}_E = -X_{-j_2}^{1,-j_2-j_4}(e)X_{j_1}^{-2,j_1+j_3}(e')\cos[j_1\lambda' + j_3\lambda + j_3\varpi' + j_4\varpi] \tag{B.11}$$

when $p = (j_2 + j_4 + 1)/2 = 0$. Because $j_2, j_4 < 0$ this last restriction only holds when $j_2 = -1$ and $j_4 = 0$. In writing $\mathcal{R}_E$, we have used the fact that $F_{1,1,0}(0) = 1$.

Appendix C

Width of a mean-motion resonance under secular forcing

Murray & Dermott (1999) presented a derivation for the width of a MMR when the resonance angle evolves due to a single resonance. In our situation with a massive third body, secular effects on the resonance angle have to be taken into account. We show here how this modifies the resonance width and resonance centroid.

The relevant resonance angle as well as its time derivatives are

$$\varphi_j = j_1 \lambda' + j_2 \lambda + j_3 \varpi' + j_4 \varpi \tag{C.1}$$

$$\dot{\varphi}_j = j_1 n' + j_2 n + j_4 \dot{\varpi} \tag{C.2}$$

$$\ddot{\varphi}_j = j_2 \dot{n} + j_4 \ddot{\varpi}. \tag{C.3}$$

The time-variations of n', e', ϖ', and ϵ' due to the influence of the planet are neglected as the planet can effectively be thought of as a test mass ($m/m_c \ll 1$). We also neglect variations in ϵ as previously mentioned.

We take the time derivative of equation (8.13), substitute equations (8.12) and (8.13) into the right-hand side, and use the resulting equations to recast equation (C.3) into the form

$$\ddot{\varphi}_j = \left[3 j_2^2 \tilde{C}_r n e^{|j_4|} - j_4 |j_4| \tilde{C}_r e^{|j_4|-2} (j_1 n' + j_2 n) - 2|j_4|^3 C_{s1} \tilde{C}_r e^{|j_4|-2} \right] \sin \varphi_j$$
$$- \left[|j_4|^3 \tilde{C}_r^2 e^{2|j_4|-4} \right] \sin 2\varphi_j, \tag{C.4}$$

where $\tilde{C}_r = C_r e'^{|j_3|}$. This reduces to equation (8.63) of MD99 when $|j_4| = 1$ and $C_{s1} = 0$. In deriving this equation, we have made some simplifying assumptions. Firstly, we have ignored the time-dependence of C_{s1} and C_r, which are in reality both functions of α. Secondly, we have neglected the C_{s2} terms in equations (8.12) and (8.13) as we expect their time-averaged contributions to be negligible.

We look for a solution of the system that is pendulum-like as in the case without secular forcing where the system satisfies a mechanical energy equation of the form $E = \frac{1}{2} \dot{\varphi}_j^2 + k \sin^2 \frac{1}{2} \varphi_j$. For a pendulum-like solution we expect that $n = n_0 + k \cos[\varphi_j/2]$, where n_0 is the mean motion associated with the nominal value

of the resonance and k is a constant that describes the amplitude of the oscillation, or equivalently, the width of the resonance. The choice of the angular form, $\cos[\varphi_j/2]$, is determined by the libration amplitude of the resonant angle φ_j ($-\pi$ to π) as well as the presumed angle where maximum change in the mean motion occurs ($\varphi_j = 0$). The latter applies when $C_r < 0$ and shifts to $\varphi_j = \pi$ when $C_r > 0$; however, the final result does not depend on the presumed sign of C_r.

Substituting the assumed form of the solution into equation (C.2) and evaluating at maximum potential energy where $(\varphi_j, \dot{\varphi}_j) = (\pi, 0)$ allows us to write

$$j_1 n' + j_2 n = j_2 k \cos \frac{1}{2}\varphi_j - 2 j_4 C_{s1} + j_4 |j_4| \tilde{C}_r e^{|j_4|-2}. \tag{C.5}$$

We obtain an energy equation by substituting this expression into equation (C.4), multiplying by $\dot{\varphi}_j$ and integrating in time to yield

$$
\begin{aligned}
\frac{1}{2}\dot{\varphi}_j^{\,2} \;=\;& \left[-6 j_2^2 \tilde{C}_r n e^{|j_4|} + 2|j_4|^4 \tilde{C}_r^{\,2} e^{2|j_4|-4} \right] \cos^2 \frac{1}{2}\varphi_j \\
&+ \frac{4}{3} j_2 j_4 |j_4| k \tilde{C}_r e^{|j_4|-2} \cos^3 \frac{1}{2}\varphi_j - |j_4|^3 \tilde{C}_r^{\,2} e^{2|j_4|-4} \sin^2 \varphi_j.
\end{aligned}
\tag{C.6}
$$

Evaluating this expression at $\varphi_j = 0$ we obtain

$$\dot{\varphi}_{j\,\varphi_j=0}^{\,2} = -12 j_2^2 \tilde{C}_r n e^{|j_4|} + 4|j_4|^4 \tilde{C}_r^{\,2} e^{2|j_4|-4} + \frac{8}{3} j_2 j_4 |j_4| k \tilde{C}_r e^{|j_4|-2}. \tag{C.7}$$

We obtain a second equation for $\dot{\varphi}_{j\,\varphi_j=0}^{\,2}$ by substituting equation (C.5) into equation (C.2), evaluating the result at $\varphi_j = 0$ and squaring:

$$\dot{\varphi}_{j\,\varphi_j=0}^{\,2} = j_2^2 k^2 + 4 j_2 j_4 |j_4| k \tilde{C}_r e^{|j_4|-2} + 4|j_4|^4 \tilde{C}_r^{\,2} e^{2|j_4|-4}. \tag{C.8}$$

Equating these two expressions we obtain the following quadratic in k:

$$k^2 + \frac{4}{3}\frac{j_4|j_4|}{j_2} k \tilde{C}_r e^{|j_4|-2} + 12\tilde{C}_r n e^{|j_4|} = 0. \tag{C.9}$$

Assuming $C_r > 0$ we would have chosen $n = n_0 + k\sin[\varphi_j/2]$ such that $\dot{\varphi}_j = 0$ at $\varphi_j = 0$ and obtained a quadratic in k with opposite signs in front of the two $\tilde{C}_r$ terms.

Solving for k we have expressions for the centroid of the resonance (eq. [C.5]) and the resonance amplitude, irrespective of the sign of C_r:

$$j_1 n' + j_2 n_0 \;=\; -2 j_4 C_{s1} - j_4 |j_4| \|\tilde{C}_r| e^{|j_4|-2}, \tag{C.10}$$

$$k \;=\; \frac{2}{3}\frac{j_4|j_4|}{j_2}|\tilde{C}_r| e^{|j_4|-2} \pm \sqrt{12|\tilde{C}_r| n e^{|j_4|}}\left(1 + \frac{j_4^4|\tilde{C}_r| e^{|j_4|-4}}{27 j_2^2 n}\right)^{1/2}. \tag{C.11}$$

The secular term is important for shifting the centroid of the resonance, but does not contribute to the width

of the resonance. In fact, the width formula is identical to equation (8.75) of MD99 where secular forcing is ignored.

The simple pendulum approach applies only when the resonant width is small, i.e., $\delta n = n_{\max} - n_{\min} \ll n$. Moreover, assuming that e is driven by the secular interaction to a value that is proportional to e' (eq. [8.15]), most MMRs have widths which approach 0 as $e' \to 0$. The first-order (e.g., 2:1) and second-order (e.g., 3:1) resonances that satisfy $j_4 \neq 0$ are exceptions; the width diverges for the former and approaches a constant for the latter.

Bibliography

Adams, F. C., Lada, C. J., & Shu, F. H., Spectral evolution of young stellar objects. 1987, ApJ, 312, 788

Adams, F. C., Shu, F. H., & Lada, C. J., The disks of T Tauri stars with flat infrared spectra. 1988, ApJ, 326, 865

Artymowicz, P., On the Wave Excitation and a Generalized Torque Formula for Lindblad Resonances Excited by External Potential. 1993, ApJ, 419, 155

Balbus, S. A. & Hawley, J. F., Instability, turbulence, and enhanced transport in accretion disks. 1998, Reviews of Modern Physics, 70, 1

Baldwin, M. P., Rhines, P. B., Huang, H.-P., & McIntyre, M. E., The Jet-Stream Conundrum. 2007, Science, 315, 467

Bally, J., O'Dell, C. R., & McCaughrean, M. J., Disks, Microjets, Windblown Bubbles, and Outflows in the Orion Nebula. 2000, AJ, 119, 2919

Beckwith, S. V. W. & Sargent, A. I., Circumstellar disks and the search for neighbouring planetary systems. 1996, Nature, 383, 139

Benest, D., Stable planetary orbits around one component in nearby binary stars. II. 1993, Celestial Mechanics and Dynamical Astronomy, 56, 45

Beron-Vera, F. J., Brown, M. G., Rypina, I. I., Olascoaga, M. J., Kocak, H., , & Udovydchenkov, I. A., Zonal jets as transport barriers in planetary atmospheres. 2007, submitted to Science

Beust, H., Symplectic integration of hierarchical stellar systems. 2003, A&A, 400, 1129

Bodenheimer, P., Hubickyj, O., & Lissauer, J. J., Models of the in Situ Formation of Detected Extrasolar Giant Planets. 2000, Icarus, 143, 2 [ADS]

Brouwer, D. & Clemence, G. M. 1961, Methods of celestial mechanics, Methods of celestial mechanics (New York: Academic Press, 1961)

Bryden, G., Chen, X., Lin, D. N. C., Nelson, R. P., & Papaloizou, J. C. B., Tidally Induced Gap Formation in Protostellar Disks: Gap Clearing and Suppression of Protoplanetary Growth. 1999, ApJ, 514, 344 [ADS]

Bryden, G., Różyczka, M., Lin, D. N. C., & Bodenheimer, P., On the Interaction between Protoplanets and Protostellar Disks. 2000, ApJ, 540, 1091 [ADS]

Chirikov, B. V., . 1979, Phys. Rep., 52, 265

Cho, J. & Polvani, M., The formation of jets and vortices from freely-evolving shallow water turbulence on the surface of a sphere. 1996, Physics of Fluids, 8, 1531

Colella, P. & Woodward, P. R., The Piecewise Parabolic Method (PPM) for Gas-Dynamical Simulations. 1984, Journal of Computational Physics, 54, 174

Crida, A. & Morbidelli, A., Cavity opening by a giant planet in a protoplanetary disc and effects on planetary migration. 2007, MNRAS, 377, 1324

Crida, A., Morbidelli, A., & Masset, F., Simulating planet migration in globally evolving disks. 2007, A&A, 461, 1173

D'Angelo, G., Lubow, S. H., & Bate, M. R., Evolution of Giant Planets in Eccentric Disks. 2006, ApJ, 652, 1698

David, E.-M., Quintana, E. V., Fatuzzo, M., & Adams, F. C., Dynamical Stability of Earth-like Planetary Orbits in Binary Systems. 2003, PASP, 115, 825

de Val-Borro, M., Artymowicz, P., D'Angelo, G., & Peplinski, A., Vortex generation in protoplanetary disks with an embedded giant planet. 2007, ArXiv e-prints, 706

de Val-Borro, M., Edgar, R. G., Artymowicz, P., Ciecielag, P., Cresswell, P., D'Angelo, G., Delgado-Donate, E. J., Dirksen, G., Fromang, S., Gawryszczak, A., Klahr, H., Kley, W., Lyra, W., Masset, F., Mellema, G., Nelson, R. P., Paardekooper, S.-J., Peplinski, A., Pierens, A., Plewa, T., Rice, K., Schäfer, C., & Speith, R., A comparative study of disc-planet interaction. 2006, MNRAS, 370, 529

Doppmann, G. W., Greene, T. P., Covey, K. R., & Lada, C. J., The Physical Natures of Class I and Flat-Spectrum Protostellar Photospheres: A Near-Infrared Spectroscopic Study. 2005, AJ, 130, 1145

Duncan, M., Quinn, T., & Tremaine, S., The long-term evolution of orbits in the solar system - A mapping approach. 1989, Icarus, 82, 402

Duncan, M. J., Levison, H. F., & Lee, M. H., A Multiple Time Step Symplectic Algorithm for Integrating Close Encounters. 1998, AJ, 116, 2067

Duquennoy, A. & Mayor, M., Multiplicity among solar-type stars in the solar neighbourhood. II - Distribution of the orbital elements in an unbiased sample. 1991, A&A, 248, 485

Dvorak, R., Numerical experiments on planetary orbits in double stars. 1984, Celestial Mechanics, 34, 369

Dwarkadas, V., Plewa, T., Weirs, G., Tomkins, C., & Marr-Lyon, M., Simulation of Vortex–Dominated Flows Using the FLASH Code. 2004, ArXiv Astrophysics e-prints

Eggenberger, A., Udry, S., & Mayor, M., Statistical properties of exoplanets. III. Planet properties and stellar multiplicity. 2004, A&A, 417, 353

Ellis, K. M. & Murray, C. D., The Disturbing Function in Solar System Dynamics. 2000, Icarus, 147, 129

Ertel, H., Ein neuer hydrodynamischer Wirbelsatz. 1942, Meteorologische Zeitschrift, 59, 277

Ford, E. B., Havlickova, M., & Rasio, F. A., Dynamical Instabilities in Extrasolar Planetary Systems Containing Two Giant Planets. 2001, Icarus, 150, 303

Gammie, C. F., Layered Accretion in T Tauri Disks. 1996, ApJ, 457, 355

Gladman, B., Dynamics of systems of two close planets. 1993, Icarus, 106, 247

Godon, P. & Livio, M., The Formation and Role of Vortices in Protoplanetary Disks. 2000, ApJ, 537, 396

Godunov, S. K., A Difference Scheme for Numerical Solution of Discontinuos Solution of Hydrodynamic Equations. 1957, Math. Sbornik, 47, 271

Goldreich, P. & Sari, R., Eccentricity Evolution for Planets in Gaseous Disks. 2003, ApJ, 585, 1024

Goldreich, P. & Tremaine, S., The excitation of density waves at the Lindblad and corotation resonances by an external potential. 1979, ApJ, 233, 857 [ADS]

—, Disk-satellite interactions. 1980, ApJ, 241, 425

Hénon, M. & Guyot, M. 1970, in Periodic Orbits Stability and Resonances, in Periodic Orbits Stability and Resonances, 349–+

Hartmann, L., Calvet, N., Gullbring, E., & D'Alessio, P., Accretion and the Evolutiion of T Tauri Disks. 1998, ApJ, 495, 385

Hartmann, L., Hewett, R., & Calvet, N., Magnetospheric accretion models for T Tauri stars. 1: Balmer line profiles without rotation. 1994, ApJ, 426, 669

Hawley, J. F. & Balbus, S. A., A Powerful Local Shear Instability in Weakly Magnetized Disks. II. Nonlinear Evolution. 1991, ApJ, 376, 223

Haynes, P. & Shuckburgh, E., Effective diffusivity as a diagnostic of atmospheric transport. 2000, JGR, 104, 22,777

Holman, M. J. & Murray, N. W., Chaos in High-Order Mean Resonances in the Outer Asteroid Belt. 1996, AJ, 112, 1278

Holman, M. J. & Wiegert, P. A., Long-Term Stability of Planets in Binary Systems. 1999, AJ, 117, 621

Juckes, M. N. & McIntyre, M. E., A high-resolution one-layer model of breaking planetary waves in the stratosphere. 1987, Nature, 328, 590

Juric, M. & Tremaine, S., The Eccentricity Distribution of Extrasolar Planets. 2007, ArXiv Astrophysics e-prints

Klahr, H. H. & Bodenheimer, P., Turbulence in Accretion Disks: Vorticity Generation and Angular Momentum Transport via the Global Baroclinic Instability. 2003, ApJ, 582, 869

Kley, W., On the treatment of the Coriolis force in computational astrophysics. 1998, A&A, 338, L37

—, Mass flow and accretion through gaps in accretion discs. 1999, MNRAS, 303, 696 [ADS]

—, On the migration of a system of protoplanets. 2000, MNRAS, 313, L47

Koller, J., Li, H., & Lin, D. N. C., Vortices in the Co-orbital Region of an Embedded Protoplanet. 2003, ApJL, 596, L91

Kundu, P. K. 1990, Fluid Mechanics, Fluid Mechanics (Academic Press)

Landau, L. D. & Lifshitz, E. M. 1959, Fluid mechanics, Fluid mechanics (Course of theoretical physics, Oxford: Pergamon Press, 1959)

Laney, C. B. 1998, Computational Gasdynamics, Computational Gasdynamics (Cambridge: University Press, 1998)

Lecar, M., Franklin, F., & Murison, M., On predicting long-term orbital instability - A relation between the Lyapunov time and sudden orbital transitions. 1992, AJ, 104, 1230

Levison, H. F., Morbidelli, A., Gomes, R., & Backman, D. 2007, 669–684

Li, H., Colgate, S. A., Wendroff, B., & Liska, R., Rossby Wave Instability of Thin Accretion Disks. III. Nonlinear Simulations. 2001, ApJ, 551, 874

Li, H., Finn, J. M., Lovelace, R. V. E., & Colgate, S. A., Rossby Wave Instability of Thin Accretion Disks. II. Detailed Linear Theory. 2000, ApJ, 533, 1023

Lissauer, J. J., Planet formation. 1993, ARAA, 31, 129 [ADS]

Lithwick, Y., Nonlinear Evolution of Hydrodynamical Shear Flows in Two Dimensions. 2007, ArXiv Astrophysics e-prints

Lovelace, R. V. E., Li, H., Colgate, S. A., & Nelson, A. F., Rossby Wave Instability of Keplerian Accretion Disks. 1999, ApJ, 513, 805

Lynden-Bell, D. & Pringle, J. E., The evolution of viscous discs and the origin of the nebular variables. 1974, MNRAS, 168, 603

Malhotra, R. 1998, in ASP Conf. Ser. 149: Solar System Formation and Evolution, in ASP Conf. Ser. 149: Solar System Formation and Evolution, 37–+

Marcus, P. S., Jupiter's Great Red SPOT and other vortices. 1993, ARAA, 31, 523

Marshall, J., Shuckburgh, E., Jones, H., & Hill, C., Estimates and Implications of Surface Eddy Diffusivity in the Southern Ocean Derived from Tracer Transport. 2006, J. Phys. Ocean., 36, 1806

Marzari, F., Weidenschilling, S. J., Barbieri, M., & Granata, V., Jumping Jupiters in Binary Star Systems. 2005, ApJ, 618, 502

Masset, F., FARGO: A fast eulerian transport algorithm for differentially rotating disks. 2000, A&AS, 141, 165

Masset, F. S. & Ogilvie, G. I., On the Saturation of Corotation Resonances: A Numerical Study. 2004, ApJ, 615, 1000

Mayor, M. & Queloz, D., A Jupiter-Mass Companion to a Solar-Type Star. 1995, Nature, 378, 355

McIntyre, M. E., On the Antarctic ozone hole. 1989, J. Atmos. Terr.l Phys., 51, 29

Murray, C. D., Solar System - Wandering on a Leash. 1992, Nature, 357, 542

Murray, C. D. & Dermott, S. F. 1999, Solar system dynamics, Solar system dynamics (Cambridge University Press)

Murray, N., Hansen, B., Holman, M., & Tremaine, S., Migrating Planets. 1998, Science, 279, 69 [ADS]

Murray, N. & Holman, M., Diffusive chaos in the outer asteroid belt. 1997, AJ, 114, 1246

Musielak, Z. E., Cuntz, M., Marshall, E. A., & Stuit, T. D., Stability of planetary orbits in binary systems. 2005, A&A, 434, 355

Muzerolle, J., Calvet, N., & Hartmann, L., Magnetospheric Accretion Models for the Hydrogen Emission Lines of T Tauri Stars. 1998a, ApJ, 492, 743

Muzerolle, J., Hartmann, L., & Calvet, N., A Brgamma Probe of Disk Accretion in T Tauri Stars and Embedded Young Stellar Objects. 1998b, AJ, 116, 2965

Nakamura, N., Two-dimensional mixing, edge formation, and permeability diagnosed in an area coordinate. 1996, JAS, 53, 1524

Nelson, R. P., Papaloizou, J. C. B., Masset, F., & Kley, W., The migration and growth of protoplanets in protostellar discs. 2000, MNRAS, 318, 18

Nesvorný, D., Alvarellos, J. L. A., Dones, L., & Levison, H. F., Orbital and Collisional Evolution of the Irregular Satellites. 2003, AJ, 126, 398

Ogilvie, G. I. & Lubow, S. H., Saturation of the Corotation Resonance in a Gaseous Disk. 2003, ApJ, 587, 398

—, The effect of planetary migration on the corotation resonance. 2006, MNRAS, 370, 784

Ou, S., Ji, J., Liu, L., & Peng, X., Disk-Planet Interaction Simulations: (I) Baroclinic Generation of Vortensity and Non-Axisymmetric Rossby-Wave-Instability. 2007, ArXiv e-prints, 706

Panetta, R. L., Zonal Jets in wide Baroclinically Unstable Regions: Persistence and Scale Selection. 1993, JAS, 50, 2073

Papaloizou, J. C. B. & Lin, D. N. C., Nonaxisymmetric instabilities in thin self-gravitating rings and disks. 1989, ApJ, 344, 645

—, Theory Of Accretion Disks I: Angular Momentum Transport Processes. 1995, ARAA, 33, 505

Pen, U.-L., A High-Resolution Adaptive Moving Mesh Hydrodynamic Algorithm. 1998, ApJS, 115, 19

Pilat-Lohinger, E. & Dvorak, R., Stability of S-type Orbits in Binaries. 2002, Celestial Mechanics and Dynamical Astronomy, 82, 143

Porco, C. C., West, R. A., McEwen, A., Del Genio, A. D., Ingersoll, A. P., Thomas, P., Squyres, S., Dones, L., Murray, C. D., Johnson, T. V., Burns, J. A., Brahic, A., Neukum, G., Veverka, J., Barbara, J. M., Denk, T., Evans, M., Ferrier, J. J., Geissler, P., Helfenstein, P., Roatsch, T., Throop, H., Tiscareno, M., & Vasavada, A. R., Cassini Imaging of Jupiter's Atmosphere, Satellites, and Rings. 2003, Science, 299, 1541

Pringle, J. E., Accretion discs in astrophysics. 1981, ARAA, 19, 137

Rabl, G. & Dvorak, R., Satellite-type planetary orbits in double stars - A numerical approach. 1988, A&A, 191, 385

Rasio, F. A. & Ford, E. B., Dynamical instabilities and the formation of extrasolar planetary systems. 1996, Science, 274, 954 [ADS]

Rhines, P. B., Waves and turbulence on a beta plane. 1975, J. Fluid. Mech., 69, 1541

Rossby, C. G., Planetary flow patterns in the atmosphere. 1940, J. Roy. Meteor. Soc., 66, 68

Roy, A. E. 1988, Orbital motion, Orbital motion (Bristol, England ; Philadelphia : A. Hilger, 1988. 3rd ed.)

Rypina, I. I., Beron-Vera, F. J., Brown, M. G., Kocak, H., Olascoaga, M. J., & Udovydchenkov, I. A., On the Lagrangian Dynamics of Atmospheric Zonal Jets and the Permeability of the Stratospheric Polar Vortex. 2006, ArXiv Physics e-prints

—, Robust transport barriers resulting from strong Kolmogorov-Arnold-Moser stability. 2007, Phys. Rev. Lett., 98, 104102

Shakura, N. I. & Sunyaev, R. A., Black holes in binary systems. Observational appearance. 1973, A&A, 24, 337 [ADS]

Sheehan, D. P., Davis, S. S., Cuzzi, J. N., & Estberg, G. N., Rossby Wave Propagation and Generation in the Protoplanetary Nebula. 1999, Icarus, 142, 238

Sommeria, J., Meyers, S. D., & Swinney, H. L., Laboratory model of a planetary eastward jet. 1989, Nature, 337, 58

Stone, J. M., Gammie, C. F., Balbus, S. A., & Hawley, J. F., Transport Processes in Protostellar Disks. 2000, Protostars and Planets IV, 589

Strang, G., On the Construction and Comparison of Difference Schemes. 1968, SIAM J. Numer. Anal., 5, 506

Sweby, P. K., High Resolution Schemes Using Flux Limiters for Hyperbolic Conservation Laws. 1984, SIAM J. Numer. Anal., 21, 995

Szebehely, V., Stability of artificial and natural satellites. 1978, Celestial Mechanics, 18, 383

Tanaka, H., Takeuchi, T., & Ward, W. R., Three-Dimensional Interaction between a Planet and an Isothermal Gaseous Disk. I. Corotation and Lindblad Torques and Planet Migration. 2002, ApJ, 565, 1257

Touma, J. & Wisdom, J., Resonances in the Early Evolution of the Earth-Moon System. 1998, AJ, 115, 1653

Trac, H. & Pen, U.-L., A Primer on Eulerian Computational Fluid Dynamics for Astrophysics. 2003, PASP, 115, 303

Ward, W. R., Density waves in the solar nebula - Differential Lindblad torque. 1986, Icarus, 67, 164

—, Protoplanet Migration by Nebula Tides. 1996, Icarus, 126, 261

—, Survival of Planetary Systems. 1997, ApJL, 482, L211+

Weidenschilling, S. J. & Marzari, F., Gravitational scattering as a possible origin for giant planets at small stellar distances. 1996, Nature, 384, 619 [ADS]

White, R. J. & Hillenbrand, L. A., On the Evolutionary Status of Class I Stars and Herbig-Haro Energy Sources in Taurus-Auriga. 2004, ApJ, 616, 998

Williams, G. P., Planetary Circulations: 1. Barotropic representation of the Jovian and Terrestrial Turbulence. 1978, JAS, 35, 1399

Wisdom, J., The resonance overlap criterion and the onset of stochastic behavior in the restricted three-body problem. 1980, AJ, 85, 1122