



On the dynamics of dust, magnetohydrodynamics of disks and atmospheric radiation of planets

Gopakumar Mohandas

Supervisors:

Martín E. Pessah and Tobias Heinemann

This thesis has been submitted to the PhD School of the Faculty of Science,
University of Copenhagen

On the dynamics of dust, magnetohydrodynamics of disks and atmospheric radiation of planets

PhD thesis in Theoretical Astrophysics

Gopakumar Mohandas

Supervisors:

Martín E. Pessah and Tobias Heinemann

Niels Bohr International Academy

Niels Bohr Institute

Faculty of Science

University of Copenhagen

Submitted June 6, 2018.

Abstract

This thesis is an anthology of three theoretical problems in astrophysics with applications to planetary atmospheres, their immediate external environment, and their birthplaces.

The first problem concerns the linear and nonlinear stability of equilibrium motions. In particular, we investigate the stability of a charged particle in a circular orbit subject to axisymmetric gravitational and electromagnetic forces. We extend previous work on this problem by including a toroidal magnetic field. In source free regions, we prove that the toroidal field has no effect on either the equilibrium orbit or its stability. However, in regions with charge or current sources, we find that the toroidal field can potentially alter the stability of the orbit although it plays no role in determining the equilibrium orbital coordinates. We show that the toroidal field enters the system as a gyroscopic force and can stabilize otherwise unstable particle orbits for a range of physical parameters. We also demonstrate that gyroscopic stability so attained is only temporary and that the slightest dissipative forces can render the system unstable again, albeit at a slower rate. Our results may apply to dust grains orbiting within a rotating planetary magnetosphere.

The second problem looks at how magnetic diffusion alters the character of the magnetorotational instability. The magnetorotational instability (MRI) is generally regarded as the foremost contender for driving magnetohydrodynamic turbulence in differentially rotating astrophysical disks thereby facilitating accretion. In disks that are poorly ionized, the three non-ideal effects of ohmic, Hall and

ambipolar diffusion take hold. We conduct a systematic analysis of the non-ideal MRI in the shearing sheet framework and elucidate the character of the eigenmodes. We derive expressions for the kinetic and magnetic stresses and uncover a new characteristic scale when the net magnetic field and angular velocity are anti-parallel. This scale may possibly signal a change in the nature of the ensuing turbulence provided dissipative effects are small. Non-ideal effects pervade disks around young stars and our results may be relevant to the dynamical evolution of certain parts of such disks.

The third problem deals with modeling irradiated atmospheres. Using the principles and methods of radiative transfer, we derive a plane-parallel equilibrium analytical model of an atmosphere that receives radiant energy from above and below. Constructing analytical atmospheric models is a challenging exercise and obtaining one with a frequency dependent opacity function is one of the main points of difficulty. By using the picket-fence technique for modeling spectral lines, we are able to derive exact analytical solutions and thereby obtain thermal profiles of an atmosphere that receives strong collimated high frequency radiation from above in addition to thermally emitted radiant energy from below. Our model also includes the effects of coherent scattering in the lines and the continuum. An obvious application of our analysis is to modeling planetary atmospheres, in particular, those outside of our own solar system that are presently being discovered in the thousands.

Resume

Denne afhandling er en antologi af tre teoretiske problemer inden for astrofysik, med anvendelse inden for planetatmosfærer, deres nærmeste omgivelser og de steder de dannes.

Det første problem omhandler den lineære og ikke-lineære stabilitet af ligevægtsbevægelser. Især undersøger vi stabiliteten af en ladet partikel, der befinner sig i en cirkulær bane og er utsat for rotationssymmetriske gravitationelle og elektromagnetiske kræfter. Vi generaliserer tidligere studier ved at inkludere et toroidalformet magnetfelt. I kildefri områder beviser vi, at det toroidale felt ikke påvirker hverken ligevægtsbanen eller dens stabilitet. Til gengæld konkluderer vi, at det toroidale felt potentielt kan påvirke stabiliteten af banen i områder med kilder af ladning eller strøm. Dette selvom disse kilder ikke spiller en rolle for bestemmelsen af ligevægtsbanekoordinaterne. Vi påviser, at det toroidale felt indgår i systemet som en gyroskopisk kraft, som kan stabilisere ellers ustabile partikelbaner for en række af fysiske parametre. Vi påpeger også, at den opnåede gyroskopiske stabilitet kun er midlertidig, og at den mindste smule dissipativ kraft kan gøre systemet ustabilt påny, omend med en lavere vækstrate. Vores resultater kan måske anvendes til forstå støvparkler, der er i omløb inde i en roterende planetmagnetosfære.

Det andet problem er et studie af hvordan magnetisk diffusion ændrer karakteren af den magnetorotationelle instabilitet (MRI). Den magnetorotationelle instabilitet betragtes generelt som en sandsynlig drivkraft af magnetohydrodynamisk turbulens i differentielt roterende massetilvækstskiver. I massetilvækstskiver, der er svagt ioniserede, er de tre ikke-ideelle effekter, Ohmisk, Hall og ambipolar diffu-

sion, til stede. Vi udfører en systematisk analyse af den ikke-ideelle MRI i “shearing sheet” frameworkt og belyser karakteren af egenfunktionerne. Vi udleder ligninger for den kinetiske og magnetiske spænding og afslører en ny karakterisk længdeskala, hvor det totale magnetfelt og angulære hastighed er anti-parallelle. Denne skala kan måske indikere en ændring i karakteren af den følgende turbulens såfremt dissipative effekter er små. Ikke-ideelle effekter er tilstede i massetilvækstskiver omkring unge stjerner og vores resultater kan potentielt være relevante for den dynamiske udvikling af visse dele af sådanne skiver.

Det tredje problem angår modellering af bestrålede atmosfærer. Ved hjælp af principper og metoder for strålingstransport, udleder vi en plan-parallel analytisk ligevægtsmodel af en atmosfære, der belyses både fra neden og fra oven. Konstruktionen af en analytisk atmosfæremodel er en udfordrende opgave, især hvis gennemsigtighedsfunktionen er frekvensafhængig. Ved hjælp af picket-fence teknikken til at modellere spektrallinjer, er vi dog i stand til at finde frem til eksakte, analytiske løsninger, og derved termiske profiler for en atmosfære, der modtager kraftig, højfrekvent, kollimeret stråling ovenfra og termisk udstråling nedefra. Vores model inkluderer også effekter fra kohærent spredning i linjer og kontinuum. En oplagt anvendelse af vores analyse er i modelleringen af planetatmosfærer, i særdeleshed dem uden for vores eget solsystem, der i disse tider bliver opdaget i tusindvis.

Acknowledgements

Gratitude is one of the least articulate of the emotions, especially when it is deep.

Felix Frankfurter

Let me try. I wish to thank:

- Martin Pessah, for his sage advice and supervision and especially for helping me take my mind off finances.
- Tobias Heinemann, for his insistence on rigor and clarity as well as teaching me some good principles every scientifically minded person must abide by.
- Omer Blaes, for his wisdom and advice, and for being a great example of one who is both a good scientist and a good human being.
- Kevin Heng, for introducing me to exoplanets and the chance to collaborate.
- Yan-Fei Jiang, Lars Bildsten and Lisa Stewart, for their help and hospitality during my time at KITP as a graduate fellow.
- Oliver Gressel, Gareth Murphy and Christian Brinch, for the interesting lunch conversations and for the annual DHL relay.
- Farrukh Nauman, Yuri Fujii and Pablo Benitez-Llambay, for the interesting discussions, chit-chat in and out of the office and the fun times abroad.
- Leonardo Krapp and Philipp Weber, for being great graduate student peers and especially for the basketball sessions during the spring.

-
- Andrew Jackson, for the many coffee time conversations and the interesting historical snippets.
 - Amel Durakovic, for the native-foreigners perspective into life in Denmark and many interesting discussions on physics.
 - Lisbeth Dilling and Kader Ahmad, for their relentless help and support throughout my studies and for being the nicest librarians I have ever known.
 - Anette Studsgård and Ulla Dahl Lindberg, for all the help with working at the NBIA and for making life as a PhD student a whole lot easier.
 - the faculty and staff at NBI, Starplan and CSH Bern, for hosting my research studies and hospitality during my time there.
 - Jytte & Jørgen Hansen, for welcoming me into their home and letting me live under their roof for the last couple of years, for treating me like family and letting me be part of the celebrations.
 - Lise Smidt and family, for their warmth and affection, and for the many meals and hyggeligt occasions together.
 - Rajan Ambat and family, for being home away from home and for being my first point of contact to Denmark.
 - Salil Joshy and Sumesh Sukumara, for the long time friendship and for the chance to converse in my native tongue away from home.
 - Thomas Berlok, for things far too numerous to count - not to mention for helping with this thesis. I could not have hoped for a better office mate, dear friend and guide during my time here.
 - my parents and my brother, for their unconditional, undying love and support. If I have been neglectful these last few years, which I no doubt have, I vow to make amends.

Publications

The following publications are part of this thesis and are listed below in the order that they are referred to and appear herein.

- Gopakumar Mohandas, Tobias Heinemann and Martin E. Pessah (2018). *Orbital Stability in Static Axisymmetric Fields*, submitted to Celestial Mechanics and Dynamical Astronomy.
- Gopakumar Mohandas and Martin E. Pessah (2017). *Spectral Analysis of Non-ideal MRI Modes: The Effect of Hall Diffusion* The Astrophysical Journal, 838, 1.
- Gopakumar Mohandas, Martin E. Pessah and Kevin Heng (2018). *Analytical Models of Exoplanetary Atmospheres. V. Non-gray Thermal Structure with Coherent Scattering* The Astrophysical Journal, 858, 1.

Outline

The outline of the thesis is as follows. The thesis is divided into three parts. Each part consists of two chapters. The first chapter of every part provides a minimum theoretical background that sets the stage for the analysis we have carried out in the respective papers. The second chapter of every part touches upon relevant applications, previous related work and a summary discussion of our main results and future prospects. Lastly, we append all three papers in the order in which they are referenced towards the end of the thesis.

Contents

I Stability of Circular Orbits	1
1 Dynamical Systems and Stability Theory	3
1.1 Dynamical Systems	3
1.2 Motion Stability	4
1.3 Lyapunov Stability	5
1.4 Lyapunov's Direct Method	7
1.5 Lyapunov's Stability Theorem	8
1.6 Linear and Spectral Stability	9
1.7 Lagrangian and Hamiltonian Mechanics	11
1.8 Stability of Conservative Systems	13
1.9 Cyclic coordinates and Routh's Theorem	14
1.10 Stability of Stationary Motions	16
1.11 Gyroscopic Stabilization	17
1.12 Dissipation Induced Instability	19
2 Charged Particle Dynamics and Stability	21
2.1 Charged Particle Dynamics	21
2.2 Circular Orbits in a Rotating Magnetosphere	23
2.3 Charged Dust in Planetary Rings	26
2.4 Paper I - Summary & Prospects	27

Contents

II Non-ideal MRI in Differentially Rotating Disks	31
3 Magnetohydrodynamics	33
3.1 Kinetics	33
3.2 Klimontovich Equation	35
3.3 Boltzmann Equation	36
3.4 MHD Equations	38
3.5 Ideal MHD	42
3.6 MHD Waves	43
3.7 Shearing Sheet Equations	45
3.8 Magnetorotational Instability	48
4 Non-ideal MRI in Accretion Disks	51
4.1 Angular Momentum Transport in Accretion Disks	51
4.2 The Eigenvalue Problem for Ideal MRI	54
4.3 Non-ideal MHD in Protoplanetary Disks	60
4.4 Nonlinear evolution of the non-ideal MRI	63
4.5 Paper II - Summary & Prospects	66
III Thermal Structure of Irradiated Atmospheres	69
5 Radiative Transfer	71
5.1 The Radiative Transfer Equation	71
5.2 Extinction: Absorption and Scattering	73
5.3 Emission: Thermal Radiation and Scattering	74
5.4 Moments of the Transfer Equation	76
5.5 The Flux	77
5.6 Model Atmospheres	79
5.7 Radiative Equilibrium	80
5.8 Optical Depth and Column Mass	81
5.9 Gray Approximation and Mean Opacities	81

5.10	Gray Atmosphere Model	83
6	Analytical Models of Irradiated Atmospheres	87
6.1	Exoplanets	87
6.2	Exoplanetary Atmospheres	88
6.3	Gray Atmosphere Model for Irradiated Atmospheres	91
6.4	Numerical Modeling of Irradiated Atmospheres	93
6.5	Paper III - Summary & Prospects	95

Contents

List of Figures

1.1	A geometric representation of Lyapunov's definition of stability. The curve traces the image point defined by the deviations as it evolves in time.	6
1.2	A geometric representation of Lyapunov's stability theorem, Theorem 1.	8
2.1	Stability diagram for equilibrium equatorial orbits. Here, the regions highlighted in the darker shade of gray correspond to potential minima and are therefore stable. The regions highlighted in light gray correspond to potential maxima including saddle points and are therefore unstable.	25
2.2	Stability diagram for equilibrium halo orbits. Here, the regions highlighted in the darker shade of gray correspond to potential minima and are therefore stable. The regions highlighted in light gray correspond to potential maxima are therefore unstable. . . .	26
2.3	A side by side comparison of the ring systems of the solar system planets, scaled to a common planetary equatorial radius, including some of the moons. Figure reprinted from Burns, Hamilton, and Showalter (2001), with permission from Springer Nature.	28
3.1	Shearing sheet schematic	47

List of Figures

4.1	The eigenvalues of the operator \mathbf{L} as a function of wavenumber k_n . The real part (in blue) of the eigenmode γ is the growth rate of the ideal magnetorotational instability.	56
4.2	The normalized per- k mean kinetic and magnetic stresses and energy densities as a function of wavenumber k_n for the ideal MRI eigenmode.	58
4.3	Recent observational image of a dusty protoplanetary disk around the young star IM Lupi as seen in reflected light captured by the SPHERE instrument on ESO’s Very Large Telescope (VLT). Figure Credit ESO/H. Avenhaus et al./DARTT-S collaboration. See also Avenhaus et al. (2018).	61
4.4	A map highlighting the dominant non-ideal effect in a parameter space given by the gas number density n and the magnetic field strength B as derived in Armitage (2015). The boundary between the Hall and ohmic regimes were derived assuming a disk temperature $T = 1000K$. The dashed line tracks the parameter space as one moves from the inner disk at 0.1 AU to the outer disk at 100 AU.	62
5.1	A beam of radiation impinging through an infinitesimal surface contained within an infinitesimal solid angle.	72
6.1	A cartoon representation of the two main methods of detecting and observing exoplanets. Figure reprinted from Hecht (2016), with permission from Springer Nature.	89
6.2	The temperature profile computed for a non-irradiated gray atmosphere (blue curve) and an irradiated dual-band gray atmosphere (black curves) for two different strengths of shortwave absorption opacity.	93

6.3 The picket fence model. A sequence of uniformly wide and uniformly spaced steps in opacity, measured in terms of the parameter β , across frequency represent the spectral lines on top of the continuum.	96
---	----

List of Figures

Part I

Stability of Circular Orbits

Chapter 1

Dynamical Systems and Stability Theory

In this chapter, we provide a brief introduction to the stability theory of dynamical systems. The purpose of this chapter is to provide a sufficient theoretical background for the analysis conducted in Mohandas, Heinemann, and Pessah (2018) which we shall refer to hereafter as Paper I. We describe the basic mathematical framework for analyzing the dynamics and stability of mechanical systems. We outline the main notions and theorems of stability that are relevant to our work. The account presented here is largely based on the excellent monograph on the subject by Merkin (2012) as well as parts of Bloch et al. (1994) and Krechetnikov and Marsden (2007). The presentation is necessarily minimal and we refer the interested reader to the above cited references for a detailed discussion of the fundamental concepts.

1.1 Dynamical Systems

The dynamics of several different kinds of mechanical systems may be described by the general first order ordinary differential equation

$$\frac{d\vec{y}}{dt} = \mathbf{Y}(\vec{y}, t; \lambda_i), \quad (1.1)$$

1. Dynamical Systems and Stability Theory

where $\vec{y} = \{y_1, y_2, \dots, y_n\}$ represent the variables characterizing the state of the system and are usually comprised of position coordinates and velocities or momenta. The time evolution of the variables given by the vector \vec{y} are governed by the n number of generally nonlinear functions denoted by \mathbf{Y} . Additionally, λ_i denotes a set of control parameters that may be present and could alter the specific nature of the dynamics depending on the constant (in time) values they assume. The system is called autonomous if none of the functions \mathbf{Y} depend on time explicitly and is called non-autonomous otherwise. Hamiltonian systems fall into the general class of systems that can be described by Equation 1.1. Given the initial conditions and with the knowledge of \mathbf{Y} , one can solve Equation 1.1, sometimes analytically but more often numerically, to determine the trajectory or orbit followed by the system in time. The long time evolution of most dynamical systems typically fall into one of three categories. They either evolve into *i*) a steady state equilibrium, *ii*) an oscillatory state, or *iii*) an irregular chaotic state.

1.2 Motion Stability

Very often, one is interested in the stability of the motion of dynamical systems to small perturbations. We refer to the solutions of Equation 1.1 given by (Greiner, 2009; Merkin, 2012)

$$y_1 = f_1(t), \dots, y_n = f_n(t), \quad (1.2)$$

whose stability we wish to investigate as the unperturbed motion. At some initial time $t = t_0$, these solutions satisfy

$$y_1 = f_1(t_0), \dots, y_n = f_n(t_0), \quad (1.3)$$

We consider small (relative to the unperturbed solution) perturbations to the unperturbed motion at t_0 such that

$$y_1 = f_1(t_0) + \varepsilon_1, \dots, y_n = f_n(t_0) + \varepsilon_n, \quad (1.4)$$

where $\varepsilon_1, \dots, \varepsilon_n$ are the perturbations and the solutions given by Equation 1.4 are referred to as the perturbed motion. The difference between the perturbed and

the unperturbed solutions are called deviations or variations and are given by

$$x_i = y_i(t) - f_i(t), \quad i = 1, \dots, n. \quad (1.5)$$

Whereas, $x_i(t_0) = \varepsilon_i$, their subsequent evolution is determined by Equation 1.1. Whether the system is stable or unstable depends on whether the deviations x_i grow or dampen as a function of time.

Let us consider the equations for the perturbed motion. Substituting Equation 1.5 into Equation 1.1, we obtain

$$\frac{df_i}{dt} + \frac{dx_i}{dt} = Y_i(f_1 + x_1, \dots, f_n + x_n). \quad (1.6)$$

Since the deviations are small by definition, we expand the right hand side in a power series and subtract the equation of unperturbed motion to get

$$\frac{dx_i}{dt} = a_{i1}x_1 + \dots + a_{in}x_n + X_i, \quad i = 1, \dots, n, \quad (1.7)$$

where

$$a_{ij} = \left(\frac{\partial Y_i}{\partial x_j} \right)_{x=0}. \quad (1.8)$$

The term X_i denotes all the terms involving powers of x_i greater than unity. If we omit this nonlinear term, we obtain what are called the equations of first approximation or the linearized system of equations.

1.3 Lyapunov Stability

The notion of stability was made precise by Lyapunov who proposed the following definition.

Definition 1 (Merkin, 2012). *If for any positive value ε , however small, one can find a positive value δ such that at $t = t_0$, for all perturbations $x_i(t_0)$ satisfying*

$$\sum_{i=1}^n x_i(t_0)^2 \leq \delta \quad (1.9)$$

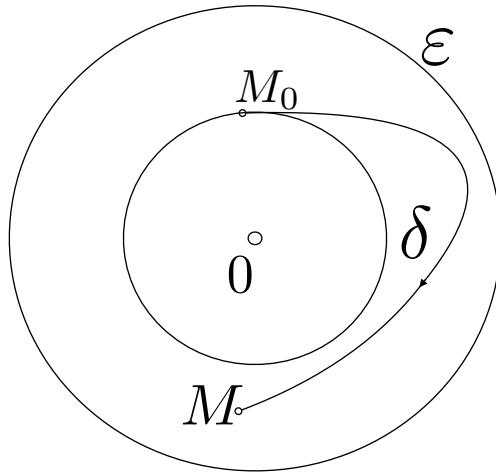


Figure 1.1: A geometric representation of Lyapunov's definition of stability. The curve traces the image point defined by the deviations as it evolves in time.

the inequality

$$\sum_{i=1}^n x_i(t)^2 < \varepsilon \quad (1.10)$$

holds true, then the unperturbed motion is stable; otherwise it is unstable.

A visual aid is very helpful in understanding the above definition and is provided in Figure 1.1. Consider a space of n variables where each “coordinate” denotes the deviations to the unperturbed motion. Therefore, the coordinate origin $x_i = 0$ represents the unperturbed motion. The time evolution of the deviations may be represented in terms of a “distance” from the origin given by the scalar

$$\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}, \quad (1.11)$$

to an image point M located by the coordinates x_i . When perturbed at the initial time t_0 , the image point is M_0 . As the system evolves, the image point traces a trajectory in the n dimensional space given by the solution to Equation 1.7. Now consider a sphere of radius $\sqrt{\varepsilon}$ in this space. If the system is stable in the Lyapunov

sense, then one can always find a smaller sphere of radius $\sqrt{\delta}$ such that starting at any image point M_0 on or within the δ sphere, the image point follows a trajectory that never crosses the ε sphere. If the system is unstable, then regardless of how small the δ sphere is, or in other words, however close the initial image point M_0 was to the origin, the image point at a later time will eventually cross the ε sphere.

If the deviations have the additional characteristic that at late times, they converge to the unperturbed motion, i.e.,

$$\lim_{t \rightarrow \infty} \sum x_i^2(t) = 0, \quad (1.12)$$

then we say that the unperturbed motion is asymptotically stable. Note however, for the system to be asymptotically stable, it must also be stable according to Lyapunov's definition in addition to satisfying Equation 1.12.

There are three restrictions on Lyapunov's definition of stability. Firstly, we require that only the initial conditions are perturbed and that the perturbations evolve under the same forces as the original unperturbed system. Secondly, the time interval over which stability is considered is arbitrarily large and in principle infinite. Finally, the perturbations are considered to be arbitrarily small.

1.4 Lyapunov's Direct Method

A common procedure with which one studies whether a system is stable in the Lyapunov's sense is the Lyapunov Direct Method or Second Lyapunov Method. The essential component to this method is the so-called Lyapunov function. We assume that the Lyapunov function $V(\vec{x})$ is real, single-valued and continuous in the domain $\sum x_i^2 \leq \mu$ where μ is a constant and that also satisfies $V(0) = 0$.

A Lyapunov function is definite if it has only one sign, either positive or negative and vanishes only at the origin. A Lyapunov function is semi-definite if it has only one sign, but may vanish at points other than the origin. Finally, a Lyapunov function is indefinite if it changes sign in the domain. A sufficient condition for positive definiteness of the Lyapunov function is given by Sylvester's

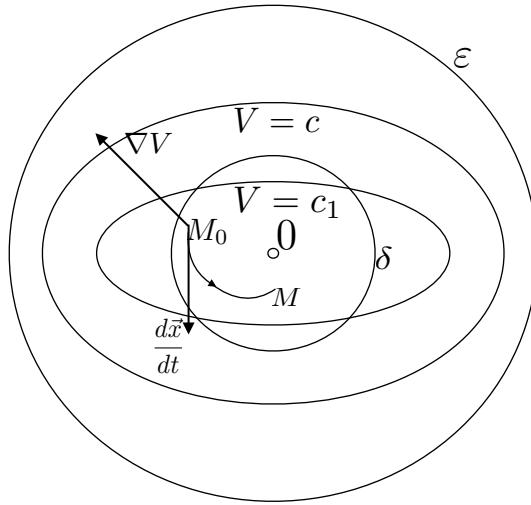


Figure 1.2: A geometric representation of Lyapunov’s stability theorem, Theorem 1.

criterion which states that the function is positive definite if all the principal diagonal minors of its coefficient matrix are positive, where the coefficient matrix C_{jk} is defined such that

$$C_{jk} = \frac{\partial^2 V}{\partial x_j \partial x_k}, \quad (1.13)$$

where $j = 1, \dots, n$ and $k = 1, \dots, n$. Negative definiteness is similarly ascertained by checking whether $-V$ is positive definite instead.

1.5 Lyapunov’s Stability Theorem

The conditions under which an unperturbed motion is stable to deviations and whose evolution is governed by Equation 1.7 were laid out by Lyapunov in his fundamental stability theorem of motion stability. We reproduce the theorem as stated in Merkin (2012) below

Theorem 1. *If for differential equations of a perturbed motion we can find a definite function V such that by virtue of the given equations its derivative \dot{V} is either identically*

equal to zero or is semidefinite with the opposite sign of V , then the unperturbed motion is stable.

We furnish a proof based on simple geometric considerations below. Consider a sphere of radius $\sqrt{\varepsilon}$ in the n dimensional deviations space as described previously and a Lyapunov definite (assumed positive without loss of generality) function to be defined in this space. Next, consider a surface with constant value of $V = c$ inside the ε sphere. Now, construct another sphere of arbitrarily small radius $\sqrt{\delta}$ that is fully contained within the surface $V = c$. Let the initial deviations to the unperturbed motion be such that the initial image point M_0 is located on or inside the δ sphere. Consider also an isosurface of the Lyapunov function $V = c_1$ coincident with the initial image point M_0 that is also contained within $V = c$. According to Theorem 1, the time derivative of the Lyapunov function

$$\dot{V} = \dot{x}_1 \frac{\partial V}{\partial x_1} + \dot{x}_2 \frac{\partial V}{\partial x_2} + \cdots + \dot{x}_n \frac{\partial V}{\partial x_n} = \frac{d\vec{x}}{dt} \cdot \nabla V \leq 0, \quad (1.14)$$

Since ∇V points outward for a positive definite function, $\dot{V} \leq 0$ only if the velocity $d\vec{x}/dt$ is at right angles or more to the constant surface $V = c_1$ containing the image point M_0 . This means that the trajectory of the image point moves further inward and never leaves $V = c_1$. Therefore, the image point never crosses δ and by Lyapunov's definition, the system is stable. Figure 1.2 provides a geometrical illustration of Lyapunov's direct method.

1.6 Linear and Spectral Stability

In many situations, one may be able to determine the stability of the system by examining the equations of first approximation alone. This is relatively easier to do and is often the first resort to addressing the question of stability. We shall list the relevant theorems that outline the method of ascertaining stability or its lack thereof and the conditions under which it is valid.

For an autonomous system, the coefficient a_{ij} in the equations of first approximation do not depend on time. Therefore, one may seek solutions of the form

1. Dynamical Systems and Stability Theory

$x_i = A_i \exp(\sigma t)$ where the A_i 's are constants. Substituting these in Equation 1.7 yields the characteristic equation which is a polynomial of degree n whose roots are σ . This method basically amounts to solving an eigenvalue problem and the determination of the structure or spectrum of eigenvalues. By examining the nature of the roots, one can address the question of stability according to two theorems due to Lyapunov as stated in Merkin (2012) which we reproduce below without proof.

Theorem 2. *If all roots of the characteristic equation of a first approximation have negative real parts, the unperturbed motion is asymptotically stable irrespective of the nonlinear terms.*

Theorem 3. *If at least one of the roots of the characteristic equation has a positive real part, then the unperturbed motion is unstable irrespective of the nonlinear terms.*

If some of the roots do not have a real part, the question of stability of the linearized system requires further examination. Systems that are determined to be stable or unstable on the basis of the spectrum of the eigenvalues alone are referred to as being spectrally stable or unstable.

Some matrix algebra machinery is necessary in order to ascertain linear stability if the characteristic polynomial has roots with zero real part. We briefly outline the key elements and steps involved in the procedure.

1. We first determine the greatest common divisor of all the minor matrices of the characteristic polynomial at all orders. The greatest common divisor at a given order k is denoted as $D_k(\sigma)$ for $k = 1, \dots, n$.
2. We then determine the invariant factors by considering the ratio of successive greatest common divisors. That is, the invariant factor at a given order k is $E_k(\sigma) = D_k/D_{k-1}$ with $D_0 = 1$.
3. We then factorize the invariant factors into their respective elementary divisors as

$$E_k = (\sigma - \sigma_1)^{e_{k1}} (\sigma - \sigma_1)^{e_{k2}} \dots (\sigma - \sigma_p)^{e_{kp}}. \quad (1.15)$$

4. The roots of the elementary divisors coincide with the roots of the characteristic polynomial but with possibly different multiplicities e_{ki} . The elementary divisor is a primary divisor if $e_{ki} = 1$.

The question of linear stability is then addressed by the following theorem as stated in Merkin (2012) which we reproduce below without proof.

Theorem 4. *If some of the roots of the characteristic equation have a zero real part and the rest of the roots all have negative real parts, then i) the unperturbed motion is stable, although not asymptotically, if all the roots with zero real part correspond to simple primary divisors, i.e., $e_{ki} = 1$, ii) the unperturbed motion is unstable if even one of the roots with zero real part is a multiple root of the corresponding primary divisor i.e., $e_{ki} > 1$.*

This extended examination of stability is only necessary if there are multiple roots with zero real part and the remainder of the roots have negative real part.

1.7 Lagrangian and Hamiltonian Mechanics

We adopt the Lagrangian or Hamiltonian description which provides an equivalent and elegant formulation of classical mechanics. We present a brief summary of the framework below.

Using d'Alembert's principle of virtual work, one derives the fundamental equation (Greiner, 2009)

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_i} - \frac{\partial T}{\partial q_i} = Q_i, \quad (1.16)$$

where T is the kinetic energy, Q_i are the generalized forces and q_i are the generalized coordinates where $i = 1, \dots, n$. Consider the case where the generalized force is derived from a scalar potential as

$$Q_i = \frac{\partial V}{\partial q_i}. \quad (1.17)$$

Such forces are called conservative forces and are an important, but not the only, type of forces that one encounters. Equation 1.16 is often expressed in terms of

1. Dynamical Systems and Stability Theory

the Lagrangian, $L = T - V$ which becomes

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0, \quad (1.18)$$

By applying a Legendre transformation, one can define a new quantity called the Hamiltonian as

$$H = \sum_{i=1}^n p_i \dot{q}_i - L, \quad (1.19)$$

where we also define the canonically conjugate momentum to q_i as

$$p_i = \frac{\partial L}{\partial \dot{q}_i}, \quad (1.20)$$

With (q_i, p_i) as the independent coordinates, the dynamics is governed by Hamilton's equations

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}. \quad (1.21)$$

When studying a larger classes of mechanical systems that includes forces other than conservative forces, it is helpful to decompose Equation 1.16 as

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_i} - \frac{\partial T}{\partial q_i} = Q_i^l + Q_i^{nl}, \quad (1.22)$$

$$\frac{dq_i}{dt} = \dot{q}_i, \quad k = 1, \dots, n \quad (1.23)$$

where Q_i^l and Q_i^{nl} are forces linear and nonlinear in the coordinates q_i and velocities \dot{q}_i respectively. The kinetic energy is a positive definite function of the velocities as given by

$$T = \frac{1}{2} a_{jk} \dot{q}_j \dot{q}_k. \quad (1.24)$$

The linear force terms may be further categorized as follows.

$$Q_i^l = -C^l q_i - B^l \dot{q}_i, \quad (1.25)$$

where C^l and B^l are $n \times n$ square matrices which we can further decompose into symmetric and skew-symmetric components as

$$c_{jk} = \frac{C_{jk}^l + C_{kj}^l}{2} = c_{kj}, \quad b_{jk} = \frac{B_{jk}^l + B_{kj}^l}{2} = b_{kj}, \quad (1.26)$$

$$p_{jk} = \frac{C_{jk}^l - C_{kj}^l}{2} = -p_{kj}, \quad g_{jk} = \frac{B_{jk}^l - B_{kj}^l}{2} = -g_{kj}. \quad (1.27)$$

With this decomposition, the general equation of motion may be expressed as (Merkin, 2012)

$$a_{ij}\ddot{q}_j + \underbrace{c_{ij}q_j}_{\text{conservative}} + \underbrace{p_{ij}q_j}_{\text{non-conservative}} + \underbrace{b_{ij}\dot{q}_j}_{\text{dissipative}} + \underbrace{g_{ij}\dot{q}_j}_{\text{gyroscopic}} = Q_i^{nl}. \quad (1.28)$$

The coefficient matrix c_{ij} represent conservative forces for which the work done is independent of the path taken. The coefficient matrix p_{ij} on the other hand represent forces for which the work done does depend on the path taken. Dissipative forces denoted by the coefficient matrix b_{ij} represents dissipative influences such as friction as the name suggests. Finally, gyroscopic forces denoted by the coefficient matrix g_{ij} represent forces that do not do any work. We refer the reader to Merkin (2012) for a more elaborate definition of the force types.

1.8 Stability of Conservative Systems

While Lyapunov's direct method offers a straightforward route to demonstrating stability, the challenge often is in constructing a suitable positive or negative definite Lyapunov function. The consideration of stability becomes greatly simplified when dealing with equilibrium states of conservative systems, that is, systems where the total energy is a conserved quantity. The potential energy function then presents itself as a suitable Lyapunov function for the investigation of stability. These are systems for which only the conservative force term in Equation 1.28 is present. A theorem due to Lagrange and rigorously proved by Dirichlet states the condition under which purely conservative systems are stable in the Lyapunov sense. We reproduce below the so-called Lagrange-Dirichlet theorem as stated in Merkin (2012).

Theorem 5. *If at the position of an isolated equilibrium of a conservative system with holonomic and scleronomous constraints¹, the potential energy has a minimum, then*

¹In many systems, the motion can be constrained where the constraining conditions may be expressed as $f_k(q_1, q_2, \dots, t) = 0$, $k = 1, \dots, s$. Such systems are called holonomic. If time does not appear explicitly in the constraint condition, the constraint is called scleronomous.

1. Dynamical Systems and Stability Theory

the equilibrium position is stable.

We sketch a quick, though not rigorous, proof of the theorem below. At the equilibrium position, the potential energy is equal to zero and is also a minimum. We may expand the potential energy function in a power series as

$$V = V(0) + \left(\frac{\partial V}{\partial q_j} \right)_0 q_j + \frac{1}{2} \left(\frac{\partial^2 V}{\partial q_j \partial q_k} \right)_0 q_j q_k + \dots \quad (1.29)$$

The first and second term vanishes at the equilibrium position. The third term is in fact the coefficient matrix as defined in Equation 1.13 which is positive at the equilibrium position since the potential is a minimum here. Therefore, according to Sylvester's criterion, the potential energy is positive definite and since it is independent of time, $\dot{V} = 0$. According to Theorem 1, the motion is stable at the minimum.

The Lagrange-Dirichlet theorem only provides sufficient conditions for stability. However, one could ask whether the inverse holds; that is, is the equilibrium unstable when it is not an isolated minimum. Chetaev proved the following generalized theorem which shows that the system is unstable when not at an isolated minimum. We state this theorem below without proof.

Theorem 6 (Merkin, 2012). *If at a position of an isolated equilibrium the potential energy V , which is an analytical function, has no minimum, then the equilibrium is unstable.*

1.9 Cyclic coordinates and Routh's Theorem

In a large class of mechanical systems, one or more coordinates may be absent in the expression for the kinetic energy and the generalized forces. Such coordinates are referred to as cyclic coordinates. Cyclic coordinates are connected to some fundamental symmetry property of the system as revealed by Noether's theorem (Goldstein, Poole, and Safko, 2014). A simple example are systems with axial symmetry. If cylindrical coordinates are used to describe the system, one finds that

the azimuthal angular coordinate is conspicuously absent from the Lagrangian and Hamiltonian. In this case, the equation of motion given by Equation 1.16 reduces to

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\varphi}_j} = 0, \quad (1.30)$$

which results in the integral of motion

$$p_{\varphi_j} = c_j = \text{const.} \quad j = 1, \dots, m, \quad (1.31)$$

where m is the number of cyclic coordinates. It would then be reasonable to work in a reduced phase space comprising only of the non-cyclic coordinates and corresponding velocities. This is easily achieved by defining the Routhian function as

$$R = L - \sum_{k=1}^m c_k \dot{\varphi}_k. \quad (1.32)$$

where R is a function of only the positional coordinates and the corresponding velocities and where we replace the $\dot{\varphi}_k$, in terms of q_k and \dot{q}_k . This yields the new equations of motion in the reduced space

$$\frac{d}{dt} \frac{\partial R}{\partial \dot{q}_j} - \frac{\partial R}{\partial q_j} = \tilde{Q}_j, \quad j = 1, \dots, s \quad (1.33)$$

where V is the potential energy associated with conservative forces and \tilde{Q}_j are the other kinds of forces present in the system. The Routhian function naturally splits into three components in powers of \dot{q}_j as (Rumiantsev, 1966)

$$R = R_2 + R_1 + R_0 - V, \quad (1.34)$$

where

$$R_2 = \frac{1}{2} a_{kj} \dot{q}_j \dot{q}_k, \quad (1.35)$$

$$R_1 = a_j \dot{q}_j, \quad (1.36)$$

and $R_0 - V$ is simply a function of q_j as are a_{jk} and a_j in general. This allows us to express Routh's equation as

$$\frac{d}{dt} \frac{\partial R_2}{\partial \dot{q}_j} - \frac{\partial R_2}{\partial q_j} = \frac{\partial R_0}{\partial q_j} - \frac{\partial V}{\partial q_j} - \left[\frac{d}{dt} \frac{\partial R_1}{\partial \dot{q}_j} - \frac{\partial R_1}{\partial q_j} \right] + \tilde{Q}_j, \quad (1.37)$$

1. Dynamical Systems and Stability Theory

The bracketed term containing R_1 can be written as

$$\left(\frac{\partial a_j}{\partial q_k} - \frac{\partial a_k}{\partial q_j} \right) \dot{q}_k = g_{jk} \dot{q}_k, \quad (1.38)$$

where g_{jk} is the skew-symmetric coefficient of the gyroscopic force. We say that the system is gyroscopically decoupled if $R_1 = 0$. Defining a new *effective* potential $U = V - R_0$, Equation 1.37 becomes

$$\frac{d}{dt} \frac{\partial R_2}{\partial \dot{q}_j} - \frac{\partial R_2}{\partial q_j} = -\frac{\partial U}{\partial q_j} - g_{jk} \dot{q}_k + \tilde{Q}_j. \quad (1.39)$$

One can easily show that the gyroscopic forces do no work (Merkin, 2012), and if $\tilde{Q}_j = 0$ we also obtain the energy integral

$$R_2 - U = \text{const.} \quad (1.40)$$

1.10 Stability of Stationary Motions

An unperturbed motion where all the non-cyclic positional coordinates and the cyclic velocities remain at their constant initial value is called stationary motion. That is, we have

$$q_j = q_{j0}, \quad \dot{q}_j = 0, \quad \varphi_j = \varphi_{j0} \quad (1.41)$$

Systems exhibiting this kind of motion are of particular interest to us and we deal with just such a system in Paper I. The stability of stationary motions was originally investigated by Routh who proposed the following theorem which we state as given in Merkin (2012) without a formal proof below.

Theorem 7. *If for a generalized system the potential energy $U = V - R_0$ attains its minimum in a stationary motion, then, at least for a perturbation in which the magnitudes of the cyclic integrals c_j are not altered, this stationary motion is stable with respect to positional coordinates q_j and velocities \dot{q}_j .*

One can prove this in the Lyapunov sense for conservative systems. However, this works only when the cyclic integrals are not themselves altered by the perturbations. This limitation is to some extent alleviated with the following theorem.

Theorem 8 (Merkin, 2012). *If the potential energy of the generalized system U has a minimum with respect to: a) the given $p_{\varphi_j} = c_j$, corresponding to the stationary motion under consideration and b) parameters $p_{\varphi_j} = c_j + \eta_j$, which are sufficiently close to the initial values for small absolute values of η_j , and moreover, if those q_k for which the potential energy is minimized are continuous functions of p_j , then the stationary motion is stable with respect to q_k and \dot{q}_k .*

The question of the whether the inverse is true, i.e., whether the equilibrium is unstable when it is not an isolated minimum, crucially depends on the presence of the gyroscopic force. However, in the absence of such a force, a theorem due to Chetaev applies and is stated without proof. This is identical to the general case of conservative systems that may lack any kind of symmetry.

Theorem 9 (Merkin, 2012). *If for an isolated stationary motion of a gyroscopically decoupled system with cyclic integrals, U is an analytic function of q_j and has no minimum, then the stationary motion is unstable.*

As alluded to earlier, stability may be possible at an equilibrium that is not an isolated minimum if the system is in fact gyroscopically coupled. We explore this possibility further below.

1.11 Gyroscopic Stabilization

A common class of dynamical systems consists of one with two essential or positional coordinates and possibly one or more cyclic coordinates (Merkin, 2012). Indeed, we shall be concerned with a prototypical example of just such a system in Paper I. An important route to stability that applies to such a system is that of gyroscopic stabilization. This involves the possible stabilization of an otherwise unstable equilibrium under the action of gyroscopic forces as introduced in Section 1.7. We examine below the conditions under which such stabilization is possible.

We consider here a system under the action of linear potential, gyroscopic and dissipative forces given by the equations of motion (Bloch et al., 1994; Merkin,

1. Dynamical Systems and Stability Theory

2012)

$$\ddot{z}_1 - g\dot{z}_2 + d\dot{z}_1 + \alpha z_1 = 0, \quad (1.42)$$

$$\ddot{z}_2 + g\dot{z}_1 + d\dot{z}_2 + \beta z_2 = 0, \quad (1.43)$$

where g is the coefficient of the gyroscopic force, d is the coefficient of the dissipative force and α, β are the coefficients of the conservative force. Here, z_1 and z_2 are the positional coordinates of the perturbed motion.

Let us first consider the case without dissipation $d = 0$. The characteristic polynomial for the undamped system is given by

$$\sigma^4 + (g^2 + \alpha + \beta)\sigma^2 + \alpha\beta = 0. \quad (1.44)$$

The coefficients α and β are referred to as stability coefficients and they may be independently positive or negative. If one of the two coefficients are negative, then we say the system has an odd degree of instability and if they are both negative, we say the system has an even degree of instability (Merkin, 2012). This concept may be generalized to systems with greater degree of freedom. In all cases, it is the degree of instability, i.e, whether they are even or odd, and not the actual number of positive or negative stability coefficients that is paramount.

We can determine stability or instability by examining the dispersion relation Equation 1.44. Three possibilities exist depending on the signs of the stability coefficients.

1. If $\alpha > 0, \beta > 0$, the roots of the characteristic polynomial or the eigenvalues are on the imaginary axis. In this case, the system is spectrally stable. However, whether they are linearly stable is to be discerned from the conditions posed by Theorem 4.
2. If α and β have opposite signs or equivalently if the degree of instability is odd, there is one eigenvalue pair on the real axis and one on the imaginary axis. One of the eigenvalues has a positive real part and the system is unstable regardless of the presence of the nonlinear terms.

3. If $\alpha < 0$ and $\beta < 0$, the degree of instability is even. Determination of instability requires closer scrutiny. The roots of the characteristic polynomial has the discriminant D as given below

$$D = g^4 + 2g^2(\alpha + \beta) + (\alpha - \beta)^2. \quad (1.45)$$

- (a) If $D < 0$, two roots occupy the right half plane and two occupy the left. The system is therefore unstable.
- (b) If $D \geq 0$ and $g^2 + \alpha + \beta > 0$, all roots lie on the imaginary axis. The system is spectrally stable here. However, one would need to check the conditions outlined in Theorem 4 in order to determine linear stability. This mode of stability is what we refer to as *gyroscopic stabilization* as stabilization is only achieved due to the presence of the gyroscopic force represented by g .
- (c) If $D \geq 0$ and $g^2 + \alpha + \beta < 0$, there are two roots on the positive real axis and two on the negative real axis and the system is unstable.

1.12 Dissipation Induced Instability

In the previous section, we demonstrated how an otherwise unstable conservative system may be stabilized due to gyroscopic forces. However, this is not the end of the story. We now look at the situation when dissipative forces are also present as they invariably would be in realistic systems. In particular, we may ask whether gyroscopic stabilization is still possible if dissipative forces are present. The characteristic polynomial for Equation 1.44 considering dissipation is given by

$$\sigma^4 + 2d\sigma^3 + (g^2 + \alpha + \beta + d^2)\sigma^2 + d(\alpha + \beta)\sigma + \alpha\beta = 0. \quad (1.46)$$

By using the Routh-Hurwitz criterion, it is easy to show that the system is unstable, i.e., there are roots on the right half plane if

$$\alpha < 0, \quad \beta < 0, \quad g^2 + \alpha + \beta > 0, \quad d > 0. \quad (1.47)$$

1. Dynamical Systems and Stability Theory

Gyroscopic stability is therefore lost despite condition 3(b) being satisfied. This loss of stability is known as *dissipation induced instability* (Bloch et al., 1994). Dissipation induced instability is a significant mode of destabilization and is observed in a number of terrestrial system (Krechetnikov and Marsden, 2007). Since dissipation is a prevalent feature of most physical systems, gyroscopic stability may only be a transient phenomenon. Indeed, stability due to gyroscopic forces is, for this reason, sometimes referred to as temporary stability in the literature (Merkin 2012; Thomson and Tait 1883a,b).

Chapter 2

Charged Particle Dynamics and Stability

In this chapter, we present a summary of the analysis and results of Paper I. The centerpiece of this paper is the study of the effect of gyroscopic and dissipative forces on the stability of a charged particle in an equilibrium circular orbit. We begin with a brief discussion of an astrophysically relevant scenario where the results of our stability analysis may apply and also point to recent work in this context. We close with the summary of Paper I and a brief mention of future prospects.

2.1 Charged Particle Dynamics

There are many situations of practical interest where one is concerned with the dynamics and stability of charged particle motion. Indeed one of the earliest and well studied applications is with regard to charged particle motion in a planetary magnetosphere. This problem was first studied in great detail by Störmer (1955) for ions and electrons, whose motion is largely determined by electromagnetic forces, in this case, the Earth's predominantly dipolar magnetic field. However, for particles with relatively lower charge to mass ratios, such as charged dust grains, the dy-

2. Charged Particle Dynamics and Stability

nodynamical properties are enriched when additional influence such as planetary gravitation, planetary rotation, radiation pressure, etc., come into play. The broader class of problems that examines charged particle motion and stability when under the influence of a larger set of forces is sometime referred to as the *general Störmer problem* (Dullin, Horányi, and Howard, 2002).

The dynamics of a charged particle subject to axisymmetric force fields is described by the Lagrangian

$$L = \frac{1}{2}\dot{\mathbf{r}}^2 + \mathbf{A} \cdot \dot{\mathbf{r}} - \Phi. \quad (2.1)$$

Here, \mathbf{r} is the particle coordinates, \mathbf{A} is a vector potential and Φ is a scalar potential that we refer to as the electromagnetic potentials. An overdot denotes the time derivative of the associated variable. The scalar potential Φ may generally be a combination of any conservative potential and so may include, in addition to the true electric field, gravity and centrifugal force in a rotating frame. Similarly, the vector potential \mathbf{A} may include contributions from any force proportional to velocity such as the Coriolis force in a rotating frame in addition to the true magnetic field. The equation of motion, which follow from the Euler-Lagrange equation, is

$$\ddot{\mathbf{r}} = \mathbf{E} + \dot{\mathbf{r}} \times \mathbf{B}, \quad (2.2)$$

where

$$\mathbf{E} = -\nabla\Phi, \quad \mathbf{B} = \nabla \times \mathbf{A}. \quad (2.3)$$

where the electric field \mathbf{E} and the magnetic field \mathbf{B} are to be understood in a general sense as suggested above.

If we conduct the analysis in cylindrical coordinates (ρ, φ, z) , the angular coordinate φ is cyclic due to axisymmetry. In these coordinates, it is useful to express the toroidal component of the vector potential in terms of a scalar flux function so that

$$\psi = \rho A_\varphi. \quad (2.4)$$

The poloidal magnetic field components are therefore

$$B_\rho = -\frac{1}{\rho} \frac{\partial \psi}{\partial z}, \quad B_z = \frac{1}{\rho} \frac{\partial \psi}{\partial \rho}, \quad (2.5)$$

and the total (true) magnetic field is

$$\mathbf{B} = \nabla \psi \times \nabla \varphi + \rho B_\varphi \nabla \varphi, \quad (2.6)$$

Axisymmetry allows us to construct a Routhian function of the form

$$R = \frac{1}{2}(\dot{\rho}^2 + \dot{z}^2) + A_\rho \dot{\rho} + A_z \dot{z} - \mathcal{U}, \quad (2.7)$$

where we define the effective potential

$$\mathcal{U} = \Phi + \frac{\rho^2 \omega^2}{2}. \quad (2.8)$$

The dynamics may therefore be described in the reduced space defined by the coordinates $(\rho, z, \dot{\rho}, \dot{z})$. The quantity ω denotes the angular velocity at a given equilibrium orbit and is related to the associated integral of motion p_φ as

$$p_\varphi = \rho^2 \omega + \psi. \quad (2.9)$$

In the absence of dissipative forces, there is one other integral of motion - the energy integral given by

$$H = \frac{1}{2}(\dot{\rho}^2 + \dot{z}^2) + U. \quad (2.10)$$

2.2 Circular Orbits in a Rotating Magnetosphere

Equilibrium solutions of the system described by the Routh equations of motion are given by critical points of the effective potential, $\delta \mathcal{U} = 0$. These correspond to circular orbits $\dot{\rho} = \dot{z} = 0$ with $\omega = \text{const.}$

We are interested in the stability of such equilibrium circular orbits in a rotating magnetosphere configuration. Here, the charged particle is subject to a central point mass gravitational potential, a dipole magnetic field and an electric field that arises due to the presence of a highly conducting and rotating plasma. This model

2. Charged Particle Dynamics and Stability

is frequently adopted for studying charged particle orbits in a planetary or stellar magnetosphere. In a series of papers (Dullin, Horányi, and Howard 2002; Howard 1999; Howard, Dullin, and Horányi 2000; Howard, Horányi, and Stewart 1999) conducted a detailed examination of the stability characteristics of circular orbits in a planetary magnetosphere. The principal application of the theoretical results were to determine the locations where charged dust grains orbits around planets such as Saturn by analyzing their stability.

We summarize the basic theoretical results of the aforementioned studies below. The rotating magnetosphere model is described in terms of the poloidal flux function ψ which in cylindrical coordinates is given by

$$\psi = \frac{\gamma\rho^2}{r^3}, \quad (2.11)$$

and the scalar potential Φ given by

$$\Phi = -\frac{\mu}{r} + \Omega\psi, \quad (2.12)$$

where γ is proportional to the magnetic dipole moment, μ is the coefficient of the Newtonian gravitational potential, and Ω is the planetary spin rate. The second term in Equation 2.12 is the electric field due to the corotating magnetosphere.

The effective potential for this system is then given by

$$\mathcal{U} = -\frac{\mu}{r} + \Omega\psi + \frac{\rho\omega^2}{2}. \quad (2.13)$$

For this system, equilibrium circular orbits can be found in the equatorial plane and in planes above or below it (halo orbits). The equilibrium equatorial orbits in cylindrical coordinates are solutions of

$$\rho^2\omega^3 = \mu + \gamma(\omega - \Omega). \quad (2.14)$$

The equilibrium halo orbits are more compactly expressed in spherical polar coordinates (r, φ, ϑ) and are solutions of

$$r^3\omega^2 = -2\gamma(\omega - \Omega), \quad \text{and} \quad \sin^2 \vartheta = -\frac{\mu}{3\gamma(\omega - \Omega)}. \quad (2.15)$$

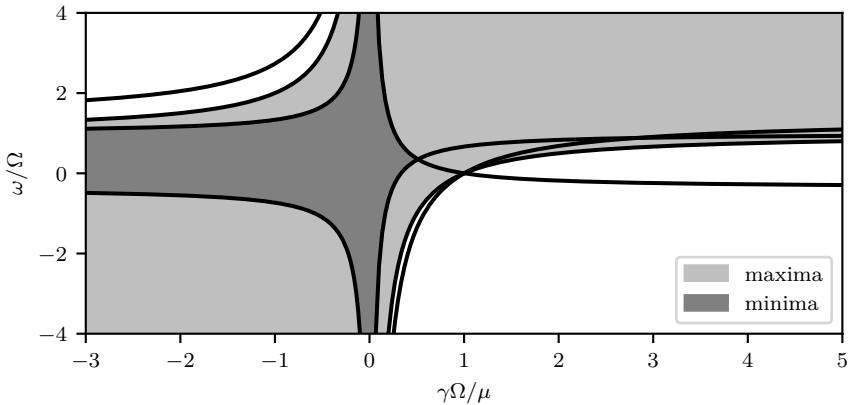


Figure 2.1: Stability diagram for equilibrium equatorial orbits. Here, the regions highlighted in the darker shade of gray correspond to potential minima and are therefore stable. The regions highlighted in light gray correspond to potential maxima including saddle points and are therefore unstable.

If the system is conservative, Routh's theorem, Theorem 7 applies. The stability of the system depends upon whether the equilibrium orbits correspond to isolated effective potential minima. That is, equilibrium orbits where $\delta^2\mathcal{U} > 0$ or which, is tantamount to requiring that the trace and determinant of the Hessian matrix \mathcal{Q} of the effective potential are positive at the equilibrium orbit in consideration. In cylindrical coordinates, this is given by

$$\text{tr } \mathcal{Q} = \frac{\partial^2 \mathcal{U}}{\partial \rho^2} + \frac{\partial^2 \mathcal{U}}{\partial z^2} > 0, \quad (2.16)$$

and

$$\det \mathcal{Q} = \frac{\partial^2 \mathcal{U}}{\partial \rho^2} \frac{\partial^2 \mathcal{U}}{\partial z^2} - \left(\frac{\partial^2 \mathcal{U}}{\partial \rho \partial z} \right)^2 > 0. \quad (2.17)$$

Figures 2.1 and 2.2 illustrate the equilibrium orbits that are stable or unstable as dictated by Routh's Theorem, Theorem 7.

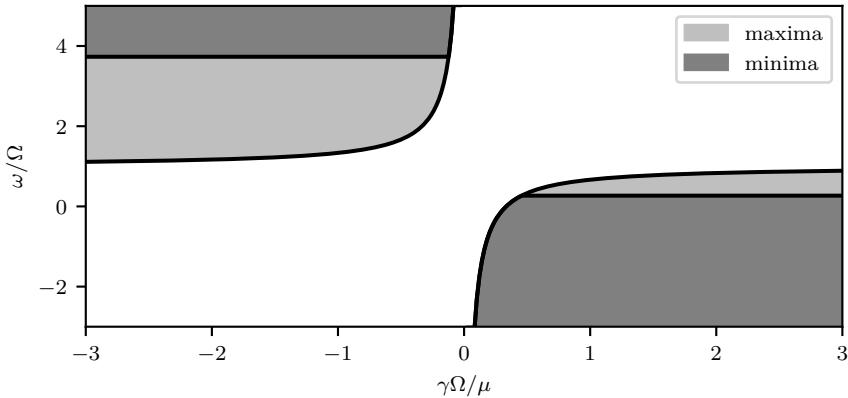


Figure 2.2: Stability diagram for equilibrium halo orbits. Here, the regions highlighted in the darker shade of gray correspond to potential minima and are therefore stable. The regions highlighted in light gray correspond to potential maxima are therefore unstable.

2.3 Charged Dust in Planetary Rings

Dust is as ubiquitous as gas in space and they often coexist and interact with each other forming what is known as a dusty plasma (Goertz, 1989). It is in fact this interaction that allows them to collect electrostatic charge and respond to electromagnetic forces in addition to other forces such as gravity, drag and radiation pressure (Horányi, 1996). The dust-plasma interaction can be quite complex and intertwined. However, when the characteristic Debye shielding distance in the plasma is much greater than the average inter dust-grain distance, the dust particle may be treated as test particles that are immersed in, but separate from, the gaseous plasma (Horányi, 1996). Dust grains accumulate charge in a number of ways. Exposure to thermal ion or electron currents, secondary electron currents due to collisional ionization, photoelectron currents due to high energy radiation, are the predominant mechanisms by which dust grains are charged. The electrostatic charging of a dust grain is a very complex process but unless the plasma parameters vary significantly along the orbit, it is reasonable to assume that the

charge varies on timescales longer than the orbital timescale.

Planetary magnetospheres are a prominent location where charged dust grains are expected to be found (Esposito, 1993, 2002). We now know that all of the solar system giant planets have ring systems of varying brightness, composition and concentration, see Figure 2.3. Jupiter is believed to have the most expansive dust rings and Saturn’s outer rings are composed primarily of micron-sized dust (Esposito, 1993). Indeed Dullin, Horányi, and Howard (2002), Howard, Dullin, and Horányi (2000), and Howard, Horányi, and Stewart (1999) carried out their analysis with Saturn’s rings as the primary application. Collective effects are expected to be low particularly in the E and G rings of Saturn, and here single particle dynamics is expected to work well. For the same reason, our extension of the stability analysis in Paper I may also find application in such planetary ring systems.

2.4 Paper I - Summary & Prospects

In Paper I, we extend the theory of the stability of circular orbits in axisymmetric fields to uncover new modes of stabilization and destabilization. The new stability properties arise from the mere inclusion of an axisymmetric toroidal magnetic field. We show that the toroidal field can enter the system as a gyroscopic force and therefore introduce the possibility of gyroscopic stabilization. This in turn suggests the possibility of dissipation induced instabilities as described in the previous chapter. In Paper I, we present the theoretical framework for the examination of gyroscopic stability and dissipation induced instability for a charged particle in circular orbit in the most general axisymmetric gravito-electromagnetic field configuration conceivable.

As a practical application, we revisit the case of a charged particle in circular orbit in a rotating magnetosphere. We extend the analysis conducted by Dullin, Horányi, and Howard (2002) to include an axisymmetric toroidal field. We also derive the conditions under which gyroscopic stabilization and dissipation induced destabilization is possible for such a system. Our findings may be of relevance to dust in planetary ring systems however, it is not clear whether the mag-

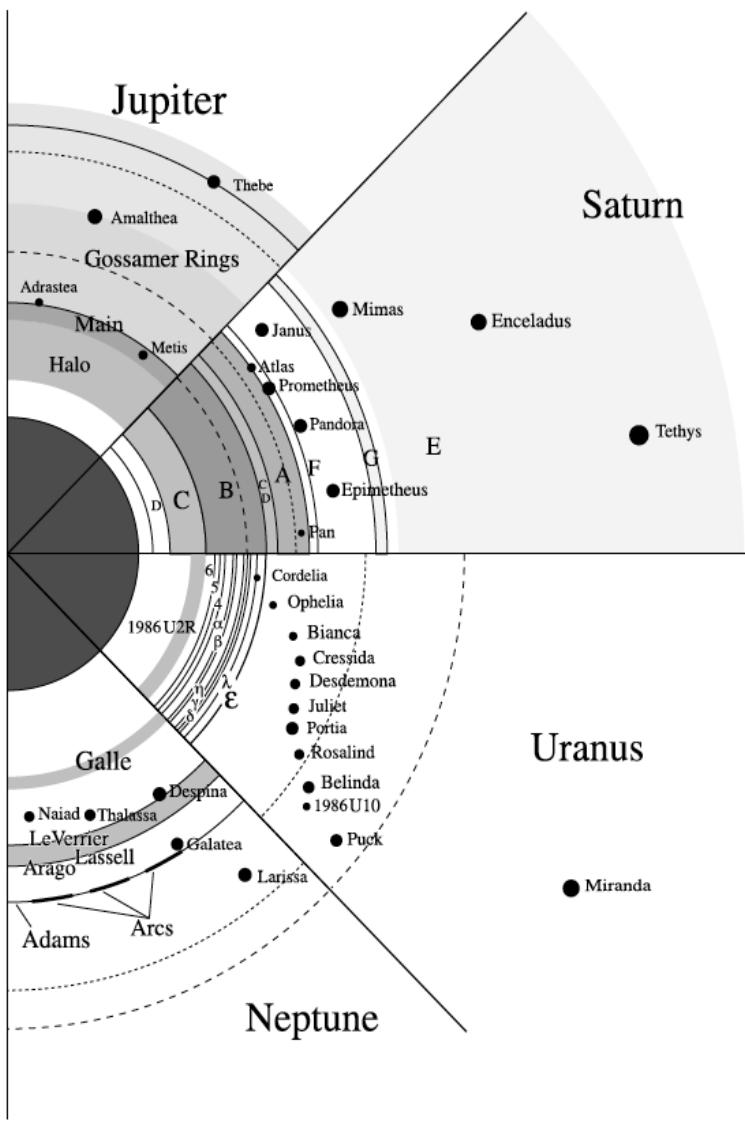


Figure 2.3: A side by side comparison of the ring systems of the solar system planets, scaled to a common planetary equatorial radius, including some of the moons. Figure reprinted from Burns, Hamilton, and Showalter (2001), with permission from Springer Nature.

netic field geometry includes a significant toroidal component at the outer parts of the magnetosphere.

We stress, however, that our theory is much more general and may encompass a wider range of astrophysical applications. For instance, protoplanetary disks present a possible avenue for application as charged dust grains tend to be a significant component of such disks and as is a torodial magnetic field configuration (Armitage, 2015). Other astrophysical possibilities may include debris disks (Wyatt, 2008) and magnetospheres around compact objects (Frank, King, and Raine, 2002). The theory of gyroscopic stabilization and dissipative destabilization extends to fluid and plasma systems as well (Krechetnikov and Marsden, 2009; Morrison, 1998). There is certainly much to explore in this regard in the astrophysical context.

2. Charged Particle Dynamics and Stability

Part II

Non-ideal MRI in Differentially Rotating Disks

Chapter 3

Magnetohydrodynamics

In this chapter, we present a succinct account of magnetohydrodynamic (MHD) theory with a view to provide a minimal theoretical background sufficient to follow the analysis and results of Mohandas and Pessah (2017), which we shall hereafter refer to as Paper II. Magnetohydrodynamic theory is a vast and rigorous scientific disciplines and several distinct flavors or variants of the theory exist (Ogilvie, 2016), each intended to serve a specific type of physical plasma or fluid¹ system. Our goal for this chapter is to introduce a non-relativistic non-ideal version of magnetohydrodynamics that has been widely employed to study the dynamics of protoplanetary disks. The contents of this chapter are mostly based on the classic texts by Krall and Trivelpiece (1973), Nicholson (1983), and Sturrock (1994) and the lecture notes by Ogilvie (2016).

3.1 Kinetics

At its most fundamental level, a plasma is simply a collection of particles, some fraction of which is electrically charged. In the next few sections, we shall layout a path to deriving the MHD equations starting from basic particle considerations. Consider a system of N_0 particles where the position of the i^{th} particle as a func-

¹We shall use the terms *plasma* and *fluid* interchangeably in this part of the thesis.

3. Magnetohydrodynamics

tion of time is denoted by $\mathbf{r}_i(t)$ and its velocity as a function of time is denoted by $\mathbf{v}_i(t)$. We shall describe the dynamics as occurring in a six dimensional phase space given by the independent coordinates (\mathbf{r}, \mathbf{v}) where \mathbf{r} and \mathbf{v} represent the three components of the position and velocity vectors respectively. We define the (number) density of this system as

$$N(\mathbf{r}, \mathbf{v}, t) = \sum_{i=1}^{N_0} \delta[\mathbf{r} - \mathbf{r}_i(t)]\delta[\mathbf{v} - \mathbf{v}_i(t)], \quad (3.1)$$

where the δ 's are Dirac delta functions (Riley, Hobson, and Bence, 2006). If there are more than one species of particle labeled by s , each containing N_{s0} particles, we may define a total density as

$$N = \sum_s N_s. \quad (3.2)$$

where N_s is given by Equation 3.1 for every species. The orbital coordinates $\mathbf{r}_i(t)$ and velocity $\mathbf{v}_i(t)$ of the particles are solutions of the Euler-Lagrange equations of motion and are given by (see Equations 2.2)

$$\dot{\mathbf{r}}_i = \mathbf{v}_i, \quad (3.3)$$

and

$$\dot{\mathbf{v}}_i = \mathbf{g}^m + q_s \mathbf{E}^m + \frac{q_s}{c} \mathbf{v}_i \times \mathbf{B}^m, \quad (3.4)$$

where \mathbf{g}^m , \mathbf{E}^m and \mathbf{B}^m are the gravitational, electric and magnetic fields respectively. The electric charge of a particle of a given species is denoted by q_s . Note that we have also prescribed to the cgs system of units in this part of the thesis. The superscript m indicates that the fields include those that are self-consistently produced by the particles themselves together with those externally applied. They

satisfy Gauss's law of gravity and Maxwell's laws of electrodynamics

$$\nabla \cdot \mathbf{g}^m = -4\pi G \rho^m \quad (3.5)$$

$$\nabla \cdot \mathbf{E}^m = 4\pi \varrho^m, \quad (3.6)$$

$$\nabla \cdot \mathbf{B}^m = 0, \quad (3.7)$$

$$\nabla \times \mathbf{E}^m = -\frac{1}{c} \frac{\partial \mathbf{B}^m}{\partial t}, \quad (3.8)$$

$$\nabla \times \mathbf{B}^m = \frac{4\pi}{c} \mathcal{J}^m. \quad (3.9)$$

where G is the gravitational constant. Note that we are interested in non-relativistic dynamics and therefore ignore the displacement current in Ampere's law, Equation 3.9. Here, we have defined the mass density

$$\rho^m(\mathbf{r}, t) = \sum_s m_s \int d\mathbf{v} N_s(\mathbf{r}, \mathbf{v}, t) \quad (3.10)$$

the charge density as

$$\varrho^m(\mathbf{r}, t) = \sum_s q_s \int d\mathbf{v} N_s(\mathbf{r}, \mathbf{v}, t), \quad (3.11)$$

and the current density as

$$\mathcal{J}^m(\mathbf{r}, t) = \sum_s q_s \int d\mathbf{v} \mathbf{v} N_s(\mathbf{r}, \mathbf{v}, t). \quad (3.12)$$

3.2 Klimontovich Equation

An evolution equation for the plasma is derived by considering the time derivative of the density N_s as

$$\frac{\partial N_s}{\partial t} = \sum_{i=1}^{N_0} \frac{\partial \delta[\mathbf{r} - \mathbf{r}_i(t)]}{\partial t} \delta[\mathbf{v} - \mathbf{v}_i(t)] + \sum_{i=1}^{N_0} \delta[\mathbf{r} - \mathbf{r}_i(t)] \frac{\partial \delta[\mathbf{v} - \mathbf{v}_i(t)]}{\partial t}, \quad (3.13)$$

$$\begin{aligned} &= - \sum_{i=1}^{N_0} \dot{\mathbf{r}}_i \cdot \nabla_{\mathbf{r}} \delta[\mathbf{r} - \mathbf{r}_i(t)] \delta[\mathbf{v} - \mathbf{v}_i(t)] \\ &\quad - \sum_{i=1}^{N_0} \dot{\mathbf{v}}_i \cdot \nabla_{\mathbf{v}} \delta[\mathbf{r} - \mathbf{r}_i(t)] \delta[\mathbf{v} - \mathbf{v}_i(t)], \end{aligned} \quad (3.14)$$

3. Magnetohydrodynamics

where ∇_r and ∇_v are gradients with respect to the position and velocity coordinates respectively. By using a property of the Dirac delta function whereby

$$a\delta(a - b) = b\delta(a - b), \quad (3.15)$$

and substituting the equations of motion given by Equations 3.3 and 3.4 into Equation 3.13, we obtain the Klimontovich equation (Nicholson, 1983)

$$\frac{\partial N_s}{\partial t} + \mathbf{v} \cdot \nabla_r N_s + \mathbf{g}^m \cdot \nabla_v N_s + \frac{q_s}{m_s} \left(\mathbf{E}^m + \frac{\mathbf{v}}{c} \times \mathbf{B}^m \right) \cdot \nabla_v N_s = 0. \quad (3.16)$$

Equation 3.16 is an exact equation which when combined with the Maxwell-Gauss equations and appropriate initial conditions describe the evolution of all the particles constituting the plasma.

3.3 Boltzmann Equation

The Klimontovich equation packs in a lot of information, considerably more than we need or can handle. It is often sufficient to describe the plasma in terms of averaged macroscopic variables such as the plasma density, pressure, velocity or temperature. The first step to arriving at such a description is by defining a suitable averaging procedure. In this case, we consider averaging over an ensemble of an infinite realizations of the plasma (Nicholson, 1983). We first define

$$N_s = f_s + \delta N_s, \quad \langle N_s \rangle = f_s, \quad \langle \delta N_s \rangle = 0, \quad (3.17)$$

$$\mathbf{g}^m = \mathbf{g} + \delta \mathbf{g}, \quad \langle \mathbf{g}^m \rangle = \mathbf{g}, \quad \langle \delta \mathbf{g} \rangle = 0, \quad (3.18)$$

$$\mathbf{E}^m = \mathbf{E} + \delta \mathbf{E}, \quad \langle \mathbf{E}^m \rangle = \mathbf{E}, \quad \langle \delta \mathbf{E} \rangle = 0, \quad (3.19)$$

$$\mathbf{B}^m = \mathbf{B} + \delta \mathbf{B}, \quad \langle \mathbf{B}^m \rangle = \mathbf{B}, \quad \langle \delta \mathbf{B} \rangle = 0. \quad (3.20)$$

where the field variables are decomposed into a mean and fluctuation (prefixed by δ). The quantity $f_s(\mathbf{r}, \mathbf{v}, t)$ is referred to as the *distribution function* and it represents the number of particles with position coordinates in the range \mathbf{r} and $\mathbf{r} + \delta \mathbf{r}$ and velocity coordinates in the range \mathbf{v} and $\mathbf{v} + \delta \mathbf{v}$ of a given species. Upon

averaging Equation 3.16, we obtain

$$\begin{aligned}\frac{\partial f_s}{\partial t} + \mathbf{v} \cdot \nabla_r f_s + \mathbf{g} \cdot \nabla_v f_s + \frac{q_s}{m_s} \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right) \cdot \nabla_v f_s \\ = -\frac{q_s}{m_s} \langle (\delta \mathbf{E} + \frac{\mathbf{v}}{c} \times \delta \mathbf{B}) \cdot \nabla_v \delta N_s \rangle.\end{aligned}\quad (3.21)$$

Equation 3.21 is known as the Boltzmann equation. The terms on the left of the equality contain the average quantities that vary smoothly in phase space whereas on the right of the equality we have sharply varying discrete quantities that are sensitive to the discrete particle nature of the constituent particles in the plasma. In this regard, the terms on the left represent the collective behavior of the plasma and the terms on the right represent the collisional aspects of the plasma.

In some situations, the collisional effects may be neglected and we then have a collisionless version of Equation 3.21 called the collisionless Boltzmann equation² given by

$$\frac{\partial f_s}{\partial t} + \mathbf{v} \cdot \nabla_r f_s + \mathbf{g} \cdot \nabla_v f_s + \frac{q_s}{m_s} \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right) \cdot \nabla_v f_s = 0.\quad (3.22)$$

Equation 3.22 along with the averaged Maxwell and Gauss equations as given by

$$\nabla \cdot \mathbf{g} = -4\pi G\rho \quad (3.23)$$

$$\nabla \cdot \mathbf{E} = 4\pi\varrho, \quad (3.24)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (3.25)$$

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}, \quad (3.26)$$

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathcal{J}, \quad (3.27)$$

²In the plasma physics and plasma astrophysics literature, this equation is sometimes referred to as the Vlasov equation or the Jeans equation. Henon (1982) makes an excellent case for why neither of the two names ought to be used when referring exclusively to Equation 3.22 and why it should be called the collisionless Boltzmann equation.

3. Magnetohydrodynamics

and

$$\rho(\mathbf{r}, t) \equiv \langle \rho^m \rangle = \sum_s m_s \int d\mathbf{v} f_s(\mathbf{r}, \mathbf{v}, t), \quad (3.28)$$

$$\varrho(\mathbf{r}, t) \equiv \langle \varrho^m \rangle = \sum_s q_s \int d\mathbf{v} f_s(\mathbf{r}, \mathbf{v}, t), \quad (3.29)$$

$$\mathcal{J}(\mathbf{r}, t) \equiv \langle \mathcal{J}^m \rangle = \sum_s q_s \int d\mathbf{v} \mathbf{v} f_s(\mathbf{r}, \mathbf{v}, t), \quad (3.30)$$

represent the Vlasov system of equations. We shall be concerned with collisional plasma systems, however, and derive the appropriate system of equations in the following sections.

3.4 MHD Equations

Let us rewrite the collisional Boltzmann equation, Equation 3.21, in a slightly different form below

$$\frac{\partial f_s}{\partial t} + \mathbf{v} \cdot \nabla f_s + \mathbf{g} \cdot \nabla_{\mathbf{v}} f_s + \frac{q_s}{m_s} \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right) \cdot \nabla_{\mathbf{v}} f_s = \left(\frac{\partial f_s}{\partial t} \right)_c. \quad (3.31)$$

Note that we have dropped the subscript \mathbf{r} from the spatial gradient and will continue to do so hereafter. We have also replaced the collisional term on the right with a crude representation of the effect of collisions on the distribution function.

In order to arrive at a set of equations with macroscopic quantities as the primary variables, we take velocity moments of the Boltzmann equation. That is, we multiply Equation 3.31 by a velocity dependent function $\psi(\mathbf{v})$ and integrate over all velocity space. The general moment equation may be expressed as (Sturrock, 1994)

$$\begin{aligned} \frac{\partial(n_s \langle \psi \rangle)}{\partial t} + \nabla \cdot (n_s \langle \psi \mathbf{v} \rangle) - \frac{q_s n_s}{m_s} \left[\left(\frac{m_s}{q_s} \mathbf{g} + \mathbf{E} \right) \left\langle \frac{\partial \psi}{\partial \mathbf{v}} \right\rangle + \frac{1}{c} \left\langle \frac{\partial \psi}{\partial \mathbf{v}} \mathbf{v} \right\rangle \times \mathbf{B} \right] \\ = \left[\frac{\partial(n_s \langle \psi \rangle)}{\partial t} \right]_c. \end{aligned} \quad (3.32)$$

Setting $\psi = 1$ in Equation 3.32, we obtain the first moment equation as

$$\frac{\partial n_s}{\partial t} + \nabla \cdot (n_s \mathbf{v}_s) = 0, \quad (3.33)$$

where we define the species averaged velocity $\mathbf{v}_s = \langle \mathbf{v} \rangle$. Ignoring ionization and recombination processes, we expect that the collisions do not change the number of particles in an infinitesimal volume by very much and so assume $\int (\partial f_s / \partial t)_c d\mathbf{v} = 0$. With $\psi = \mathbf{v}$ in Equation 3.32, we obtain the second moment equation as

$$\frac{\partial(m_s n_s \mathbf{v}_s)}{\partial t} + \nabla \cdot (m_s n_s \langle \mathbf{v}_s \mathbf{v}_s \rangle) = m_s n_s \mathbf{g} + q_s n_s \mathbf{E} + q_s n_s \mathbf{v}_s \times \mathbf{B} + \mathbf{K}_{ss'}, \quad (3.34)$$

We have also defined the second moment of the collisional integral as

$$\mathbf{K}_{ss'} = m_s \int \mathbf{v} \left(\frac{\partial f_s}{\partial t} \right)_c d\mathbf{v}. \quad (3.35)$$

While \mathbf{K}_s for a given species may be non-negligible, by conservation of momentum we expect that

$$\mathbf{K}_{ss'} = -\mathbf{K}_{s's}. \quad (3.36)$$

We shall now define a number of variables starting with the total mass density

$$\rho = \sum_s m_s n_s, \quad (3.37)$$

the total charge density as

$$\varrho = \sum_s q_s n_s, \quad (3.38)$$

the bulk or center of mass velocity as

$$\rho \mathbf{V} = \sum_s m_s n_s \mathbf{v}_s, \quad (3.39)$$

and the bulk fluid current density as

$$\mathbf{J} = \sum_s q_s n_s \mathbf{v}_s. \quad (3.40)$$

Adding together Equation 3.33 for the different plasma species, we obtain the single fluid continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = 0. \quad (3.41)$$

3. Magnetohydrodynamics

It is also useful to work with the relative or peculiar velocity such that its average is given by

$$\langle \mathbf{w}_s \rangle = \mathbf{v}_s - \mathbf{V}. \quad (3.42)$$

Adding together Equation 3.34 of each species and using Equations 3.42 and 3.41, we obtain the single fluid momentum equation as

$$\rho \frac{\partial \mathbf{V}}{\partial t} + \rho (\mathbf{V} \cdot \nabla) \mathbf{V} = \rho \mathbf{g} + \varrho \mathbf{E} + \mathbf{J} \times \mathbf{B} - \nabla \cdot \mathbf{P}, \quad (3.43)$$

where we define the total pressure tensor as $\mathbf{P} = \sum_s \mathbf{P}_s$ and the pressure tensor for a given species as

$$\mathbf{P}_s = \int m_s n_s \mathbf{w}_s \mathbf{w}_s f_s d\mathbf{v}. \quad (3.44)$$

For an isotropic velocity distribution, the pressure tensor is diagonal such that $\mathbf{P}_s = p_s \mathbf{I}_3$ where \mathbf{I}_3 is the three dimensional identity matrix. For a general velocity distribution, the pressure tensor is normally split into a purely diagonal component and a traceless component so that $\mathbf{P}_s = p_s \mathbf{I}_3 + \boldsymbol{\Pi}_s$ (Krall and Trivelpiece, 1973) where

$$\boldsymbol{\Pi}_s = \int m_s n_s \left[\mathbf{w}_s \mathbf{w}_s - \frac{1}{3} \mathbf{w}_s \cdot \mathbf{w}_s \right] f_s d\mathbf{v}. \quad (3.45)$$

By setting $\psi = \mathbf{v} \cdot \mathbf{v}$, one can derive a third moment equation governing the transport of thermodynamic energy. The system of equations is ultimately closed by adopting some closure relation usually specified in the form of an equation of state relating the pressure to density and temperature. We will not derive the third moment equation here since we shall not have occasion to use it in this work and refer the reader to Sturrock (1994) for a detailed derivation.

The magnetic field evolution is fundamentally governed by Faraday's law as given by Equation 3.26. We seek to derive a version of Faraday's law in which the only dependent field variables are the bulk velocity field and the magnetic field vector itself. Such an equation is generally referred to as an *Induction equation* in magnetohydrodynamics. A crucial step in deriving the induction equation is to relate the Electric field to the bulk velocity and magnetic field in terms of a generalized version of Ohm's law. A rigorous derivation of the generalized Ohm's

law is beyond the scope of this work, so we provide here a only crude derivation suitable for a weakly ionized three component (electrons, singly charged ions and neutrals) plasma where the collisions are primarily between the ions and neutral species.

We follow the procedure by Balbus (2009) and start by considering the momentum equations for each plasma species as given by

$$m_s n_s \left(\frac{\partial \mathbf{v}_s}{\partial t} + \mathbf{v}_s \cdot \nabla \mathbf{v}_s \right) = -\nabla \cdot \mathbf{P}_s + \rho_s \mathbf{g} + q_s n_s (\mathbf{E} + \mathbf{v}_s \times \mathbf{B}) + \mathbf{K}_{ss'}. \quad (3.46)$$

Assuming that the time scale for the evolution of the macroscopic variables are generally much longer than the collisional time scale and the electromagnetic interactions, we have for the ions and electrons

$$en_i \left(\mathbf{E} + \frac{1}{c} \mathbf{v}_i \times \mathbf{B} \right) + \mathbf{K}_{in} = 0, \quad (3.47)$$

$$-en_e \left(\mathbf{E} + \frac{1}{c} \mathbf{v}_e \times \mathbf{B} \right) + \mathbf{K}_{en} = 0, \quad (3.48)$$

where we have ignored inter-charge species collisions by virtue of the weak ionization. Assuming charge neutrality, $n_e \approx n_i$, we get

$$\mathbf{K}_{in} + \mathbf{K}_{en} = \mathbf{J} \times \mathbf{B}, \quad \text{where } \mathbf{J} = en_e(\mathbf{v}_i - \mathbf{v}_e). \quad (3.49)$$

In a plasma dominated by the neutral species, it is reasonable to assume $\rho \approx m_n n_n$ and $\mathbf{V} \approx \mathbf{v}_n$. Using Ampere's law, the momentum equation for the neutrals become

$$\rho \left(\frac{\partial \mathbf{V}}{\partial t} + \mathbf{V} \cdot \nabla \mathbf{V} \right) = -\nabla \cdot \mathbf{P}_s + \rho \mathbf{g} + \frac{1}{4\pi} (\nabla \times \mathbf{B}) \times \mathbf{B}, \quad (3.50)$$

Thus the momentum equation for the neutrals in a neutral-dominated plasma is the same as the bulk momentum equation derived in Equation 3.43. Now, expressing the collisional terms as (Shu, 1991)

$$\mathbf{K}_{en} = n_e \nu_{en} m_e (\mathbf{V} - \mathbf{v}_e), \quad (3.51)$$

$$\mathbf{K}_{in} = \rho \rho_i \gamma (\mathbf{V} - \mathbf{v}_i). \quad (3.52)$$

3. Magnetohydrodynamics

Using Equations 3.52 and 3.51 in Equation 3.48 yields

$$\mathbf{E} + \frac{1}{c} [\mathbf{V} + (\mathbf{v}_e - \mathbf{v}_i) + (\mathbf{v}_i - \mathbf{V})] + \frac{\nu_{en} m_e}{e} [\mathbf{v}_e - \mathbf{v}_i + \mathbf{v}_i - \mathbf{V}] = 0. \quad (3.53)$$

If $|\mathbf{K}_{in}| \gg |\mathbf{K}_{en}|$, then we are left with

$$\mathbf{E} + \frac{1}{c} \mathbf{V} \times \mathbf{B} = \frac{\nu_{en} m_e}{e^2 n_e} \mathbf{J} + \frac{1}{cen_e} \mathbf{J} \times \mathbf{B} - \frac{1}{c^2 \gamma \rho \rho_i} (\mathbf{J} \times \mathbf{B}) \times \mathbf{B}. \quad (3.54)$$

Equation 3.54 is the version of the generalized Ohm's law that we require and by using it in Faraday's law, we finally obtain

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times \left[\mathbf{V} - \frac{1}{en_e} \mathbf{J} + \frac{1}{c\gamma\rho\rho_i} (\mathbf{J} \times \mathbf{B}) \right] \times \mathbf{B} - \nabla \times \left(\frac{c^2 \nu_{en} m_e}{4\pi e^2 n_e} \right) \mathbf{J}. \quad (3.55)$$

The first term in square brackets on the right side of Equation 3.55 is the induction term from which the equation derives its name. The second, third and fourth terms on the right are the Hall diffusion, ambipolar diffusion and ohmic diffusion terms respectively.

As mentioned earlier, a number of simplifying assumptions have gone into the derivation of Equation 3.54. We refer the reader to Freidberg (1987), Hazeltine (2018), and Krall and Trivelpiece (1973) for a more rigorous derivation of the generalized Ohm's law. However, to our knowledge, only Pandey and Wardle (2008) have presented a moderately rigorous derivation of the generalized Ohm's law for a weakly ionized neutral species dominated plasma.

3.5 Ideal MHD

Let us briefly comment on a simple yet powerful formulation of magnetohydrodynamics, that of ideal MHD. Ideal MHD serves as a suitable approximation when the hydrodynamic and magnetic dissipative effects are negligible. This situation arises in highly ionized plasma systems. Of course, the approximation depends very much on the scales of interest as even in highly conducting plasmas, dissipation begins to matter at small enough scales.

In the ideal MHD approximation, the induction equation, Equation 3.55 reduces to

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{V} \times \mathbf{B}). \quad (3.56)$$

A remarkable consequence of this equation is the flux freezing effect demonstrated by Alfvén in his now eponymous theorem (Nicholson, 1983). This theorem implies that the magnetic field is “frozen-in” to the fluid in the absence of magnetic diffusion. A magnetic field line therefore acts as a material line element and twists and folds in harmony with the fluid.

An interesting, albeit partial, analogy exists between the ideal induction equation and the vorticity equation for an inviscid barotropic fluid which has the form

$$\frac{\partial \boldsymbol{\omega}}{\partial t} = \nabla \times (\mathbf{V} \times \boldsymbol{\omega}), \quad (3.57)$$

where $\boldsymbol{\omega} = \nabla \times \mathbf{V}$ is the fluid vorticity. Therefore, under these conditions, the vortex lines also exhibit the same frozen-in behavior.

3.6 MHD Waves

The complexities of fluid systems stems from the different variety of wave phenomena they can support. Waves are oscillations in the fluid that occur when some disturbance causes the fluid to deviate from its equilibrium state transporting energy away from the source of disturbance. Magnetic fields only add to the richness of the wave families and their properties. Examples of hydrodynamic waves include acoustic or sound waves and gravity waves (Pringle and King, 2007). We derive here the basic MHD waves under ideal fluid conditions and briefly discuss their attributes.

Here, we look at the basic types of MHD waves and describe their main characteristics. For simplicity, we consider an ideal fluid in static equilibrium, i.e., $\mathbf{V} = 0$. Disturbances to the equilibrium state are represented by the perturbed velocity vector $\delta \mathbf{u}$ as well as the perturbed density, pressure and magnetic fields $\delta \rho$, δp and $\delta \mathbf{B}$. The calculations become somewhat simpler if we work in terms of a

3. Magnetohydrodynamics

displacement vector defined such that $\delta\mathbf{u} = \partial\boldsymbol{\xi}/\partial t$. The continuity and induction equation then have the form

$$\delta\rho = \nabla \cdot (\rho\boldsymbol{\xi}), \quad (3.58)$$

$$\delta\mathbf{B} = \mathbf{B} \cdot \nabla \boldsymbol{\xi} - \mathbf{B} \nabla \cdot \boldsymbol{\xi} - \boldsymbol{\xi} \cdot \nabla \mathbf{B}. \quad (3.59)$$

The momentum equation is given by

$$\rho \frac{\partial^2 \boldsymbol{\xi}}{\partial t^2} = -\nabla \left(\delta p + \frac{\mathbf{B} \cdot \delta \mathbf{B}}{4\pi} \right) + \frac{1}{4\pi} \mathbf{B} \cdot \nabla [\mathbf{B} \cdot \nabla \boldsymbol{\xi} - \mathbf{B}(\nabla \cdot \boldsymbol{\xi})]. \quad (3.60)$$

We shall consider a homogeneous system where the equilibrium pressure, density and magnetic field are uniform. Assuming a barotropic equation of state such that $\delta p = a^2 \delta\rho$ where a is the speed of sound in the fluid, the gradient of the kinetic and magnetic pressure perturbations become

$$\nabla \left(\delta p + \frac{\mathbf{B} \cdot \delta \mathbf{B}}{4\pi} \right) = \left(a^2 \rho + \frac{B^2}{4\pi} \right) \nabla \cdot \boldsymbol{\xi} + \frac{1}{4\pi} \mathbf{B} \cdot \mathbf{B} \cdot \nabla \boldsymbol{\xi}. \quad (3.61)$$

Spatial homogeneity of the basic state allows us to assume a plane waveform for the displacement vector, $\boldsymbol{\xi} = \text{Re}[\tilde{\boldsymbol{\xi}} \exp(i\mathbf{k} \cdot \mathbf{r} - i\omega t)]$. With this, the equation of motion becomes

$$\begin{aligned} \rho\omega^2 \boldsymbol{\xi} &= \mathbf{k} \left[\left(a^2 \rho + \frac{B^2}{4\pi} \right) (\mathbf{k} \cdot \boldsymbol{\xi}) - \frac{1}{4\pi} (\mathbf{k} \cdot \mathbf{B})(\mathbf{B} \cdot \boldsymbol{\xi}) \right] \\ &\quad + \frac{(\mathbf{k} \cdot \mathbf{B})}{4\pi} [(\mathbf{k} \cdot \mathbf{B})\boldsymbol{\xi} - (\mathbf{k} \cdot \boldsymbol{\xi})\mathbf{B}]. \end{aligned} \quad (3.62)$$

where we omit the tilde for convenience. If the displacement is orthogonal to the wavevector and the magnetic field, i.e., $\mathbf{k} \cdot \boldsymbol{\xi} = \mathbf{B} \cdot \boldsymbol{\xi} = 0$, we obtain the wave solution given by the dispersion relation

$$\omega = \pm \mathbf{k} \cdot \mathbf{v}_A, \quad (3.63)$$

These solutions are called Alfvén waves and \mathbf{v}_A is the Alfvén velocity defined as

$$v_A = \frac{B}{\sqrt{4\pi\rho}} \quad (3.64)$$

Alfvén waves are incompressible waves that oscillate transverse to and propagate along the magnetic field and are much like waves on a string with the restoring force provided here by magnetic tension.

Two other wave modes are obtained by considering displacements that have a parallel component with respect to the wavevector and the magnetic field. By taking the dot product of Equation 3.62 with \mathbf{k} and \mathbf{B} , we obtain the system of equations

$$\begin{bmatrix} \rho\omega^2 - \left(a^2\rho + \frac{B^2}{4\pi}\right)k^2 & \frac{k^2}{4\pi}(\mathbf{k} \cdot \mathbf{B}) \\ -a^2\rho(\mathbf{k} \cdot \mathbf{B}) & \rho\omega^2 \end{bmatrix} \begin{bmatrix} \mathbf{k} \cdot \boldsymbol{\xi} \\ \mathbf{B} \cdot \boldsymbol{\xi} \end{bmatrix} = 0. \quad (3.65)$$

Non-trivial solutions require that the following equation is satisfied

$$\omega^4 - \omega^2 k^2(a^2 + v_A^2) + k^2 a^2 (\mathbf{k} \cdot \mathbf{v}_A)^2 = 0. \quad (3.66)$$

Without loss of generality, let us assume $\mathbf{k} \parallel \mathbf{v}_A$. We then obtain two additional wavemodes

$$2\omega_{\pm}^2 = k^2(a^2 + v_A^2) \pm \sqrt{k^4(a^2 + v_A^2)^2 - 4k^2 a^2 (\mathbf{k} \cdot \mathbf{v}_A)^2}. \quad (3.67)$$

ω_+ represents fast magnetoacoustic waves and ω_- represents slow magnetoacoustic waves. Both the fast and slow waves constitute compressible longitudinal oscillations. The fast magnetoacoustic wave is a quasi-isotropic wave in which the magnetic and gas pressure act in concert. The slow magnetoacoustic wave is an acoustic-like wave that is strongly oriented along the magnetic field and where the magnetic and gas pressure forces act in opposition.

3.7 Shearing Sheet Equations

Ideal or non-ideal MHD theory is frequently used to describe the dynamics of stars and accretion disks. The magnetized plasma or fluid that makes up an accretion disk is predominantly in rotational motion around the central object as a result of angular momentum conservation. Therefore, the dominant velocity component

3. Magnetohydrodynamics

is azimuthal in cylindrical coordinates (r, φ, z) and is simply

$$\mathbf{V} = r\Omega(r, z)\hat{\varphi}. \quad (3.68)$$

The balance of forces in the radial direction is assumed to be largely between the centrifugal force of the rotating fluid and the gravitational force due to the central object. Thus, we have

$$\Omega = \sqrt{-\frac{1}{r} \frac{\partial \Phi}{\partial r} + \mathcal{O}\left(\frac{H}{R}\right)^2}. \quad (3.69)$$

where Φ is the gravitational potential and H and R are characteristic vertical and radial length scales for the disk. The $\mathcal{O}(H/R)$ terms may be due to hydrodynamic and magnetic pressure gradients. In cold protostellar or protoplanetary disks, $H/R \ll 1$ as a result of which, disks are largely centrifugally supported although small gradients in the thermodynamic properties could potentially be relevant to and affect disk stability (Barker and Latter, 2015). In the absence of self-gravity, Φ is just the Newtonian potential. Consequently, the disk plasma rotates at a near Keplerian rate so that fluid elements closer to the central object rotate faster than those farther away resulting in shear between adjacent layers of the disk fluid.

The plasma or fluid in an accretion disk is spread across a truly astronomical extent. Disks around young stellar objects span 100 AU or more. Capturing the dynamics across the vast range of scales is extremely difficult if not impossible. A very useful technique that allows one to study the local physics is afforded by the shearing sheet or shearing box system of equations (Binney and Tremaine 2011; Goldreich and Lynden-Bell 1965). Here, one derives equations of motion in a localized Cartesian frame centered around a fiducial radius where the differential rotation is represented as a planar shear.

The shearing sheet equations are derived by considering a small patch of the disk around an arbitrarily chosen fiducial point in the disk labeled by the coordinates $(r_0, \varphi_0 + \Omega_0 t, z_0)$. Without loss of generality, we set the reference angle $\varphi_0 = 0$ and the height $z_0 = 0$. A local representation of the rotational velocity

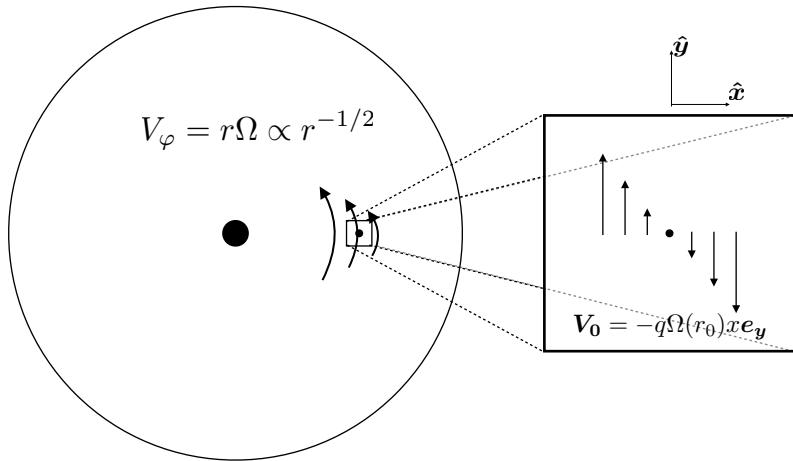


Figure 3.1: Shearing sheet schematic

is obtained by an expansion around the fiducial radius r_0 such that in Cartesian coordinates with origin at r_0 , we have

$$\mathbf{V}_0(r_0 + x) = (r_0 + x)\Omega(r_0 + x, z)\mathbf{e}_\varphi = -q\Omega_0 x \mathbf{e}_y, \quad (3.70)$$

where we have defined the dimensionless shear rate

$$q \equiv - \left. \frac{\partial \ln \Omega}{\partial \ln r} \right|_{r=r_0}. \quad (3.71)$$

that is related to Oort's first parameter A (Binney and Tremaine, 2011). We transform the MHD equations to a rotating frame around r_0 . In particular the momentum equations with the Coriolis and centrifugal forces become

$$\rho \left(\frac{\partial \mathbf{V}}{\partial t} + \mathbf{V} \cdot \nabla \mathbf{V} \right) = 2\rho \mathbf{V} \times \boldsymbol{\Omega} - \rho \nabla \Phi_{cg} - \nabla \cdot \mathbf{P} + \nabla \cdot \mathbf{M}, \quad (3.72)$$

where we have defined the combined gravitation-centrifugal potential as

$$\Phi_{cg} = \Phi - \frac{\Omega_0^2 r^2}{2}, \quad (3.73)$$

3. Magnetohydrodynamics

and the general magnetic stress tensor

$$\mathbf{M} = \mathbf{B}\mathbf{B} - \frac{(\mathbf{B} \cdot \mathbf{B})}{2} \mathbf{I}_3. \quad (3.74)$$

Assuming dominant centrifugal support in Equation 3.69, we also expand the combined gravito-centrifugal potential around r_0 radially and vertically in a power series. If we also assume axisymmetry of the fields, we obtain the following system of equations in Cartesian coordinates (x, y, z)

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \quad (3.75)$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = 2\mathbf{u} \times \boldsymbol{\Omega}_0 + 2q\Omega_0^2 x \mathbf{e}_x - \Omega_0^2 z \mathbf{e}_z - \frac{1}{\rho} \nabla \cdot \mathbf{P} + \frac{1}{\rho} \nabla \cdot \mathbf{M}, \quad (3.76)$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times \left[\mathbf{u} - \frac{1}{en_e} \mathbf{J} + \frac{1}{c\gamma\rho\rho_i} (\mathbf{J} \times \mathbf{B}) \right] \times \mathbf{B} - \nabla \times \left(\frac{c^2 \nu_{en} m_e}{4\pi e^2 n_e} \right) \mathbf{J}. \quad (3.77)$$

Equations 3.75 -3.77 represent the shearing box or shearing sheet(ignoring vertical stratification) system of equations. They may be further simplified by assuming axisymmetry and incompressibility. Note that \mathbf{u} represents the total velocity field and may consist of perturbations in addition to the equilibrium solution \mathbf{V}_0 . Without axisymmetry, an additional spatial gradient of \mathbf{u} representing advection by the shear flow \mathbf{V}_0 would be present. This gradient has an x dependence which may be eliminated by a further transformation of coordinates at the cost of an additional time derivative.

3.8 Magnetorotational Instability

As with individual particles, wave or fluid motion may be stable or unstable. Instabilities in fluids or plasmas can have very important consequences and play a vital role in determining the dynamics of many physical systems. In this section, we describe one such instability called the magnetorotational instability (MRI) that is of immense significance to astrophysical disks and is the subject of our investigation in Paper II.

We introduce below the simplest version of the MRI derived in the shearing-sheet approximation. We will continue to work with the displacement vector. However, due to the background planar shear, the displacement vector is related to the perturbed velocity as

$$\delta \mathbf{u} = \frac{\partial \boldsymbol{\xi}}{\partial t} + \mathbf{V}_0 \cdot \nabla \boldsymbol{\xi} - \boldsymbol{\xi} \cdot \nabla \mathbf{V}_0. \quad (3.78)$$

On the other hand, the induction equation as given by Equation 3.59 remains unchanged despite the background shear. We assume the plane waveform solution for the perturbations as in the previous section but only consider a uniform equilibrium field and wavevector that are both aligned in the z direction. The equations of motion are then given by

$$\omega^2 \xi_x - 2i\omega\Omega\xi_y = -(2q\Omega - k^2 v_A^2) \xi_x, \quad (3.79)$$

$$\omega^2 \xi_y + 2i\omega\Omega\xi_x = k^2 v_A^2 \xi_y, \quad (3.80)$$

From Equations 3.79 and 3.80, we obtain the dispersion relation

$$\omega^4 - \omega^2(\kappa^2 + 2k^2 v_A^2) + k^2 v_A^2(k^2 v_A^2 - 2q\Omega^2) = 0, \quad (3.81)$$

where we have defined the radial epicyclic frequency $\kappa = \sqrt{2(2-q)}\Omega$. We shall discuss the characteristic traits of the different solutions to the dispersion relation in greater detail in the following chapter. However, let us point out a very interesting feature of Equation 3.81. Note that unlike Equation 3.66 whose solutions were the standard MHD waves, one set of the two distinct pairs of solutions of Equation 3.81 can be purely imaginary. That is, for

$$kv_A < \sqrt{2q}\Omega, \quad (3.82)$$

one of the solutions represents exponentially growing perturbations. Thus, for $q > 0$, it is possible to find a finite range of wavenumbers which grow exponentially provided the magnetic field is weak. This leads to an unstable situation and is called the magnetorotational instability (Balbus and Hawley, 1998).

3. Magnetohydrodynamics

The instability has a maximum growth rate of $\omega_{\max} = iq\Omega/2$ which occurs at wavenumber

$$k_{\max}v_A = \frac{q\Omega}{2} \sqrt{\frac{4}{q} - 1}. \quad (3.83)$$

This unstable mode was first discovered by Chandrasekhar (1960b) and Velikhov (1959) in the context of Taylor-Couette flow experiments. However, its astrophysical significance was not appreciated until Balbus and Hawley (1991) proposed the instability as the driver of angular momentum transport in accretion disks as we shall discuss in the following chapter.

The most remarkable aspect of the dispersion relation above is that the stability characteristics change abruptly with the addition of a magnetic field. Without a magnetic field, the solutions are all stable waves and stability is governed by Rayleigh's criterion which requires $\kappa^2 > 0$ or $q > 2$. However, there is a discontinuous change in stability criteria, given by Equation 3.82, that arises with the introduction of the slightest magnetic field. One can also derive the MRI in global calculations (Gammie and Balbus, 1994; Latter, Lesaffre, and Balbus, 2009). Therefore, there is little doubt over the potential of the MRI to drive a rapidly growing disk wide MHD instability. As we will see in the next chapter, this instability is widely considered to be the main vehicle of angular momentum transport in most types of accretion disks.

Chapter 4

Non-ideal MRI in Accretion Disks

In this chapter, we provide the relevant context for our study of non-ideal effects on the magnetorotational instability. We provide an overview of the significance of the MRI in accretion disks as a driver of angular momentum transport. This is followed with a discussion of the non-ideal conditions expected in disks around young stars and how they alter the character of the MRI. Following a short account of nonlinear numerical simulations of disks, we conclude with a summary of our work as presented in Paper II.

4.1 Angular Momentum Transport in Accretion Disks

The universe abounds with accretion disks. They are often present around young stars as they are born and can be found around them in their twilight years. Accretion disks form around stellar objects because much of the disk material has excess angular momentum that it cannot easily rid itself of (Pringle, 1981). The time scale for angular momentum exchange is usually much longer than the free-fall time scale and so the infalling gas settles into a disk like structure with the star in the center. Eventually, gravity prevails and much of the disk material finds its

4. Non-ideal MRI in Accretion Disks

way onto the central object. More than being mere buffers of gas, accretion disks play very important roles. Disks around younger stars crucially set the stage for the formation and evolution of planetary systems (Armitage, 2015). Disks around older stars such as white-dwarfs, neutron stars and black holes are responsible for powering the most energetic phenomena in the universe (Frank, King, and Raine, 2002).

Accretion disks can be classified both in terms of the type of central stellar object around which they orbit and also by the nature of the plasma that constitute the disk matter. The two classifications are not unrelated. Disks around young stars mostly comprise of a weakly ionized cold plasma whereas disks around compact stellar objects tend to be hotter and strongly ionized. Dust is also an important component of disks around young stars (Dullemond and Monnier, 2010) as well as those around active galactic nuclei (Netzer, 2015).

Observing accretion disks is a challenging enterprise. Disks around binary systems are compact in size. Nevertheless, their presence is revealed by the hotspot emanating from the stream of material crashing onto itself as it is drawn from the secondary star. The double peaked profile in the Balmer emission lines and the spectral energy density in the optical and ultraviolet are other strong indications of a disk structure (Lin and Papaloizou, 1996). Disks around active galactic nuclei are very far away and lie shrouded in an obscure haze of dust. Their presence is betrayed however by the excess in the blue and ultraviolet continuum and also by MASER activity (Lin and Papaloizou, 1996). Disks around young stars have presented the most promising sources for direct imaging of a disk and significant strides have been taken in this direction recently with very large baseline instruments becoming operational. Remarkably, the disk shadows or silhouettes in a bright background first revealed the presence of protostellar disks in the optical wavelength range. Scattered light from dust has, however, produced the best images of a protoplanetary disk to date (Lin and Papaloizou, 1996; Williams and Cieza, 2011).

The fundamental problem in accretion disk theory, for a long time, was to

identify the means by which angular momentum could be transported away from the disk enabling matter to fall inwards (Pringle, 1981). The radially varying angular velocity profile causes adjacent layers of the disk fluid to shear. However, the kinematic viscosities of the disk gas are far too feeble to enable significant angular momentum exchange. On the other hand, the small viscosities coupled with the immense range of spatial and temporal scales associated with a disk implies gigantic Reynolds numbers by any chosen measure. Such high Reynolds numbers would naturally suggest that the disk fluid or plasma is prime for turbulence. If correlated turbulent fluctuations could sustain and grow to substantial amplitude, they could act as an *effective viscosity* to drive accretion at realistic rates (Shakura and Sunyaev, 1973).

The trouble lies in the fact that Keplerian disks are hydrodynamically stable by the Rayleigh criterion. This fact poses a great hurdle to achieving outward angular momentum transport by purely hydrodynamic means, though it is by no means ruled out entirely¹. However, as we saw in the previous chapter, the stability properties of a differentially rotating disk are dramatically altered by the presence of a weak magnetic field. With a weak field, there is a singular shift in the criterion of stability. Weak magnetic fields are expected to be a prevalent feature in astrophysical fluid systems and for sufficiently ionized Keplerian disks, the magnetorotational instability is an inevitable outcome. With a growth rate of the order of the angular frequency, the perturbations would rapidly grow to saturation resulting in magnetohydrodynamic turbulence and possibly driving mass accretion.

Assuming complete axisymmetry, the mean angular momentum equation for a differentially rotating (ideal) disk is

$$\frac{\partial \bar{l}}{\partial t} + \nabla \cdot \bar{l}\bar{V} = -\nabla \cdot (r\bar{F}_i), \quad (4.1)$$

¹Hydrodynamic instability could arise in the form of the vertical shear overstability (Barker and Latter, 2015) or the convective overstability (Klahr and Hubbard, 2014) provided the disk fluid is baroclinic, i.e. isobars and isopyynes are not parallel. However, these instabilities grow at slow rates and are sensitive to the rate at which the fluid cools.

4. Non-ideal MRI in Accretion Disks

where $\bar{l} = r\rho V_\varphi$ is the specific angular momentum. The right hand side of Equation 4.1 is the gradient of the flux of angular momentum brought about by stresses due to velocity and magnetic perturbations

$$\bar{F}_i = R_{i\varphi} - M_{i\phi}, \quad i = r, z, \quad (4.2)$$

where

$$R_{i\varphi} = \overline{\delta u_i \delta u_\varphi}, \quad M_{i\varphi} = \frac{\overline{\delta B_i \delta B_\varphi}}{4\pi\rho}. \quad (4.3)$$

Assuming steady state accretion, radially integrating Equation 4.1 and specifying the boundary conditions results in an equation in which $\bar{F}_i > 0$ suggests an outward angular momentum flux (Balbus and Hawley, 1998; Popham and Narayan, 1991).

4.2 The Eigenvalue Problem for Ideal MRI

As a precursor to describing the analysis carried out in Paper II, we shall look at the ideal MRI eigenvalue problem. A thorough analysis of this problem in the shearing sheet framework was carried out by Pessah, Chan, and Psaltis (2006b) for the ideal MRI and by Pessah and Chan (2008) for viscous-resistive MRI. We build upon this formalism in Paper II to include all three non-ideal effects.

Assuming homogeneity and periodicity over a domain of vertical extent $2H$, we may express the perturbed fluid quantities in Fourier series as (Boas, 2006)

$$\delta f(z, t) = \sum_{n=-\infty}^{\infty} \delta \hat{f} \exp(ik_n z), \quad (4.4)$$

where $k_n = n\pi/H$ and δf represents any of the perturbed variables.

Assuming a uniform vertical background magnetic field, uniform density, an isothermal equation of state and non-diffusive conditions, Equations 3.76 and 3.77 may be compactly expressed as a system of first order differential equations in time as

$$\frac{\partial}{\partial t} \hat{\boldsymbol{\delta}}(k_n, t) = \mathbf{L} \hat{\boldsymbol{\delta}}(k_n, t), \quad (4.5)$$

where δ represents the vector of perturbations

$$\hat{\boldsymbol{\delta}} = \begin{bmatrix} \delta\hat{u}_x \\ \delta\hat{u}_y \\ \delta\hat{b}_x \\ \delta\hat{b}_y \end{bmatrix}. \quad (4.6)$$

We also define the linear operator

$$\mathbf{L} = \begin{bmatrix} 0 & 2\Omega & i\omega_A & 0 \\ (q-2)\Omega & 0 & 0 & i\omega_A \\ i\omega_A & 0 & 0 & 0 \\ 0 & i\omega_A & -q\Omega & 0 \end{bmatrix}, \quad (4.7)$$

where $\omega_A = k_n v_A$ is the Alfvén frequency. The operator \mathbf{L} has four eigenvalues σ_j and four corresponding eigenvectors e_j for $j = 1, \dots, 4$. The characteristic polynomial of \mathbf{L} yields the dispersion relation of Equation 3.81 (with $\sigma = -i\omega$). Since the dispersion relation only contains even powers of σ , there are only two distinct pairs of modes $\sigma = \pm\nu, \pm\gamma$ that are solutions of Equation 3.81. These are

$$\nu = (\sqrt{\Delta} + \Lambda)^{1/2}, \quad (4.8)$$

$$\gamma = (\sqrt{\Delta} - \Lambda)^{1/2}, \quad (4.9)$$

where

$$\Lambda = k_n^2 v_A^2 + (2 - q)\Omega^2, \quad (4.10)$$

$$\Delta = 4k_n^2 v_A^2 + (2 - q)^2 \Omega^2. \quad (4.11)$$

The solutions $\pm\nu$ are purely imaginary and represent wave modes that oscillate with approximately the epicyclic frequency κ at very low wavenumbers and behave like Alfvén waves at high wavenumbers. The solutions $\pm\gamma$ are purely real for a finite range of wavenumbers and represent exponentially growing (the MRI) and decaying perturbations respectively. Figure 4.1 shows the eigenvalues plotted as a function of wavenumber.

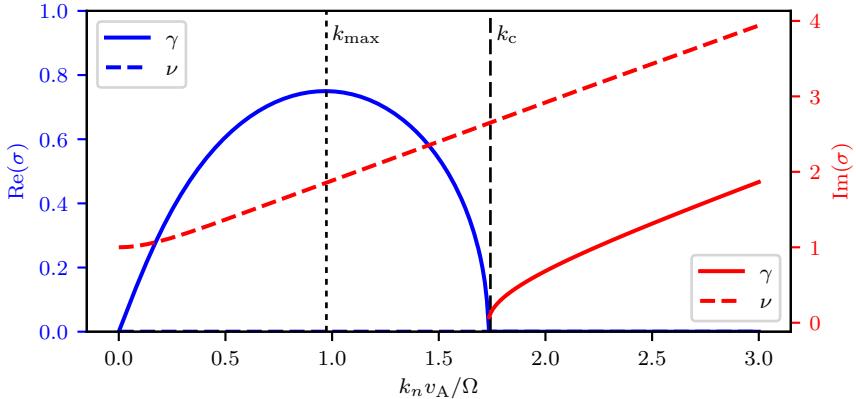


Figure 4.1: The eigenvalues of the operator \mathbf{L} as a function of wavenumber k_n . The real part (in blue) of the eigenmode γ is the growth rate of the ideal magnetorotational instability.

As \mathbf{L} is a normal operator (Riley, Hobson, and Bence, 2006), the eigenvectors of \mathbf{L} constitute a linearly independent basis and we may represent any vector $\hat{\delta}$ as a linear combination

$$\hat{\delta} = \sum_{j=1}^4 a_j(t) e_j, \quad a_j \in \mathbb{C}_4, \quad (4.12)$$

where \mathbb{C}_4 represents the four dimensional space of complex numbers. Substituting Equation 4.12 in the evolution equation, Equation 4.5, we obtain the solution

$$\hat{\delta}(k_n, t) = \sum_{j=1}^4 a_j(0) \exp(\sigma_j t) e_j, \quad (4.13)$$

where the eigenvector e_j may be expressed as

$$e_j = \begin{bmatrix} \sigma_j \\ (\omega_A^2 + \sigma_j^2)/2\Omega \\ i\omega_A \\ -2i\Omega\omega_A\sigma_j/(\omega_A^2 + \sigma_j^2) \end{bmatrix}. \quad (4.14)$$

For any given k_n , it is easy to see that the eigenvector components satisfy the

relationship

$$-\frac{e_j^4}{e_j^1} = \frac{e_j^3}{e_j^2} = \frac{i2\Omega\omega_A}{(\omega_A^2 + \sigma_j^2)}. \quad (4.15)$$

Compact expressions for the eigenvector components allows us to deduce key relationships between the kinetic and magnetic stresses as we show below. We define the mean kinetic (Reynolds) and magnetic (Maxwell) perturbation or fluctuation stresses as

$$R_{ij}(t) = \overline{\delta u_i(z, t)\delta u_j(z, t)}, \quad M_{ij}(t) = \overline{\delta b_i(z, t)\delta b_j(z, t)}, \quad (4.16)$$

where the overbar represents a vertical average which is defined for the product of any two real functions f and g as

$$\overline{fg}(t) \equiv \frac{1}{2H} \int_{-H}^H f(z, t)g(z, t)dz. \quad (4.17)$$

In terms of the Fourier series of the individual velocity and magnetic perturbations, the mean stresses are expressed as (Pessah, Chan, and Psaltis, 2006b)

$$R_{ij}(t) = 2 \sum_{n=1}^{\infty} \text{Re}[\delta \hat{u}_i(k_n, t)\delta \hat{u}_j^*(k_n, t)], \quad (4.18)$$

$$M_{ij}(t) = 2 \sum_{n=1}^{\infty} \text{Re}[\delta \hat{b}_i(k_n, t)\delta \hat{b}_j^*(k_n, t)], \quad (4.19)$$

where the asterisk denotes the complex conjugate.

The off-diagonal Reynolds and Maxwell stress components are given by

$$R_{xy}(t) = 2 \sum_{n=1}^{\infty} \mathcal{R}_{xy}(k_n) \exp(2\sigma t), \quad (4.20)$$

$$M_{xy}(t) = 2 \sum_{n=1}^{\infty} \mathcal{M}_{xy}(k_n) \exp(2\sigma t), \quad (4.21)$$

where

$$\mathcal{R}_{xy} = \frac{\text{Re}[e_{\sigma_j}^1 e_{\sigma_j}^{2*}]}{\|e_{\sigma_j}\|^2}, \quad \mathcal{M}_{xy} = \frac{\text{Re}[e_{\sigma_j}^3 e_{\sigma_j}^{4*}]}{\|e_{\sigma_j}\|^2}. \quad (4.22)$$

4. Non-ideal MRI in Accretion Disks

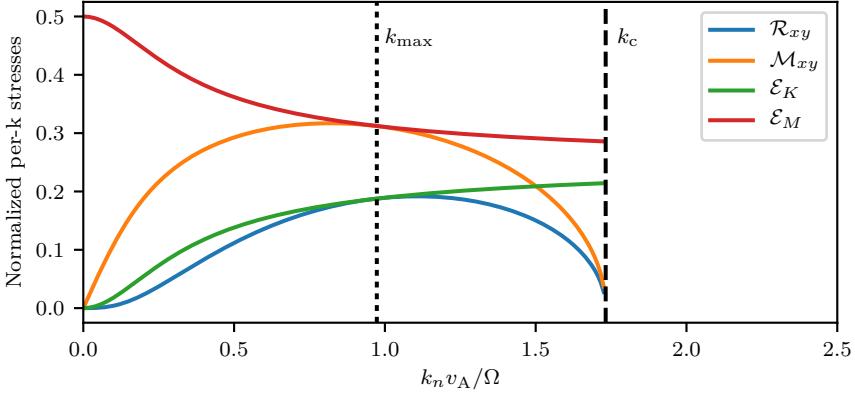


Figure 4.2: The normalized per- k mean kinetic and magnetic stresses and energy densities as a function of wavenumber k_n for the ideal MRI eigenmode.

The trace of the stress tensors yields expressions for the mean kinetic and magnetic energy densities

$$E_K(t) = 2 \sum_{n=1}^{\infty} \mathcal{E}_K(k_n) \exp(2\sigma_j t), \quad (4.23)$$

$$E_M(t) = 2 \sum_{n=1}^{\infty} \mathcal{E}_M(k_n) \exp(2\sigma_j t), \quad (4.24)$$

where

$$\mathcal{E}_K = \frac{\mathcal{R}_{xx} + \mathcal{R}_{yy}}{2}, \quad \mathcal{E}_M = \frac{\mathcal{M}_{xx} + \mathcal{M}_{yy}}{2}. \quad (4.25)$$

At late times, the MRI eigenmode dominates the growth of the perturbations and therefore provides the dominant contribution to the kinetic and magnetic stresses. Thus, the off-diagonal component of the kinetic and magnetic stresses may be expressed as

$$R_{xy}(t) = 2 \sum_{n=1}^{n_c} \mathcal{R}_{xy}(k_n) \exp(2\gamma t) + \dots, \quad (4.26)$$

$$M_{xy}(t) = 2 \sum_{n=1}^{n_c} \mathcal{M}_{xy}(k_n) \exp(2\gamma t) + \dots, \quad (4.27)$$

where the dots represent the negligible contributions from the other three modes and where n_c represents the wavenumber $k_c v_A = \sqrt{2q}\Omega$ at which the MRI is cut-off.

Using the respective eigenvector components in Equation 4.14 yields

$$R_{xy}(t) = \sum_{n=1}^{n_c} \gamma(\gamma^2 + k_n^2) \frac{a_1}{\|e_1\|^2} \exp(2\gamma t) + \dots, \quad (4.28)$$

$$M_{xy}(t) = -4 \sum_{n=1}^{n_c} \frac{\gamma k_n^2}{\gamma^2 + k_n^2} \frac{a_1}{\|e_1\|^2} \exp(2\gamma t) + \dots. \quad (4.29)$$

From Equations 4.28 and 4.29, we can easily see that $R_{xy} > 0$ and $M_{xy} < 0$ for any k_n . As a result, the total mean perturbation stress is also positive, $T_{xy} = R_{xy} - M_{xy} > 0$. This conforms with expectations and would suggest an outward flux of angular momentum under steady state conditions (Balbus and Hawley, 1998; Popham and Narayan, 1991). A property of interest is the ratio of the xy component of the Maxwell to Reynolds stress. At late times, this is dominated by the wavenumber for maximum growth k_{\max} and is

$$\lim_{t \gg 1} \frac{-M_{xy}}{R_{xy}} = -\frac{\text{Re}[e_\gamma^3 e_\gamma^{4*}]}{\text{Re}[e_\gamma^1 e_\gamma^{2*}]} = \frac{4-q}{q} \quad (4.30)$$

Pessah, Chan, and Psaltis (2006b) showed that this relationship carries over, to a large extent, into the nonlinear regime of the MRI.

An interesting feature of the ideal MRI modes considered here, is that they are also exact solutions of the full nonlinear (local) system of equations. This is true of the non-ideal MRI modes that we consider in Paper II as well. Several studies have explored this property in detail and have suggested possible routes to instability saturation (Goodman and Xu, 1994; Latter, Lesaffre, and Balbus, 2009; Pessah, 2010; Pessah and Goodman, 2009). More recently however, Latter, Fromang, and Faure (2015) illuminated the connection between the local ideal MRI modes and their global counterparts. With regard to the local channel modes themselves, Latter, Fromang, and Faure (2015) showed that these modes correspond to the evanescent (decaying) part of a global mode appearing at larger disk radii while a radially varying version of the local mode was associated with the same global

4. Non-ideal MRI in Accretion Disks

mode at smaller radii where the instability showed vigorous activity. A radially varying local mode distinct from the channel mode was associated with the inner rapidly varying part of the global mode however. Latter, Fromang, and Faure (2015) also demonstrated that the global modes, for the most part, did not share the property of being exact nonlinear solutions leading one to strongly suspect that such nonlinear solutions are an artifact of the local approximation. This may quite possibly be the case for the local non-ideal modes as well although a similar study has yet to be performed in this context.

4.3 Non-ideal MHD in Protoplanetary Disks

The plasma conditions in an accretion disk around a compact object are generally favorable to the disk-wide occurrence of the MRI due to the high conductivity of the plasma. However, disks around young and forming stars, called circumstellar disks or protoplanetary disks, present an environment where different diffusive effects play an important role in determining the dynamical evolution of both the gas and the intermixed dust. See Figure 4.3 for a recent observational image of a protoplanetary disk. All three non-ideal effects, i.e., ohmic, Hall and ambipolar diffusion, depend inversely on the fraction of charged particles in the disk fluid, see Equation 3.55. This means that the weaker the ionization of the disk plasma, the stronger the non-ideal effects become. Their different dependencies on the ionization fraction and also on different disk parameters implies that the relative strength of the non-ideal effects varies depending on the location in the disk. Close to the inner edge of the disk, the gas is expected to be sufficiently ionized so that ideal MHD is a good approximation.

Based on simple scaling arguments and crude estimates for the disk parameters such as density, temperature and magnetic field strength, one can form a rough map highlighting the dominant non-ideal effects in the disk. Figure 4.4 presents just such a plot in the parameter space defined by the density n and the magnetic field strength B illustrating where the different non-ideal effects are dominant based on the estimates derived in Armitage (2015). The general picture that

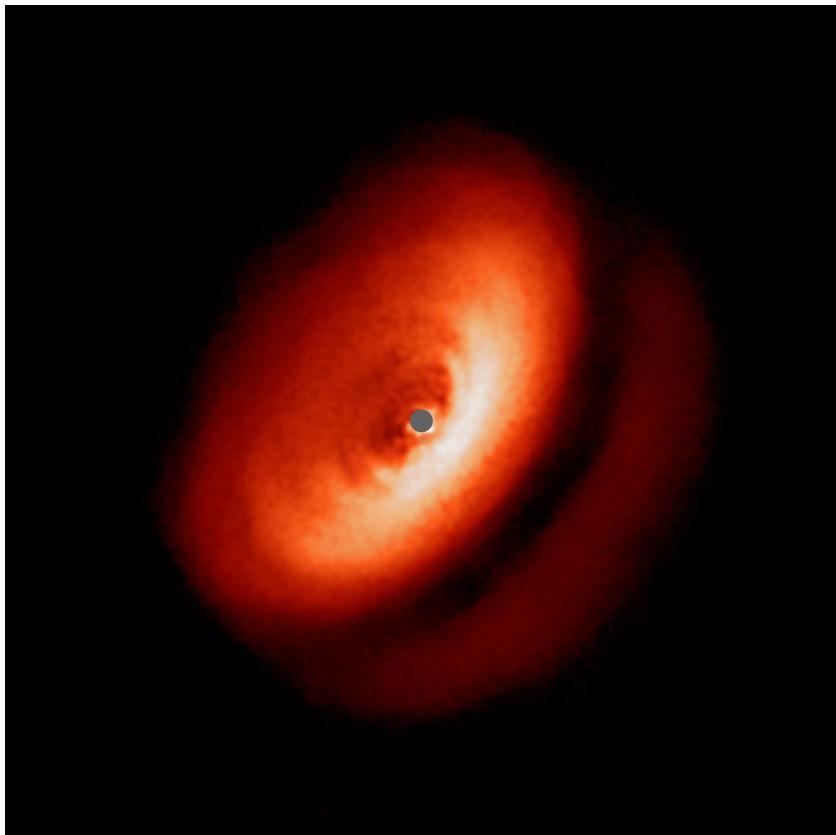


Figure 4.3: Recent observational image of a dusty protoplanetary disk around the young star IM Lupi as seen in reflected light captured by the SPHERE instrument on ESO's Very Large Telescope (VLT). Figure Credit ESO/H. Avenhaus et al./DARTT-S collaboration. See also Avenhaus et al. (2018).

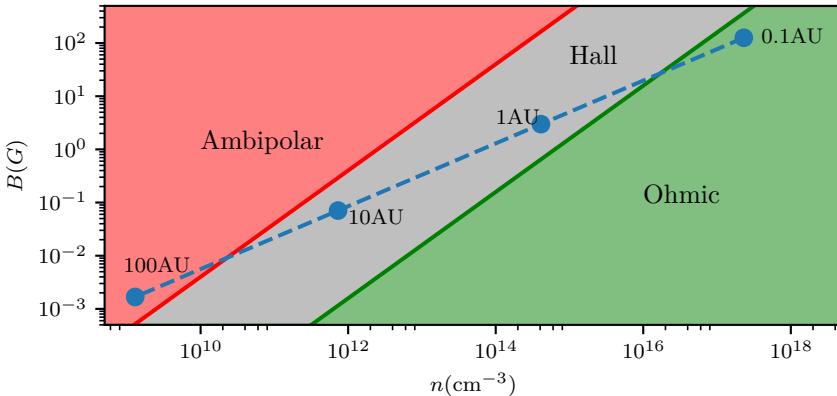


Figure 4.4: A map highlighting the dominant non-ideal effect in a parameter space given by the gas number density n and the magnetic field strength B as derived in Armitage (2015). The boundary between the Hall and ohmic regimes were derived assuming a disk temperature $T = 1000K$. The dashed line tracks the parameter space as one moves from the inner disk at 0.1 AU to the outer disk at 100 AU.

emerges from this simplified calculation is one wherein ohmic diffusion dominates in the inner regions of the disk, Hall diffusion dominates at intermediate distances and ambipolar diffusion dominates in the farther reaches or the low density upper layers of the disk.

Linear analysis also reveals the effect that each non-ideal term has on the MRI. Ohmic diffusion acts as a purely dissipative process. This has the effect of curtailing the range of wavenumbers susceptible to the instability, primarily inhibiting growth at smaller scales and generally reducing the growth rate across the range of unstable wavenumbers. Ambipolar diffusion behaves very much like ohmic diffusion except that it does not dissipate currents parallel to the magnetic field. Such currents are absent for strictly vertical magnetic fields and so the ambipolar diffusion simply contributes to an added dissipation in this setting. On the other hand, if the initial magnetic field has both toroidal and vertical components, am-

bipolar diffusion can act as a destabilizing agent (Desch, 2004; Kunz and Balbus, 2004). However, combined with the dissipative action, the instability is not expected to lead to significant dynamical consequences. Hall diffusion stands out in its capacity to modify the MRI when it comes to non-ideal magnetic diffusivities (Wardle, 1999). Unlike ohmic and ambipolar diffusion, Hall diffusion does not lead to magnetic dissipation. Its effect on the plasma may be best described as being analogous to a magnetic version of the Coriolis force (Balbus and Terquem, 2001). A deciding factor is the relative orientation of the disk angular velocity (or more generally, the disk vorticity (Kunz, 2008) and the direction of the net magnetic field. In the absence of dissipative forces, the Hall diffusion has a markedly different effect on the MRI when the field is aligned to the rotation velocity compared to when it is anti-aligned. In the case of aligned field, the Hall effect shrinks the range of unstable wavenumbers whereas in the case of an anti-aligned field, the Hall effect can render an infinite range of wavenumbers unstable for a narrow range of parameters. Since it is not a dissipative influence, the maximum growth rate attainable is the same as that for ideal MRI with only the Hall effect. In a highly diffusive situation, the Hall effect takes on a character of its own and one finds a destabilizing mode called the Hall-shear instability that operates regardless of rotation (Kunz, 2008; Rüdiger and Kitchatinov, 2005; Wardle and Salmeron, 2012). This mode is best likened to a non-ideal version of Rayleigh's instability, i.e, when $q > 2$ in the absence of magnetic fields (Mohandas and Pessah, 2017).

4.4 Nonlinear evolution of the non-ideal MRI

While linear theory can furnish us with the conditions under which the disk is expected to be stable or unstable and determine the possibility of turbulence, nonlinear numerical simulations are in order to determine the presence and character of the ensuing magnetohydrodynamic turbulence². Nevertheless, linear theory

²Note that the nonlinear regime is not completely outside of the purview of analytical methods. Interesting work in this area include Knobloch and Julien (2005) and Umurhan, Menou, and Regev (2007).

4. Non-ideal MRI in Accretion Disks

has been shown to predict basic features of the saturated state as evidenced by the results in Pessah, Chan, and Psaltis (2006a,b), however, a detailed understanding and exploration of the fully nonlinear state requires numerical simulations.

There are two main approaches to numerical simulations when studying disk dynamics that mirror analytical approaches, namely, local shearing box simulations and global disk simulations. Local shearing box simulations can solve the set of Equations 3.75 - Equations 3.77 with or without i) vertical stratification, ii) a net magnetic flux, iii) non-ideal terms including viscosity and magnetic diffusion. The advantages of the shearing box setup are the high numerical resolutions attainable, speed of computation, and the simplicity in specifying initial or boundary conditions. They cannot however account for curvature effects and radial structure in the background fluid variables. Shearing box simulations have been the prominent workhorse of numerical astrophysicists for studying differentially rotating disks. Ideal MHD simulations in the shearing box (Hawley, Gammie, and Balbus, 1995; Stone et al., 1996) have long since established the possibility of the MRI as the most promising candidate facilitating sustained MHD turbulence sufficient to drive accretion at observed rates. However, issues related to obtaining numerical convergence depending on presence or absence of a net magnetic flux (Davis, Stone, and Pessah, 2010; Fromang and Papaloizou, 2007; Guan et al., 2009; Pessah, Chan, and Psaltis, 2007) as well as on the box size (Shi, Stone, and Huang, 2015) still seek a final resolution.

Inclusion of dissipative effects such as viscosity and resistivity have been relatively straightforward in shearing box simulations (Lesur and Longaretti, 2007) although the issue of physical, as opposed to numerical, convergence arises here as well with respect to the magnetic Prandtl number (Longaretti and Lesur, 2010; Nauman and Pessah, 2016). Guided by expectations of disk structure from models such as that used for Figure 4.4, several shearing box simulations have been performed specifically with the inclusion of one or more of the non-ideal magnetic diffusivities and we refer here to just a few (Bai, 2014a; Bai and Stone, 2011, 2013a; Fleming and Stone, 2003; Hirose and Turner, 2011; Simon et al., 2015).

The first shearing box simulations including Hall diffusion were carried out by Sano and Stone (2002a,b). These simulations did not report any significant modification to the nature of MRI driven turbulence due to the Hall effect. However, Wardle and Salmeron (2012) pointed out that the simulations by Sano and Stone (2002a,b) precludes the strong Hall dominated parameter regime which they argue was more relevant to protoplanetary disks. Recent local simulations (Kunz and Lesur 2013; Lesur, Kunz, and Fromang 2014) have managed to explore this parameter space and found peculiarities in the nonlinear state. While the simulations exhibited the bi-modality of outcomes expected of the Hall effect, the system was observed to settle into a state consisting of long-lived axisymmetric structures called zonal flows in an unstratified disk and one with large toroidal fields and attendant laminar Maxwell stresses with stratification. In either case, turbulence was observed to be too weak to enable significant mass transport. Simulations with a dominant Hall effect are not without difficulties. Numerical stability issues have been a concern in these simulations (Bai, 2014b; Krapp et al., 2018) and so results derived from simulations to date are clearly not the final word on the matter. The dynamical picture as it currently stands is one where MHD turbulence is substantially weakened due to non-ideal effects. This has sparked renewed interest in possible hydrodynamic modes of instability (Barker and Latter, 2015; Klahr and Hubbard, 2014; Lesur and Papaloizou, 2010) and wind-driven angular momentum transport (Bai, 2013; Bai and Stone, 2013b).

As computational power has become more accessible and numerical schemes more efficient, global simulations of accretion disks are now well within grasp. While they are better at capturing large scale effects due to curvature effects and radial stratification, resolution is still limited. Global simulations can now include ohmic diffusion (Dzyurkevich et al., 2010), ohmic and ambipolar diffusion (Gressel et al., 2015), and Hall diffusion (Bai and Stone, 2017; Béthune, Lesur, and Ferreira, 2016; Krapp et al., 2018). These simulations are particularly useful and especially relevant in that they can be used to study the conditions for planet formation and growth (Kley and Nelson, 2012). That said, it is still some time

4. Non-ideal MRI in Accretion Disks

away before we have reliable and realistic full disk simulations at our disposal.

4.5 Paper II - Summary & Prospects

In Paper II, we have undertaken a comprehensive analysis of the linear eigenmodes for the axisymmetric MRI subject to all three non-ideal effects. Our analysis has enabled us to sketch a geometric representation of the eigenmodes, thereby clarifying certain physical characteristics and also elucidating the polarization properties of the oscillatory modes acting particularly under the combined influence of the Coriolis and Hall effects. Using the formalism of Pessah and Chan (2008) and Pessah, Chan, and Psaltis (2006b), we derive closed form expressions for the kinetic and magnetic stresses: the off-diagonal component of the Reynolds and Maxwell stresses and the kinetic and magnetic energy densities associated with the non-ideal MRI in different parts of the parameter space defined by Hall and Pedersen (ohmic + ambipolar) diffusivities.

Our calculations have also enabled us to discern a previously unknown length-scale associated with a characteristic wavenumber that we designate k_{inv} . The wavenumber k_{inv} signals a role reversal in the dominant stress component when the magnetic field and the angular velocity vector are anti-aligned. We find that when the dissipative effects (ohmic and ambipolar) are small, the dominant contribution to the off-diagonal total perturbation stress T_{xy} can be due to correlated velocity perturbations rather than their magnetic counterpart. We have also detected this reversal in well resolved shearing box simulations. While we have not explored the consequences of this feature in the nonlinear regime, we surmise that the nature of the ensuing turbulence would be noticeably different should the Reynolds stress component maintain dominance. This is a prediction that could be tested with local non-ideal shearing box simulations.

The parameter space wherein the Reynolds stress dominates over the Maxwell stress is narrow and may not be widespread within a protoplanetary disk in the current paradigm. Nevertheless, it is an interesting theoretical result that may prove relevant when a better understanding of disk conditions emerge or one that may

find potential application in other astrophysical contexts. Additional avenues to explore would be the connection between local and global non-ideal MRI modes. As mentioned earlier, global simulations, especially those including the Hall effect have only become possible recently. As much of the physics with regard to non-ideal modes have been informed by local calculations, both analytical and numerical; it would be useful to perform a rigorous comparison of the local and global aspects of the non-ideal MRI along the lines of Latter, Fromang, and Faure (2015).

4. Non-ideal MRI in Accretion Disks

Part III

Thermal Structure of Irradiated Atmospheres

Chapter 5

Radiative Transfer

In this chapter, we present a brief introduction to the theory of radiative transfer. As with any of the other subjects dealt with herein, this chapter is only intended to provide an introduction sufficient to follow the technical aspects of the analysis and results presented in Mohandas, Pessah, and Heng (2018) which we shall refer to hereafter as Paper III. For a thorough treatment of the topic, we refer the reader to the classic texts by Chandrasekhar (1960a), Mihalas (1970), and Mihalas and Mihalas (1984). This chapter is largely based on sections of the books by Bodenheimer (2006), Mihalas (1970), and Thomas and Stamnes (2002).

5.1 The Radiative Transfer Equation

The fundamental goal of radiative transfer theory is to describe the flow of radiation and the change in its properties as it interacts with the matter in its path. When we describe radiation here, we generally mean a bundle of rays of light. As we have derived for a gas of massive particles in Chapter 3, here we shall derive an equation that describes the evolution of a “gas” of massless photons starting from a Boltzmann equation for radiation of the form

$$\frac{\partial f}{\partial t} + \dot{\mathbf{x}} \cdot \nabla f + \dot{\mathbf{p}} \cdot \nabla f = \left[\frac{\partial f}{\partial t} \right]_{\text{coll}}. \quad (5.1)$$

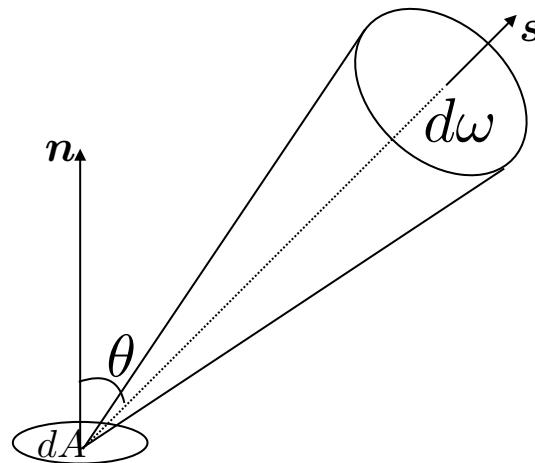


Figure 5.1: A beam of radiation impinging through an infinitesimal surface contained within an infinitesimal solid angle.

Here, f is the photon distribution function such that

$$dN = f(\mathbf{x}, \mathbf{p}, t) d\mathbf{x} d\mathbf{p}, \quad (5.2)$$

represents the number of photons with position between \mathbf{x} and $\mathbf{x} + d\mathbf{x}$ and momentum between \mathbf{p} and $\mathbf{p} + d\mathbf{p}$ at time t . For a system of photons, we have $\dot{x} = c$ where c is the speed of light in the medium. In the absence of general relativistic effects, we may set $\dot{p} = 0$.

Owing to its roots in geometrical optics, the fundamental quantity in radiative transfer theory is the specific intensity. The specific intensity is defined as the amount of energy passing through an infinitesimal area element dA in the time interval dt by rays contained within the solid angle $d\omega$, and in the frequency interval $d\nu$ and is given by

$$dE = I(\mathbf{x}, \mathbf{s}, \nu, t) \mathbf{s} \cdot \mathbf{n} dA d\omega d\nu dt, \quad (5.3)$$

where \mathbf{n} is the unit vector normal to the area element and \mathbf{s} is the unit vector in the direction of the beam of photons. A visual illustration of the concept is provided

in Figure 5.1. How does the specific intensity relate to the photon distribution function? From Equation 5.2, we obtain the total radiant energy contained in the unit phase space volume as

$$\begin{aligned}
 dE &= h\nu f dx dp, \\
 &= h\nu f \mathbf{s} \cdot \mathbf{n}, dA c dt p^2 dp d\omega, \\
 &= h^2 \nu \left(\frac{h\nu}{c} \right)^2 f \mathbf{s} \cdot \mathbf{n} dA dt d\nu d\omega, \\
 &= \frac{h^4 \nu^3}{c^2} f \mathbf{s} \cdot \mathbf{n}, dA dt d\nu d\omega,
 \end{aligned} \tag{5.4}$$

where we have used the relation $p = h\nu/c$. Using Equations 5.4 and 5.3, the Boltzmann equation for the photon gas may be expressed as

$$\frac{\partial I_\nu}{\partial t} + c(\mathbf{s} \cdot \nabla) I_\nu = \left[\frac{\partial I_\nu}{\partial t} \right]_{\text{coll}}. \tag{5.5}$$

where we use the subscript ν to denote frequency dependence of the intensity.

Since the photons do not interact with themselves, the collisional term on the right hand side of Equation 5.5 represents only the interactions of the photons with matter and consists of two types of contributions. The first denotes the loss of radiant intensity from its original trajectory and the second denotes in the addition of radiant intensity to the original beam. We describe the two types of interaction processes in the next section.

5.2 Extinction: Absorption and Scattering

When a beam of radiation encounters matter in its path, the resulting interaction can lead to the removal of some photons from the original beam of radiation. This may happen either through the process of absorption or the process of scattering. Absorption takes place when some of the photons in the beam are lost to the thermal energy of the intervening matter. Scattering occurs when the photon is redirected away from its original path leading to a reduction in energy from the incident beam of radiation. The resulting change in energy or intensity due to

5. Radiative Transfer

absorption and/or scattering is expressed as

$$\delta(dE_\nu) = \delta I_\nu dA dt d\nu d\omega, \quad (5.6)$$

One expects that the intensity of the radiation after absorption or scattering is diminished by an amount proportional to the incident beam intensity and the amount of material present. The quantitative expression of this empirical result is known as Beer's law and is given by

$$\delta I_\nu = -\rho k_\nu I_\nu \delta x \quad (5.7)$$

where

$$k_\nu = \kappa_\nu + \sigma_\nu \quad (5.8)$$

and ρ is the density of the absorbing or scattering matter along the path of the radiant beam under consideration. Here, κ_ν and σ_ν are the absorption and scattering opacities respectively and k_ν is referred to as the extinction opacity.

5.3 Emission: Thermal Radiation and Scattering

The other major contribution to the collisional term comes about when matter in the path of a given beam of radiation adds energy to the beam. The processes by which this occurs may be thought as being reciprocal to extinction processes and is termed emission. Emission may be due to thermal radiation or scattering processes and the resulting energy enhancement can also be generally expressed by Equation 5.6. However, the increase in specific intensity due to emission is independent of the intensity of the incident beam. Therefore

$$\delta I_\nu = -\rho j_\nu \delta x, \quad (5.9)$$

where j_ν is called the emissivity.

If the material is in thermodynamic equilibrium with the surrounding radiation field, then according to the Kirchhoff-Planck law, as much radiation is released as is absorbed.

$$j_\nu^t = \kappa_\nu B_\nu(T), \quad (5.10)$$

where j_ν^t is the thermal emissivity and B_ν is the Planck function. In reality, thermodynamic equilibrium is a poor approximation in any appreciable astrophysical domain consisting of interacting matter and photons. However, for an arbitrarily small volume, one may assume thermodynamic equilibrium to hold locally and therefore apply a localized version of the Kirchhoff-Planck law. Such an approximation is commonly referred to as Local Thermodynamic Equilibrium or LTE.

The other possible means by which radiant intensity in a beam is enhanced is when radiation that is originally traveling in another direction is scattered into the solid angle containing the beam under consideration. Scattering is generally described in terms of a redistribution function $R(\nu', \mathbf{n}'; \nu, \mathbf{n})$, which gives the probability that a photon with frequency in the range $(\nu', \nu' + d\nu')$ and within the solid angle $d\omega'$ is scattered into the solid angle $d\omega$ with a different frequency in the range $(\nu, \nu + d\nu)$.

If the frequency of the photon changes as it is scattered, we refer to scattering as being non-coherent whereas if the frequency of the scattered photon is unchanged, we refer to it as coherent scattering. The energy scattered by the material into the beam of radiation is given by

$$\delta(dE) = \sigma_\nu d\nu d\omega \frac{1}{4\pi} \oint \int_0^\infty I_{\nu'} R(\nu', \mathbf{s}'; \nu, \mathbf{s}) d\nu' d\omega'. \quad (5.11)$$

In the case of coherent scattering, the redistribution function reduces to

$$R(\nu', \mathbf{s}'; \nu, \mathbf{s}) = \mathcal{P}(\mathbf{s}', \mathbf{s}) \delta(\nu - \nu'), \quad (5.12)$$

where \mathcal{P} is the scattering phase function. In the case of isotropic scattering, the phase function is trivial, $\mathcal{P} = 1$ and has no dependence on the angle between scattered and emergent beam. The emissivity due to scattering is then

$$j_\nu^s = \sigma_\nu J_\nu. \quad (5.13)$$

Other analytic scattering phase functions include the dipole phase function (Mihalas, 1970)

$$\mathcal{P} = \frac{3}{4}(1 + \cos^2 \Theta), \quad (5.14)$$

5. Radiative Transfer

and the Henyey-Greenstein phase function (Seager, 2010)

$$\mathcal{P} = \frac{1 - g^2}{(1 + g^2 - 2g \cos \Theta)^{1/2}}, \quad (5.15)$$

where $\Theta = \arccos(\mathbf{s}', \mathbf{s})$ and g (not to be confused with gravity) is the anisotropy parameter. The anisotropy parameter occupies the range $-1 \leq g \leq 1$ such that $g = -1$ refers to backward scattering, $g = 1$ refers to forward scattering and $g = 0$ refers to isotropic scattering. We will only consider the simple case of coherent isotropic scattering in Paper III.

Thus, the total emissivity containing contributions from both thermal emission and scattering is given by $j_\nu = j_\nu^t + j_\nu^s$.

5.4 Moments of the Transfer Equation

With the general form of the collisional contributions as given above, the radiative transfer equation may be expressed as (Bodenheimer, 2006)

$$\begin{aligned} \frac{1}{c} \frac{\partial I_\nu}{\partial t} + \mathbf{s} \cdot \nabla I_\nu &= -\rho k I_\nu + \rho j_\nu, \\ &= -\rho k_\nu (I_\nu - S_\nu), \end{aligned} \quad (5.16)$$

where we define the source function, $S_\nu = j_\nu / k_\nu$. In principle, one need only solve the above differential equation for the intensity to obtain a complete solution to the problem of radiative transfer. This, however, requires knowledge of the opacities and the source function together with the required number of initial and boundary conditions. In most applications, the opacities and emissivity are functions of temperature, composition, frequency and density which are in turn functions of space and time. While the opacities themselves are rarely ever dependent on the intensity of the impinging beam, the source function is inextricably tied to the incident intensity if scattering is present as given by Equation 5.11. This complicates the process of obtaining a solution to the transfer problem as one now has to deal with an integro-differential equation. In steady-state, one can

write down an expression constituting a formal solution of the transfer equation (Mihalas, 1970; Thomas and Stammes, 2002).

Alternatively, one may choose to solve a system of equations involving the transfer equation and its moments. We define below the moments of the specific intensity as (Bodenheimer, 2006)

$$(J_\nu, \mathbf{H}_\nu, \mathbf{K}_\nu) \equiv \frac{1}{4\pi} \oint (1, \mathbf{s}, \mathbf{ss}) I_\nu d\omega, \quad (5.17)$$

where J_ν , \mathbf{H}_ν and \mathbf{K}_ν are the mean intensity, the Eddington flux and the K integral respectively. Integrating Equation 5.16 over all solid angles, we obtain the first moment equation

$$\frac{1}{c} \frac{\partial J_\nu}{\partial t} + \nabla \cdot \mathbf{H}_\nu = -\frac{1}{4\pi} \oint \rho k_\nu (I_\nu - S_\nu) d\omega. \quad (5.18)$$

Multiplying Equation 5.18 with s_i and integrating over all solid angles, we obtain the second moment equation

$$\frac{1}{c} \frac{\partial \mathbf{H}_\nu}{\partial t} + \nabla \cdot \mathbf{K}_\nu = -\frac{1}{4\pi} \oint s \rho k_\nu (I_\nu - S_\nu) d\omega, \quad (5.19)$$

where x_j represents the components of an orthogonal coordinate system. The K -integral is a symmetric tensor with $\text{tr} \mathbf{K}_\nu = J_\nu$ since $\sum \mathbf{ss} = 1$. Equations 5.18 and 5.19 constitute four equations in nine unknowns. We therefore require five additional equations to close the system. These take the form of the so-called closure relations

$$\mathbf{K}_\nu = \mathbf{f}_\nu J_\nu, \quad (5.20)$$

where \mathbf{f} represents the matrix of the so-called Eddington coefficients and are symmetric with unit trace. The values assigned to the Eddington coefficients are usually based on considerations of geometry and symmetry.

5.5 The Flux

The most relevant quantity with regard to astrophysical observations is the flux. The flux as defined by Equation 5.17 is a vector. However, because virtually all

5. Radiative Transfer

the flux that we collect from astronomical bodies passes nearly parallel through a very small (relative to the size of the source and the distance to it) plane area, we can orient our coordinate system in such a way that we measure only a scalar component. This scalar flux is given by

$$F_\nu = \mathbf{H}_\nu \cdot \mathbf{e}_z = \oint I_\nu(z, \theta, \varphi) \cos \theta d\omega, \quad (5.21)$$

$$= 2\pi \int_{-1}^1 I_\nu(z, \mu) \mu d\mu, \quad (5.22)$$

where $\mu = \cos \theta$. In the literature, one often finds an equivalent quantity called the *astrophysical flux* defined as $\mathcal{F}_\nu = \frac{F_\nu}{\pi}$. A basic first approximation used in modeling stars or planets is to assume that they radiate as a uniformly bright sphere. Let us calculate the flux received from a uniformly bright sphere of radius R at a distance D from the detector. A uniformly bright sphere has an isotropic intensity emanating from its surface, that is the intensity is the same in all directions. The flux received is then

$$F_\nu = 2\pi \int_{\cos^{-1}\left(\frac{\sqrt{D^2 - R^2}}{D}\right)}^1 I_S \mu d\mu \quad (5.23)$$

$$= \pi I_S \left(\frac{R}{D}\right)^2 = F_S \left(\frac{R}{D}\right)^2 \quad (5.24)$$

where $F_S = \pi I_S$ where I_S and F_S are the specific intensity and the associated flux (magnitude) at the surface of the bright sphere.

We can define an equivalent temperature called the effective temperature T_{eff} that is related to the flux by considering a blackbody of the same shape with the same total flux. Therefore,

$$F_\nu = 2\pi \int_{-1}^1 I_\nu(\mu) \mu d\mu = 2\pi B_\nu(T) \int_0^1 \mu d\mu = \pi B_\nu(T) \quad (5.25)$$

On integrating across all frequencies, we obtain

$$F = \int_0^\infty F_\nu d\nu = \pi \int_0^\infty B_\nu(T) d\nu = \sigma T_{\text{eff}}^4 \quad (5.26)$$

5.6 Model Atmospheres

The development of radiative transfer as a rigorous discipline was driven by the efforts to devise a model for the atmospheres of distant stars and planets. The construction of such model atmospheres allows us to determine atmospheric conditions - the temperatures, pressures and composition, of stellar or sub-stellar objects by examining their spectrum. In the most general circumstance, this is an enormously complicated enterprise even with the latest numerical techniques available to us today. Deriving purely analytical models is a difficult to accomplish except under a number of simplifying assumptions. We consider here a common set of assumptions that usually make the problem analytically tractable (Mihalas, 1970)

- Homogeneous plane-parallel atmosphere: This is an assumption by which one can reduce the dimensionality of the problem. Here, we treat the atmosphere as a one dimensional slab where the physical variables only exhibit non-negligible changes vertically and are uniform horizontally. The one-dimensional approximation is a reasonable assumption due to the fact that most atmospheric domains under consideration have a horizontal extent that is many times larger than their vertical extent. In other words, atmospheres generally possess a small aspect ratio $H/L \ll 1$ where H is the height of the atmosphere and L the horizontal extent usually of the order of the radius of the star or planet in question.
- Statistical, hydrostatic and radiative equilibrium: This is an assumption of steady-state conditions. Most astronomical objects that we observe are assumed to be in a state of equilibrium. We expect statistical equilibrium in the sense that the microphysical processes that populate and depopulate atomic levels are in balance. Hydrostatic equilibrium implies that pressure gradients balance gravity with negligible vertical accelerations whereas radiative equilibrium assumes negligible transport of energy due to convective or other hydrodynamical motions.

5.7 Radiative Equilibrium

The atmospheres of most stars and planets transport energy from the base of the atmosphere where it emanates to the top of the atmosphere where it is eventually lost to space as radiation. The general expectation is that on a global average sense, a balance is attained between the incoming energy from the insolation and the energy emitted and reradiated by the planet.

The two main mechanisms of energy transport in the atmosphere are radiative and convective transport. Radiative equilibrium arises when transport is entirely due to radiation and convective equilibrium comes about when transport is due to convective motions alone. The atmospheres of both stars and planets may have regions unstable to convection (Mihalas, 1970; Seager, 2010). Typically, these correspond to deeper regions close to the base with high pressures and high enough opacities. Convective instability is usually determined by Schwarzschild's criterion (Mihalas, 1970; Pringle and King, 2007).

A mathematical expression for radiative equilibrium is derived by equating the total energy removed against the total energy added, integrated across all frequencies and all solid angles. This implies

$$\int_0^\infty d\nu \oint k_\nu I_\nu d\omega = \int_0^\infty d\nu \oint j_\nu d\omega, \quad (5.27)$$

$$\int_0^\infty d\nu k_\nu J_\nu = \int_0^\infty d\nu k_\nu S_\nu. \quad (5.28)$$

Equivalently, the condition of radiative equilibrium may be expressed by the frequency integrated left hand side of the transfer equation, Equation 5.16 in the steady-state

$$\int_0^\infty \nabla \cdot \mathbf{H}_\nu d\nu = 0. \quad (5.29)$$

This implies that the flux is divergenceless. In a planar atmosphere, this reduces to

$$\frac{dH}{dz} = 0, \quad (5.30)$$

where z is the vertical coordinate and $H = \int H_\nu d\nu$.

5.8 Optical Depth and Column Mass

A natural measure of distance when dealing with radiative transport is optical depth which is defined here as

$$d\tau \equiv -\rho k_\nu dx. \quad (5.31)$$

There are three things to note in this definition. Firstly, the optical depth is a dimensionless measure of distance. We will shortly clarify what we mean by distance in this context. Secondly, the popular convention is such that the optical depth increases in a direction opposite to the real spatial extent when considering a one dimensional planar atmosphere. Finally, the inclusion of density in Equation 5.31 implies that unlike physical distance, the optical depth measures distance in relation to the amount of material in the path of the beam. Therefore, the optical depth is larger in a region with a greater concentration of absorbing matter than where it is not. The mean free path of photons is much less in a denser region and this effectively increases the distance a photon has to travel to emerge unscathed. In this sense, the optical depth is larger, $\tau \gg 1$, in the interior of stars, smaller in interstellar space, $\tau \ll 1$, and of order unity, $\tau \sim 1$, at the stellar photosphere (what we perceive as the surface of the star).

A related quantity that we shall use is the column mass defined here as

$$dm \equiv -\rho dx. \quad (5.32)$$

The main advantage of the column mass variable is that it allows us to derive a simple relation connecting atmospheric depth to atmospheric pressure via the hydrostatic equilibrium condition (assuming constant gravity) $P = mg$. This allows us to express the temperature as a function of pressure.

5.9 Gray Approximation and Mean Opacities

The fundamental assumption in the gray approximation approach to the solution of the transfer equation is in treating the opacity as being independent of frequency. In reality, opacities exhibit substantial frequency dependence and as such

5. Radiative Transfer

the gray approximation may be severely limited in parts of the atmosphere. To be clear, the gray approximation does not posit that one assume an ad-hoc constant value of opacity. Instead, a suitably averaged opacity is considered that substitutes as reasonable approximation in certain parts of the atmosphere. This affords one the luxury to exactly solve the transfer equation and construct an analytical thermal profile.

The choices for the mean opacities differ with respect to the choices for the weight function used and the mode of averaging. Conventionally, the choice of mean opacities are motivated with a view to preserve the form of one of the moment equations. Two such mean opacities are

1. Flux-Weighted Mean:

$$\kappa_H \equiv \frac{\int_0^\infty \kappa_\nu H_\nu d\nu}{\int_0^\infty H_\nu d\nu}. \quad (5.33)$$

2. Absorption Mean:

$$\kappa_J \equiv \frac{\int_0^\infty \kappa_\nu J_\nu d\nu}{\int_0^\infty J_\nu d\nu}. \quad (5.34)$$

The flux-weighted and absorption mean opacities are simple arithmetic averages over the moments of the intensity function. The trouble with using these opacities is that we do not have prior knowledge of the intensity moments, indeed, this is what we set out to solve for in the first place. Unless, further assumptions can be made, the utility of the above choices are limited.

Two popular choices of mean opacities that overcome this difficulty are introduced below

1. Rosseland Mean:

$$\frac{1}{\kappa_R} \equiv \frac{\int_0^\infty \frac{1}{\kappa_\nu} \frac{dK_\nu}{dz} d\nu}{\int_0^\infty \frac{dK_\nu}{dz} d\nu} = \frac{\int_0^\infty \frac{1}{\kappa_\nu} \frac{dB_\nu}{dT} d\nu}{\int_0^\infty \frac{dB_\nu}{dT} d\nu}, \quad (5.35)$$

where in deriving the second equality, we have used the Eddington approximation so that $K_\nu = J_\nu/3$ and $J_\nu = B_\nu$. The Eddington approximation is valid in the limit of large optical depths where the radiation field is effectively isotropic. As the Rosseland mean is a harmonic average, it attaches greater weight to those regions where the opacity is lowest.

2. Planck Mean:

$$\kappa_P \equiv \frac{\int_0^\infty \kappa_\nu B_\nu d\nu}{\int_0^\infty B_\nu d\nu}, \quad (5.36)$$

This is a mean opacity that is easily calculable and does not require prior solution of the radiative transfer equation. It is defined so that it yields the correct value of the integrated thermal emission. It is most relevant in optically thin regions which generally correspond to the upper regions in a one dimensional model atmosphere.

5.10 Gray Atmosphere Model

We now use the machinery laid out in the previous sections to construct a model atmosphere in the gray approximation. Consider a plane-parallel homogeneous atmosphere under equilibrium conditions. The equation of transfer is given by

$$\mu \frac{dI_\nu}{dm} = k_\nu(I_\nu - S_\nu). \quad (5.37)$$

The source function, assuming LTE and coherent isotropic scattering, is given by

$$S_\nu = \kappa_\nu B_\nu + \sigma_\nu J_\nu. \quad (5.38)$$

The first moment equation is

$$\frac{dH_\nu}{dm} = k_\nu(J_\nu - S_\nu), \quad (5.39)$$

$$= \kappa_\nu(J_\nu - B_\nu). \quad (5.40)$$

5. Radiative Transfer

Integrating over all frequencies, this becomes

$$\frac{dH}{dm} = \kappa_J J - \kappa_P B, \quad (5.41)$$

where J and B are the frequency integrated mean intensity and the blackbody function respectively. The radiative equilibrium equilibrium under the gray approximation becomes

$$\kappa_J J = \kappa_P B. \quad (5.42)$$

If we make the additional assumption that $\kappa_J = \kappa_P$, we get the simple relation $J = B$.

The second moment equation is given by

$$\frac{dK_\nu}{dm} = k_\nu H_\nu, \quad (5.43)$$

which upon integrating across all frequencies yields

$$\frac{dK}{dm} = \kappa_R H. \quad (5.44)$$

Since H is a constant according to the radiative equilibrium condition, we can readily solve Equation 5.44 to obtain

$$K = \kappa_R H \tau + c, \quad (5.45)$$

where we have changed the independent variable to optical depth $d\tau = \kappa_R dm$ here defined in terms of the Rosseland mean opacity κ_R . To proceed we need to connect the moments of the specific intensity to temperature. We invoke the standard Eddington approximation whereby $J = 3K$ and use $J = B$. We also relate the Eddington flux to the physical flux $\mathcal{F} = 4H$ such that Equation 5.45 becomes

$$B = \frac{3}{4} \mathcal{F} \tau + 3c. \quad (5.46)$$

Using the formal solution of the radiative transfer equation for a semi-infinite planar atmosphere, the constant factor can be calculated (see Mihalas (1970)) and is given by $c = \mathcal{F}/6$. Substituting the fluxes with effective temperatures, we obtain

$$T^4 = \frac{3}{4} T_{\text{eff}}^4 \left(\tau + \frac{2}{3} \right), \quad (5.47)$$

where T gives the full run of temperatures with depth in terms of the effective temperature associated with the internal flux or energy emanating from the bottom of the atmosphere.

5. Radiative Transfer

Chapter 6

Analytical Models of Irradiated Atmospheres

This chapter contains a summary of the analysis carried out in Paper III. We begin by motivating the chief astrophysical application for the model developed in Paper III, namely exoplanetary atmospheres, and briefly describe the main methods by which they are detected and characterized. This is followed by a derivation of a basic two-step gray atmospheric model of irradiated atmospheres. We close with a summary discussion our extension to this rudimentary model derived in Paper III.

6.1 Exoplanets

The age of extra-solar planets or exoplanets was heralded with the discovery by Mayor and Queloz (1995) of the first planet outside of the solar system, designated 51 Pegasi b, orbiting a main-sequence star, 51 Pegasi. Since then, there has been an explosion in the number of exoplanets discovered with around 3000 confirmed planets at the time of writing, see exoplanets.org. That number is set to increase manifold in the near future with new space and ground based facilities. We have reached the point in the human scientific enterprise wherein an answer to the

6. Analytical Models of Irradiated Atmospheres

age-old question of "are we alone in the universe?" is nearly within our grasp. A crucial step in arriving at an affirmative answer to the presence of Earth-like life¹ is to understand the properties of its atmosphere. Although the reliable detection of an Earth-analog or Earth-twin is one of the main drivers of exoplanet science, understanding the physical conditions that prevail among the wide diversity of exoplanets discovered is of enormous scientific value in and of itself.

6.2 Exoplanetary Atmospheres

Our primary window to the atmospheres of exoplanets are the spectral features that we can extract from the light that is collected. There are two prominent methods by which we obtain such spectral information - transit spectroscopy and direct imaging (Madhusudhan et al., 2014). The transit method involves the observations of those exoplanets that cross their host star's disk orthogonal to the line of sight connecting the star to the detector. Depending on the orbital phase of the transiting exoplanet, the spectra that we obtain can be further categorized into three kinds - transmission spectra, thermal spectra and thermal phase curves (Madhusudhan et al., 2014).

We obtain the transmission spectrum as the planet passes in front of and cuts through the disk of the host star end to end. This phase of the planet's orbit is often referred to as the primary eclipse and degeneracy free observations are obtained when the planet is on a non-inclined orbit. As the planet passes through the disk of the star, some of the starlight is blocked by the disk of the planet and there is a dip in brightness and the flux received as a result. The depth of this dip is proportional to the ratio of the size of the planet to the star. However, if the planet has a substantial atmosphere, the light of the star is diminished by differing amounts at different wavelengths depending on the composition and concentration of absorbers present in the atmosphere. Transmission spectrum typically

¹While this is a restrictive perception of the nature of life, it is the only one we have to work with.

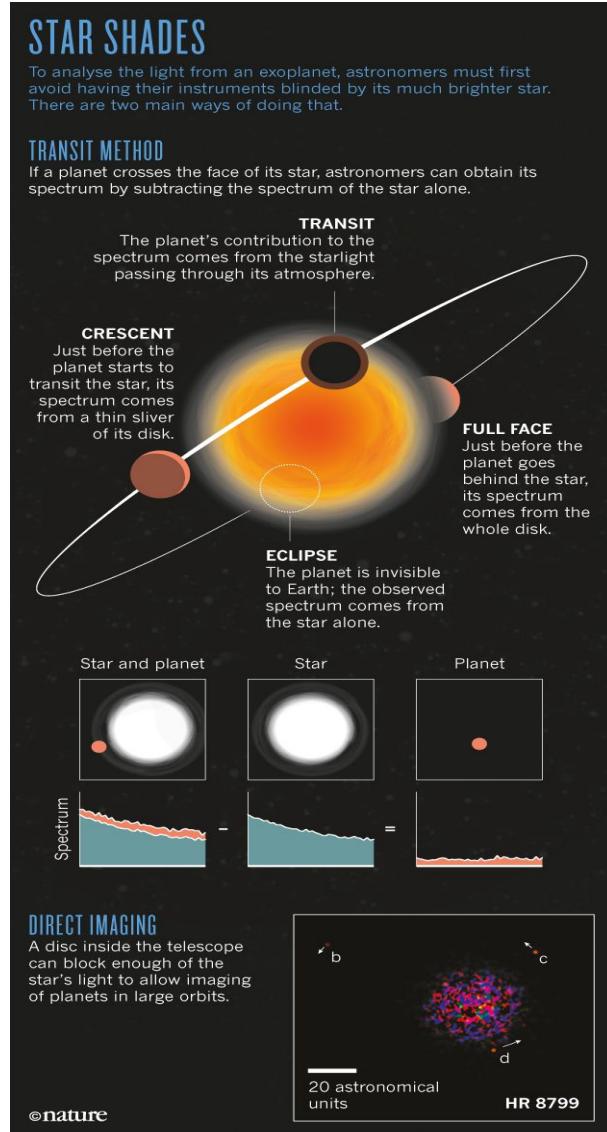


Figure 6.1: A cartoon representation of the two main methods of detecting and observing exoplanets. Figure reprinted from Hecht (2016), with permission from Springer Nature.

6. Analytical Models of Irradiated Atmospheres

captures information relating to the atmospheric properties at the day-night terminator of the transiting planet.

As the planet slips behind its host star, there is again a smaller but measurable dip in the flux observed. This phase of the planet's orbit is commonly referred to as the secondary eclipse and yields information about the day-side composition and temperature pressure profile. This dip is greatest in the near-to-mid IR wavelength range as this is the wavelength band in which the planet's emission is most intense and where the stellar emission is expected to be relatively lower in comparison. Thermal spectra yields information about the day-side atmospheric features of the planet.

The transmission spectroscopy method of exoplanet observations is most sensitive to closely orbiting giant planets. For this reason, such planets are the most amenable to atmospheric modeling. These planets are also expected to be tidally locked to their host star with permanent day sides and night sides. This in turn allows us to extract longitudinal information about the planet's atmosphere as it reveals different phases during its orbit. Observations from different phases also enables us to extract information about the atmospheric dynamics. The presence of super-rotating equatorial winds were confirmed by using global circulation models in conjunction with observations of hot and cold spots on the planet shifted from its sub-stellar and anti-stellar points (Showman, Cho, and Menou, 2010).

Direct imaging of planets is a more challenging endeavor. It has been possible however to achieve this for a smaller set of planets and is another useful method by which we can study the planet's atmosphere. The planets that are best suited for direct imaging are young massive planets that are widely separated from their host star. This is because younger and massive planets are likely to be hotter and brighter (Madhusudhan et al., 2014). While the reach of direct imaging will extend to relatively lower mass and older planets with the advancement of newer ground based and space based instruments, currently transit spectroscopy remains the primary workhorse of exoplanet atmosphere observations.

6.3 Gray Atmosphere Model for Irradiated Atmospheres

We now present the simplest analytical model of an irradiated atmosphere by using the gray approximation. In this scenario, a complex interplay of absorption, re-emission and reflection lead to energy balance between external radiative forcing and the emergent internal radiation. We will construct a very basic model that describes the vertical thermal structure of such an atmosphere. A key assumption is the separation in wavelength or frequency of the external and internal radiation. This separation in frequency is a good assumption for Earth and distant planets in general but becomes questionable the closer the planet is to the star. Nevertheless, it serves as a useful first approximation. This also enables us to carry forward the calculation in parallel on two fronts: the shortwave set of equations connected to irradiation and the longwave set of equations connected to internal radiation. We then combine the two using the principle of energy conservation or radiative equilibrium to arrive at a complete solution.

For simplicity, we also assume only absorption in the derivation. The moments of the radiative transfer equation in the longwave band denoted by the subscript L are

$$\frac{dH_L}{d\tau} = J_L - \lambda \mathcal{B}, \quad (6.1)$$

and

$$\frac{dK_S}{d\tau} = H_L, \quad (6.2)$$

where the optical depth is defined as $d\tau = \kappa_L dm$. Here, $\lambda = \kappa_P / \kappa_L$ where κ_P is the Planck mean opacity and κ_L is the absorption mean opacity in the longwave band. The moments of the radiative transfer equation in the shortwave band denoted by the subscript S are

$$\frac{dH_S}{d\tau} = \gamma J_S, \quad (6.3)$$

and

$$\frac{dK_S}{d\tau} = \gamma H_S, \quad (6.4)$$

where we have assumed negligible thermal emission in the shortwave and have also defined $\gamma = \kappa_S / \kappa_L$. Using the closure relation $K_S = \bar{\mu}^2 J_S$ (Guillot, 2010;

6. Analytical Models of Irradiated Atmospheres

Heng, Mendonça, and Lee, 2014) where $\bar{\mu}$ is the angle of the collimated irradiation beam from the star, we combine the two first order moment equations into the single second order equation

$$\frac{d^2 J_S}{dm^2} = \gamma_\mu^2 J_S, \quad (6.5)$$

which has the solution

$$J_S = J_{S0} \exp(-\gamma_\mu \tau), \quad (6.6)$$

where $\gamma_\mu = \gamma/|\bar{\mu}|$ and $\bar{\mu} < 0$. We also obtain the Eddington flux in the shortwave as $H_S = -|\bar{\mu}| J_S$. The radiative equilibrium condition is given by

$$\frac{d\mathcal{H}}{d\tau} = J_L - \gamma_\mu H_S - \lambda \mathcal{B} = 0, \quad (6.7)$$

where $\mathcal{H} = H_L + H_S$ is the total Eddington flux. We readily obtain the solution to Equation 6.2 in terms of the fluxes as

$$K_L = \mathcal{H}\tau + c + \frac{H_S}{\gamma_\mu}. \quad (6.8)$$

Using the Eddington approximation to the right hand side of Equation 6.7 as well as in Equation 6.8, we have

$$\mathcal{B} = \frac{3}{\gamma_\mu} \left[\mathcal{H}\tau + c + \left(\frac{1}{\gamma_\mu} - \frac{\gamma_\mu}{3} \right) H_S \right]. \quad (6.9)$$

It only remains to determine the constant of integration c to obtain the complete solution. We may use the same procedure as in the non-irradiated gray case to obtain

$$c = \frac{2}{3} \mathcal{H} - \left(\frac{2}{3} + \frac{1}{\gamma_\mu} \right) H_S(0). \quad (6.10)$$

Substituting Equation 6.10 and replacing the Eddington fluxes with the internal and irradiation astrophysical fluxes using the relations $\mathcal{F} = 4\mathcal{H}$ and $\mathcal{F}_*(0) = 4H_S(0)$, we get

$$\mathcal{B} = \frac{3}{4} \frac{\mathcal{F}}{\lambda} \left(\tau + \frac{2}{3} \right) - \frac{3}{4} \frac{\mathcal{F}_*(0)}{\lambda} \left[\frac{2}{3} + \frac{1}{\gamma_\mu} + \left(\frac{\gamma_\mu}{3} - \frac{1}{\gamma_\mu} \right) \exp(-\gamma_\mu \tau) \right]. \quad (6.11)$$

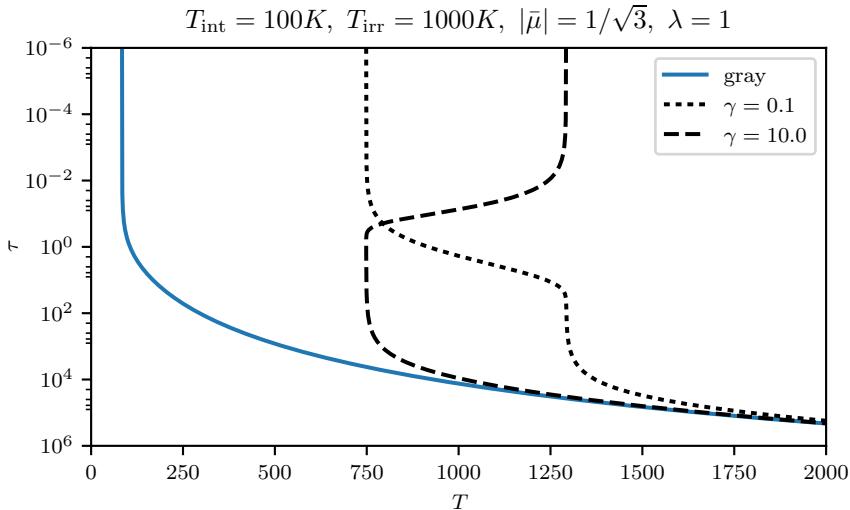


Figure 6.2: The temperature profile computed for a non-irradiated gray atmosphere (blue curve) and an irradiated dual-band gray atmosphere (black curves) for two different strengths of shortwave absorption opacity.

Finally, substituting the Planck function and the astrophysical fluxes with the respective effective temperatures yields the temperature profile

$$T^4 = \frac{3}{4} \frac{T_{\text{int}}^4}{\lambda} \left(\tau + \frac{2}{3} \right) - \frac{3}{4} \frac{|\mu| T_{\text{irr}}^4}{\lambda} \left[\frac{2}{3} + \frac{1}{\gamma_\mu} + \left(\frac{\gamma_\mu}{3} - \frac{1}{\gamma_\mu} \right) \exp(-\gamma_\mu \tau) \right]. \quad (6.12)$$

Figure 6.2 shows the temperature profile for an irradiated atmosphere as a function of optical depth given by Equation 6.12 and for a non-irradiated atmosphere as given by Equation 5.47.

6.4 Numerical Modeling of Irradiated Atmospheres

We briefly mention the present state of affairs with regard to numerical efforts to model irradiated atmospheres with applications to exoplanets. There are currently two ways of numerically modeling irradiated atmospheres: the *forward* modeling technique and the *atmospheric retrieval* method (Hubeny, 2017). In the former

6. Analytical Models of Irradiated Atmospheres

method, the objective is to self-consistently solve the equations of radiative transfer by adopting suitable physical prescriptions for atomic and molecular abundances, their opacities as well as cloud and haze opacities, together with considerations of equilibrium, in order to predict the temperature structure and atmospheric spectrum of the planet. In the latter method, one solves the radiative transfer equations with the exception that the temperature and pressure profiles are treated as free parameters along with atomic and molecular abundances and is used in conjunction with statistical algorithms to simultaneously retrieve the temperature profile and chemical composition of a planetary atmosphere from observational data. A few numerical codes that perform retrieval computations are NEMESIS (Irwin et al., 2008), CHIMERA (Line et al., 2013), TAU-REX (Waldmann et al., 2015). Some numerical codes that perform forward modeling include, but are not limited to, COOLTLUSTY (Hubeny, Burrows, and Sudarsky, 2003), EXO-REM (Goukenleuque et al., 2000) and HELIOS (Malik et al., 2017).

In addition to the forward modeling and retrieval techniques, there is an additional duality of approaches to modeling exoplanetary or, more generally, sub-stellar mass objects (SMO) atmospheres. By adapting either, i) a stellar atmosphere code, or ii) a solar system planetary atmosphere code, one can take advantage of the extensive historical development in modeling stars and solar system planets. (Hubeny, 2017). Both approaches offer their unique set of advantages as well as challenges. Stellar atmosphere codes have a long history of development and have achieved great levels of sophistication. They are capable of incorporating NLTE effects which are, however, not very important in the context of planetary atmospheres. Moreover, spectral lines in SMO's tend to be mostly organized in bands and are of molecular origin. Adapting stellar atmosphere codes therefore boils down to removing procedures for dealing with NLTE, and implementing opacities obtained from pre-calculated tables. On the other hand, one has to include a reliable treatment of cloud formation and also methods to determine concentrations of molecular species as functions of temperature and pressure.

Atmospheric modeling codes that were originally used for modeling solar

system planet (or satellite) atmospheres also presents a natural choice of a tool that could be adapted to model exoplanets. However, most exoplanets that are amenable to atmospheric characterization are closely orbiting highly irradiated exoplanets. As these are planets for which a solar system analog is unavailable, codes used to model solar-system planets would have to be adjusted accordingly.b

6.5 Paper III - Summary & Prospects

In this paper, we extend the analytical treatment of modeling exoplanetary atmospheres by accommodating for the presence of spectral lines that may be due to absorption and/or scattering processes, thereby, taking further departures from the strict gray approximation compared to previous calculations. We accomplish this by using the picket fence model devised by Chandrasekhar (1935). The picket fence model was originally used to derive an analytical model of non-irradiated stellar atmospheres (Mihalas, 1970), see Figure 6.3 for a rough schematic of the model concept. Recently, it was applied to model irradiated atmospheres by Parmentier and Guillot (2014), however, they only considered the case of lines resulting from pure absorption. We have extended the model to include the possibility of lines resulting from a combination of absorption and scattering processes. While exoplanetary atmospheres may be the most obvious and popular application, we propose the possibility of using our model to derive thermal profiles of irradiated binary companions such as a brown-dwarf in a binary system orbiting a white-dwarf companion (Santisteban et al., 2016; Showman, 2016).

Although such analytical models cannot possibly match or compete with the latest numerical atmospheric models in terms of sophistication they serve important and useful purposes (Hubeny, 2017; Hubeny and Mihalas, 2014). Analytical models such as the one we have derived can provide a quick means of estimating approximate atmospheric features for a minimal set of parameters. The physical insight gained from analytical models is without parallel and their pedagogical value is not to be under-appreciated. These models can also be used as initial approximations for numerical schemes and can also help benchmark results obtained

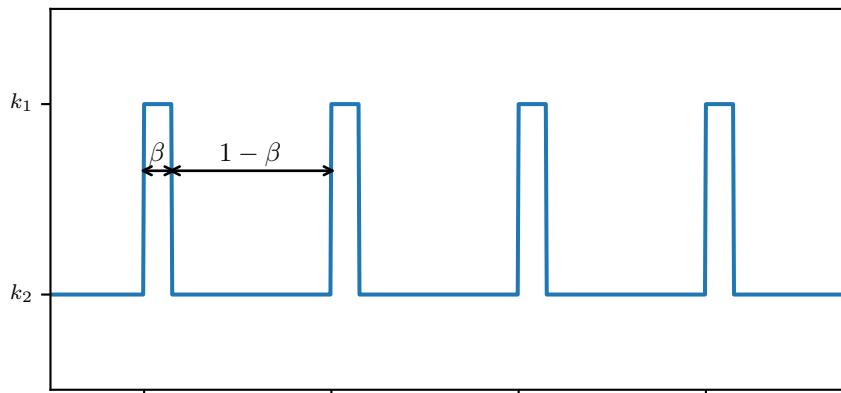


Figure 6.3: The picket fence model. A sequence of uniformly wide and uniformly spaced steps in opacity, measured in terms of the parameter β , across frequency represent the spectral lines on top of the continuum.

from computational models.

Bibliography

- Armitage, Philip J (2015). “Physical processes in protoplanetary disks”. In: *arXiv preprint arXiv:1509.06382*.
- Avenhaus, Henning, Sascha P Quanz, Antonio Garufi, Sebastian Perez, Simon Casassus, Christophe Pinte, Gesa H-M Bertrang, Claudio Caceres, Myriam Benisty, and Carsten Dominik (2018). “Disks ARound TTauri Stars with Sphere (DARTTS-S) I: Sphere/IRDIS Polarimetric Imaging of 8 prominent TTauri Disks”. In: *arXiv preprint arXiv:1803.10882*.
- Bai, Xue-Ning (2013). “Wind-driven accretion in protoplanetary disks. II. Radial dependence and global picture”. In: *The Astrophysical Journal* 772.2, p. 96.
- (2014a). “Hall effect controlled gas dynamics in protoplanetary disks. II. Full 3D simulations toward the outer disk”. In: *The Astrophysical Journal* 798.2, p. 84.
- (2014b). “Hall effect controlled gas dynamics in protoplanetary disks. II. Full 3D simulations toward the outer disk”. In: *The Astrophysical Journal* 798.2, p. 84.
- Bai, Xue-Ning and James M Stone (2011). “Effect of ambipolar diffusion on the nonlinear evolution of magnetorotational instability in weakly ionized disks”. In: *The Astrophysical Journal* 736.2, p. 144.
- (2013a). “Local study of accretion disks with a strong vertical magnetic field: magnetorotational instability and disk outflow”. In: *The Astrophysical Journal* 767.1, p. 30.

Bibliography

- Bai, Xue-Ning and James M Stone (2013b). “Wind-driven accretion in protoplanetary disks. I. Suppression of the magnetorotational instability and launching of the magnetocentrifugal wind”. In: *The Astrophysical Journal* 769.1, p. 76.
- (2017). “Hall Effect–Mediated Magnetic Flux Transport in Protoplanetary Disks”. In: *The Astrophysical Journal* 836.1, p. 46.
- Balbus, S. A. and J. F. Hawley (1991). “A powerful local shear instability in weakly magnetized disks. I - Linear analysis. II - Nonlinear evolution”. In: *The Astrophysical Journal* 376, pp. 214–233. DOI: 10.1086/170270.
- (1998). “Instability, turbulence, and enhanced transport in accretion disks”. In: *Reviews of Modern Physics* 70, pp. 1–53. DOI: 10.1103/RevModPhys.70.1.
- Balbus, Steven A (2009). “Magnetohydrodynamics of protostellar disks”. In: *arXiv preprint arXiv:0906.0854*.
- Balbus, Steven A and Caroline Terquem (2001). “Linear analysis of the Hall effect in protostellar disks”. In: *The Astrophysical Journal* 552.1, p. 235.
- Barker, Adrian J and Henrik N Latter (2015). “On the vertical-shear instability in astrophysical discs”. In: *Monthly Notices of the Royal Astronomical Society* 450.1, pp. 21–37.
- Béthune, William, Geoffroy Lesur, and Jonathan Ferreira (2016). “Self-organisation in protoplanetary discs-Global, non-stratified Hall-MHD simulations”. In: *Astronomy & Astrophysics* 589, A87.
- Binney, James and Scott Tremaine (2011). *Galactic dynamics*. Princeton university press.
- Bloch, A. M., P. S. Krishnaprasad, J. E. Marsden, and T. S. Ratiu (1994). “Dissipation Induced Instabilities”. In: *Annales de L'Institut Henri Poincare Section (C) Non Linear Analysis* 11, pp. 37–90. DOI: 10.1016/S0294-1449(16)30196-2.
- Boas, Mary L (2006). *Mathematical methods in the physical sciences*. Wiley.
- Bodenheimer Peter, Laughlin Gregory P (2006). *Numerical methods in astrophysics: an introduction*. Taylor & Francis.

- Burns, Joseph A, Douglas P Hamilton, and Mark R Showalter (2001). “Dusty rings and circumplanetary dust: Observations and simple physics”. In: *Interplanetary Dust*. Springer, pp. 641–725.
- Chandrasekhar, S. (1935). “The radiative equilibrium of the outer layers of a star, with special reference to the blanketing effect of the reversing layer”. In: *MNRAS* 96, p. 21. doi: [10.1093/mnras/96.1.21](https://doi.org/10.1093/mnras/96.1.21).
- (1960a). *Radiative transfer*.
- (1960b). “The Stability of Non-Dissipative Couette Flow in Hydromagnetics”. In: *Proceedings of the National Academy of Science* 46, pp. 253–257. doi: [10.1073/pnas.46.2.253](https://doi.org/10.1073/pnas.46.2.253).
- Davis, S. W., J. M. Stone, and M. E. Pessah (2010). “Sustained Magnetorotational Turbulence in Local Simulations of Stratified Disks with Zero Net Magnetic Flux”. In: *The Astrophysical Journal* 713, pp. 52–65. doi: [10.1088/0004-637X/713/1/52](https://doi.org/10.1088/0004-637X/713/1/52).
- Desch, SJ (2004). “Linear analysis of the magnetorotational instability, including ambipolar diffusion, with application to protoplanetary disks”. In: *The Astrophysical Journal* 608.1, p. 509.
- Dullemond, C. P. and J. D. Monnier (2010). “The Inner Regions of Protoplanetary Disks”. In: *Annual Review of Astronomy and Astrophysics* 48, pp. 205–239. doi: [10.1146/annurev-astro-081309-130932](https://doi.org/10.1146/annurev-astro-081309-130932).
- Dullin, H. R., M. Horányi, and J. E. Howard (2002). “Generalizations of the Störmer problem for dust grain orbits”. In: *Physica D Nonlinear Phenomena* 171, pp. 178–195. doi: [10.1016/S0167-2789\(02\)00550-X](https://doi.org/10.1016/S0167-2789(02)00550-X).
- Dzyurkevich, Natalia, Mario Flock, Neal J Turner, Hubert Klahr, and Th Henning (2010). “Trapping solids at the inner edge of the dead zone: 3-D global MHD simulations”. In: *Astronomy & Astrophysics* 515, A70.
- Esposito, Larry W (1993). “Understanding planetary rings”. In: *Annual Review of Earth and Planetary Sciences* 21.1, pp. 487–523.
- (2002). “Planetary rings”. In: *Reports on Progress in Physics* 65.12, p. 1741.

Bibliography

- Fleming, Timothy and James M. Stone (2003). “Local Magnetohydrodynamic Models of Layered Accretion Disks”. In: *The Astrophysical Journal* 585, pp. 908–920. DOI: 10.1086/345848.
- Frank, Juhan, Andrew King, and Derek Raine (2002). *Accretion power in astrophysics*. Cambridge university press.
- Freidberg, Jeffrey P (1987). “Ideal magnetohydrodynamics”. In:
- Fromang, S. and J. Papaloizou (2007). “MHD simulations of the magnetorotational instability in a shearing box with zero net flux. I. The issue of convergence”. In: *Astronomy and Astrophysics* 476, pp. 1113–1122. DOI: 10.1051/0004-6361:20077942.
- Gammie, C. F. and S. A. Balbus (1994). “Quasi-Global Linear Analysis of a Magnetized Disc”. In: *Monthly Notices of the Royal Astronomical Society* 270, p. 138. DOI: 10.1093/mnras/270.1.138.
- Goertz, CK (1989). “Dusty plasmas in the solar system”. In: *Reviews of Geophysics* 27.2, pp. 271–292.
- Goldreich, Peter and D Lynden-Bell (1965). “II. Spiral arms as sheared gravitational instabilities”. In: *Monthly Notices of the Royal Astronomical Society* 130.2, pp. 125–158.
- Goldstein, Herbert, Charles P Poole, and John L Safko (2014). *Classical Mechanics: Pearson New International Edition*. Pearson Higher Ed.
- Goodman, J. and G. Xu (1994). “Parasitic instabilities in magnetized, differentially rotating disks”. In: *The Astrophysical Journal* 432, pp. 213–223. DOI: 10.1086/174562.
- Goukenleuque, Cedric, Bruno Bezard, Benoit Joguet, Emmanuel Lellouch, and Richard Freedman (2000). “A radiative equilibrium model of 51 Peg b”. In: *Icarus* 143.2, pp. 308–323.
- Greiner, Walter (2009). *Classical mechanics: systems of particles and Hamiltonian dynamics*. Springer Science & Business Media.

- Gressel, Oliver, Neal J Turner, Richard P Nelson, and Colin P McNally (2015). “Global simulations of protoplanetary disks with ohmic resistivity and ambipolar diffusion”. In: *The Astrophysical Journal* 801.2, p. 84.
- Guan, X., C. F. Gammie, J. B. Simon, and B. M. Johnson (2009). “Locality of MHD Turbulence in Isothermal Disks”. In: *The Astrophysical Journal* 694, pp. 1010–1018. DOI: 10.1088/0004-637X/694/2/1010.
- Guillot, T. (2010). “On the radiative equilibrium of irradiated planetary atmospheres”. In: *AAP* 520, A27, A27. DOI: 10.1051/0004-6361/200913396.
- Hawley, J. F., C. F. Gammie, and S. A. Balbus (1995). “Local Three-dimensional Magnetohydrodynamic Simulations of Accretion Disks”. In: *The Astrophysical Journal* 440, p. 742. DOI: 10.1086/175311.
- Hazeltine, Richard D (2018). *The framework of plasma physics*. CRC Press.
- Hecht, J. (2016). “The truth about exoplanets”. In: *Nature* 530.
- Heng, K., J. M. Mendonça, and J.-M. Lee (2014). “Analytical Models of Exoplanetary Atmospheres. II. Radiative Transfer via the Two-stream Approximation”. In: *The Astrophysical Journal Supplement* 215, 4, p. 4. DOI: 10.1088/0067-0049/215/1/4.
- Henon, M. (1982). “Vlasov equation”. In: *Astronomy and Astrophysics* 114, p. 211.
- Hirose, S and NJ Turner (2011). “Heating and Cooling Protostellar Disks”. In: *The Astrophysical Journal Letters* 732.2, p. L30.
- Horányi, Mihály (1996). “Charged dust dynamics in the solar system”. In: *Annual Review of Astronomy and Astrophysics* 34.1, pp. 383–418.
- Howard, J. E. (1999). “Stability of relative equilibria in arbitrary axisymmetric gravitational and magnetic fields”. In: *Celestial Mechanics and Dynamical Astronomy* 74, pp. 19–57. DOI: 10.1023/A:1008388105585.
- Howard, J. E., H. R. Dullin, and M. Horányi (2000). “Stability of Halo Orbits”. In: *Physical Review Letters* 84, pp. 3244–3247. DOI: 10.1103/PhysRevLett.84.3244.

Bibliography

- Howard, J. E., M. Horányi, and G. R. Stewart (1999). “Global Dynamics of Charged Dust Particles in Planetary Magnetospheres”. In: *Physical Review Letters* 83, pp. 3993–3996. doi: 10.1103/PhysRevLett.83.3993.
- Hubeny, I. (2017). “Model atmospheres of sub-stellar mass objects”. In: *MNRAS* 469, pp. 841–869. doi: 10.1093/mnras/stx758.
- Hubeny, I. and D. Mihalas (2014). *Theory of Stellar Atmospheres*.
- Hubeny, Ivan, Adam Burrows, and David Sudarsky (2003). “A possible bifurcation in atmospheres of strongly irradiated stars and planets”. In: *The Astrophysical Journal* 594.2, p. 1011.
- Irwin, PGJ, NA Teanby, R De Kok, LN Fletcher, CJA Howett, CCC Tsang, CF Wilson, SB Calcutt, CA Nixon, and PD Parrish (2008). “The NEMESIS planetary atmosphere radiative transfer and retrieval tool”. In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 109.6, pp. 1136–1150.
- Klahr, Hubert and Alexander Hubbard (2014). “Convective overstability in radially stratified accretion disks under thermal relaxation”. In: *The Astrophysical Journal* 788.1, p. 21.
- Kley, W. and R. P. Nelson (2012). “Planet-Disk Interaction and Orbital Evolution”. In: *Annual Review of Astronomy and Astrophysics* 50, pp. 211–249. doi: 10.1146/annurev-astro-081811-125523.
- Knobloch, Edgar and Keith Julien (2005). “Saturation of the magnetorotational instability”. In: *Physics of Fluids* 17.9, p. 094106.
- Krall, Nicholas A and Alvin W Trivelpiece (1973). “Principles of plasma physics”. In: *American Journal of Physics* 41.12, pp. 1380–1381.
- Krapp, L., O. Gressel, P. Benitez-Llambay, T. Downes, G. Mohandas, and M. E. Pessah (2018). “Dust segregation in Hall-dominated turbulent protoplanetary disks”. In: *The Astrophysical Journal(to be submitted)*.
- Krechetnikov, R. and J. E. Marsden (2007). “Dissipation-induced instabilities in finite dimensions”. In: *Reviews of Modern Physics* 79, pp. 519–553. doi: 10.1103/RevModPhys.79.519.

- Krechetnikov, Rouslan and Jerrold E Marsden (2009). “Dissipation-induced instability phenomena in infinite-dimensional systems”. In: *Archive for rational mechanics and analysis* 194.2, pp. 611–668.
- Kunz, M. W. (2008). “On the linear stability of weakly ionized, magnetized planar shear flows”. In: *Monthly Notices of the Royal Astronomical Society* 385, pp. 1494–1510. DOI: 10.1111/j.1365-2966.2008.12928.x.
- Kunz, Matthew W and Steven A Balbus (2004). “Ambipolar diffusion in the magnetorotational instability”. In: *Monthly Notices of the Royal Astronomical Society* 348.1, pp. 355–360.
- Kunz, Matthew W and Geoffroy Lesur (2013). “Magnetic self-organization in Hall-dominated magnetorotational turbulence”. In: *Monthly Notices of the Royal Astronomical Society* 434.3, pp. 2295–2312.
- Latter, H. N., P. Lesaffre, and S. A. Balbus (2009). “MRI channel flows and their parasites”. In: *Monthly Notices of the Royal Astronomical Society* 394, pp. 715–729. DOI: 10.1111/j.1365-2966.2009.14395.x.
- Latter, Henrik N., Sébastien Fromang, and Julien Faure (2015). “Local and global aspects of the linear MRI in accretion discs”. In: *Monthly Notices of the Royal Astronomical Society* 453, pp. 3257–3268. DOI: 10.1093/mnras/stv1890.
- Lesur, G. and P. Y. Longaretti (2007). “Impact of dimensionless numbers on the efficiency of magnetorotational instability induced turbulent transport”. In: *Monthly Notices of the Royal Astronomical Society* 378, pp. 1471–1480. DOI: 10.1111/j.1365-2966.2007.11888.x.
- Lesur, Geoffroy, Matthew W Kunz, and Sébastien Fromang (2014). “Thanatology in protoplanetary discs-The combined influence of Ohmic, Hall, and ambipolar diffusion on dead zones”. In: *Astronomy & Astrophysics* 566, A56.
- Lesur, Geoffroy and John CB Papaloizou (2010). “The subcritical baroclinic instability in local accretion disc models”. In: *Astronomy & Astrophysics* 513, A60.
- Lin, D. N. C. and J. C. B. Papaloizou (1996). “Theory of Accretion Disks II: Application to Observed Systems”. In: *Annual Review of Astronomy and Astrophysics* 34, pp. 703–748. DOI: 10.1146/annurev.astro.34.1.703.

Bibliography

- Line, Michael R, Aaron S Wolf, Xi Zhang, Heather Knutson, Joshua A Kammer, Elias Ellison, Pieter Deroo, Dave Crisp, and Yuk L Yung (2013). “A systematic retrieval analysis of secondary eclipse spectra. I. A comparison of atmospheric retrieval techniques”. In: *The Astrophysical Journal* 775.2, p. 137.
- Longaretti, P-Y and G Lesur (2010). “MRI-driven turbulent transport: the role of dissipation, channel modes and their parasites”. In: *Astronomy & Astrophysics* 516, A51.
- Madhusudhan, N., H. Knutson, J. J. Fortney, and T. Barman (2014). “Exoplanetary Atmospheres”. In: *Protostars and Planets VI*, pp. 739–762. doi: 10.2458/azu_uapress_9780816531240-ch032.
- Malik, Matej, Luc Grosheintz, João M Mendonca, Simon L Grimm, Baptiste Lavie, Daniel Kitzmann, Shang-Min Tsai, Adam Burrows, Laura Kreidberg, Megan Bedell, et al. (2017). “HELIOS: An Open-source, GPU-accelerated Radiative Transfer Code for Self-consistent Exoplanetary Atmospheres”. In: *The Astronomical Journal* 153.2, p. 56.
- Mayor, M. and D. Queloz (1995). “A Jupiter-mass companion to a solar-type star”. In: *Nature* 378, pp. 355–359. doi: 10.1038/378355a0.
- Merkin, David R (2012). *Introduction to the Theory of Stability*. Vol. 24. Springer Science & Business Media.
- Mihalas, D. (1970). *Stellar atmospheres*. W. H. Freeman and Company.
- Mihalas, D. and B. W. Mihalas (1984). *Foundations of radiation hydrodynamics*.
- Mohandas, G., T. Heinemann, and M. E. Pessah (2018). “Orbital stability in static axisymmetric fields”. In: *ArXiv e-prints*.
- Mohandas, G. and M. E. Pessah (2017). “Spectral Analysis of Non-ideal MRI Modes: The Effect of Hall Diffusion”. In: *The Astrophysical Journal* 838, 48, p. 48. doi: 10.3847/1538-4357/aa6118.
- Mohandas, G., M. E. Pessah, and K. Heng (2018). “Analytical Models of Exoplanetary Atmospheres. V. Non-gray Thermal Structure with Coherent Scattering”. In: *The Astrophysical Journal* 858, 1, p. 1. doi: 10.3847/1538-4357/aab35d.

- Morrison, Philip J (1998). "Hamiltonian description of the ideal fluid". In: *Reviews of modern physics* 70.2, p. 467.
- Nauman, Farrukh and Martin E Pessah (2016). "Sustained Turbulence in Differentially Rotating Magnetized Fluids at a Low Magnetic Prandtl Number". In: *The Astrophysical Journal* 833.2, p. 187.
- Netzer, Hagai (2015). "Revisiting the Unified Model of Active Galactic Nuclei". In: *Annual Review of Astronomy and Astrophysics* 53, pp. 365–408. doi: 10.1146/annurev-astro-082214-122302.
- Nicholson, Dwight Roy (1983). *Introduction to plasma theory*. Wiley New York.
- Ogilvie, Gordon I (2016). "Astrophysical fluid dynamics". In: *Journal of Plasma Physics* 82.3.
- Pandey, B. P. and M. Wardle (2008). "Hall magnetohydrodynamics of partially ionized plasmas". In: *Monthly Notices of the Royal Astronomical Society* 385, pp. 2269–2278. doi: 10.1111/j.1365-2966.2008.12998.x.
- Parmentier, V. and T. Guillot (2014). "A non-grey analytical model for irradiated atmospheres. I. Derivation". In: *Astronomy and Astrophysics* 562, A133, A133. doi: 10.1051/0004-6361/201322342.
- Pessah, M. E. (2010). "Angular Momentum Transport in Protoplanetary and Black Hole Accretion Disks: The Role of Parasitic Modes in the Saturation of MHD Turbulence". In: *The Astrophysical Journal* 716, pp. 1012–1027. doi: 10.1088/0004-637X/716/2/1012.
- Pessah, M. E. and C.-k. Chan (2008). "Viscous, Resistive Magnetorotational Modes". In: *The Astrophysical Journal* 684, pp. 498–514. doi: 10.1086/589915.
- Pessah, M. E. and J. Goodman (2009). "On the Saturation of the Magnetorotational Instability Via Parasitic Modes". In: *The Astrophysical Journal, Letters* 698, pp. L72–L76. doi: 10.1088/0004-637X/698/1/L72.
- Pessah, Martin E., Chi-Kwan Chan, and Dimitrios Psaltis (2006a). "Local Model for Angular-Momentum Transport in Accretion Disks Driven by the Magnetorotational Instability". In: *Physical Review Letters* 97, 221103, p. 221103. doi: 10.1103/PhysRevLett.97.221103.

Bibliography

- Pessah, Martin E., Chi-Kwan Chan, and Dimitrios Psaltis (2006b). “The signature of the magnetorotational instability in the Reynolds and Maxwell stress tensors in accretion discs”. In: *Monthly Notices of the Royal Astronomical Society* 372, pp. 183–190. doi: 10.1111/j.1365-2966.2006.10824.x.
- Pessah, Martin E., Chi-kwan Chan, and Dimitrios Psaltis (2007). “Angular Momentum Transport in Accretion Disks: Scaling Laws in MRI- driven Turbulence”. In: *The Astrophysical Journal* 668, pp. L51–L54. doi: 10.1086/522585.
- Popham, Robert and Ramesh Narayan (1991). “Does Accretion Cease When a Star Approaches Breakup?” In: *The Astrophysical Journal* 370, p. 604. doi: 10.1086/169847.
- Pringle, James E and Andrew King (2007). *Astrophysical flows*. Cambridge University Press.
- Pringle, JE (1981). “Accretion discs in astrophysics”. In: *Annual review of astronomy and astrophysics* 19.1, pp. 137–160.
- Riley, Kenneth Franklin, Michael Paul Hobson, and Stephen John Bence (2006). *Mathematical methods for physics and engineering: a comprehensive guide*. Cambridge university press.
- Rüdiger, G. and L. L. Kitchatinov (2005). “The influence of the Hall effect on the global stability of cool protostellar disks”. In: *Astronomy and Astrophysics* 434, pp. 629–635. doi: 10.1051/0004-6361:20042499.
- Rumiantsev, Valentin V (1966). “On the stability of steady motions”. In: *Journal of Applied Mathematics and Mechanics* 30.5, pp. 1090–1103.
- Sano, Takayoshi and James M Stone (2002a). “The effect of the Hall term on the nonlinear evolution of the magnetorotational instability. I. Local axisymmetric simulations”. In: *The Astrophysical Journal* 570.1, p. 314.
- (2002b). “The effect of the Hall term on the nonlinear evolution of the magnetorotational instability. II. Saturation level and critical magnetic Reynolds number”. In: *The Astrophysical Journal* 577.1, p. 534.
- Santisteban, Juan V Hernández, Christian Knigge, Stuart P Littlefair, Rene P Breton, Vikram S Dhillon, Boris T Gänsicke, Thomas R Marsh, Magaretha L

- Pretorius, John Southworth, and Peter H Hauschmidt (2016). “An irradiated brown-dwarf companion to an accreting white dwarf”. In: *Nature* 533.7603, p. 366.
- Seager, S. (2010). *Exoplanet Atmospheres: Physical Processes*.
- Shakura, N and Rashid Alievich Sunyaev (1973). “Black holes in binary systems. Observational appearance.” In: *Astronomy and Astrophysics* 24, pp. 337–355.
- Shi, Ji-Ming, James M Stone, and Chelsea X Huang (2015). “Saturation of the magnetorotational instability in the unstratified shearing box with zero net flux: convergence in taller boxes”. In: *Monthly Notices of the Royal Astronomical Society* 456.3, pp. 2273–2289.
- Showman, A. P., J. Y.-K. Cho, and K. Menou (2010). “Atmospheric Circulation of Exoplanets”. In: *Exoplanets*. Ed. by S. Seager, pp. 471–516.
- Showman, Adam P (2016). “Astrophysics: Illuminating brown dwarfs”. In: *Nature* 533.7603, p. 330.
- Shu, Frank H (1991). *The Physics of Astrophysics: Gas Dynamics*. Vol. 2. University Science Books.
- Simon, Jacob B, Geoffroy Lesur, Matthew W Kunz, and Philip J Armitage (2015). “Magnetically driven accretion in protoplanetary discs”. In: *Monthly Notices of the Royal Astronomical Society* 454.1, pp. 1117–1131.
- Stone, James M, John F Hawley, Charles F Gammie, and Steven A Balbus (1996). “Three-dimensional magnetohydrodynamical simulations of vertically stratified accretion disks”. In: *The Astrophysical Journal* 463, p. 656.
- Störmer, Carl (1955). *The polar aurora*. Clarendon Press.
- Sturrock, Peter Andrew (1994). *Plasma Physics: an introduction to the theory of astrophysical, geophysical and laboratory plasmas*. Cambridge University Press.
- Thomas, Gary E and Knut Stamnes (2002). *Radiative transfer in the atmosphere and ocean*. Cambridge University Press.
- Thomson, W. and P. G. Tait (1883a). *Treatise on Natural Philosophy (Vol 1)*. Cambridge: Cambridge University Press.

Bibliography

- Thomson, W. and P. G. Tait (1883b). *Treatise on Natural Philosophy (Vol 2)*. Cambridge: Cambridge University Press.
- Umurhan, Orkan M, K Menou, and O Regev (2007). “Weakly nonlinear analysis of the magnetorotational instability in a model channel flow”. In: *Physical review letters* 98.3, p. 034501.
- Velikhov, EP (1959). “Stability of an ideally conducting liquid flowing between cylinders rotating in a magnetic field”. In: *Sov. Phys. JETP* 36.9, pp. 995–998.
- Waldmann, Ingo P, Giovanna Tinetti, Marco Rocchetto, Emma J Barton, Sergey N Yurchenko, and Jonathan Tennyson (2015). “Tau-REx I: A next generation retrieval code for exoplanetary atmospheres”. In: *The Astrophysical Journal* 802.2, p. 107.
- Wardle, Mark (1999). “The Balbus-Hawley instability in weakly ionized discs”. In: *Monthly Notices of the Royal Astronomical Society* 307.4, pp. 849–856.
- Wardle, Mark and Raquel Salmeron (2012). “Hall diffusion and the magnetorotational instability in protoplanetary discs”. In: *Monthly Notices of the Royal Astronomical Society* 422.4, pp. 2737–2755.
- Williams, J. P. and L. A. Cieza (2011). “Protoplanetary Disks and Their Evolution”. In: *Annual Review of Astronomy and Astrophysics* 49, pp. 67–117. doi: 10.1146/annurev-astro-081710-102548.
- Wyatt, Mark C (2008). “Evolution of debris disks”. In: *Annu. Rev. Astron. Astrophys.* 46, pp. 339–383.

Paper I

Orbital Stability in Static Axisymmetric Fields

Submitted to

Celestial Mechanics & Dynamical Astronomy

arxiv:1801.07106

Orbital stability in static axisymmetric fields

Gopakumar Mohandas,¹ Tobias Heinemann,¹ and Martín E. Pessah¹

¹*Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark*
(Dated: January 23, 2018)

We investigate the stability of test-particle equilibrium orbits in axisymmetric, but otherwise arbitrary, gravitational and electromagnetic fields. We extend previous studies of this problem to include a toroidal magnetic field. We find that, even though the toroidal magnetic field does not alter the location of the circular orbits, it enters the problem as a gyroscopic force with the potential to provide gyroscopic stability. This is in essence similar to the situation encountered in the reduced three-body problem where rotation enables stability around the local maxima of the effective potential. Nevertheless, we show that gyroscopic stabilization by a toroidal magnetic field is impossible for axisymmetric force fields in source-free regions because in this case the effective potential does not possess any local maxima. As an example of an axisymmetric force field with sources, we consider the classical problem of a rotating, aligned magnetosphere. By analyzing the dynamics of halo and equatorial particle orbits we conclude that axisymmetric toroidal fields that are antisymmetric about the equator are unable to provide gyroscopic stabilization. On the other hand, a toroidal magnetic field that does not vanish at the equator can provide gyroscopic stabilization for positively charged particles in prograde equatorial orbits.

1. INTRODUCTION

The study of particle dynamics in axisymmetric fields is important in a variety of problems spanning a wide range of scales in nature. The axial symmetry can be exploited to reduce the number of degrees of freedom. In the reduced phase space, the scalar potential, which conservative forces derive from, is replaced by an effective potential that includes the toroidal kinetic energy. Critical points of the effective potential are equilibria of the reduced system. These correspond to circular orbits in three-dimensional space. Howard (1999) provides a thorough overview of the stability of circular orbits in axisymmetric gravitational and electromagnetic fields. The magnetic fields considered there are assumed to be purely poloidal. While this is a reasonable assumption to make in many axisymmetric systems, it is not difficult to think of examples where toroidal magnetic fields play a significant role. Such examples include toroidal fusion devices and galactic disks.

In this paper, we generalize Howard (1999) analysis addressing the stability of circular orbits by including a toroidal axisymmetric magnetic field. Axisymmetric poloidal and toroidal magnetic fields enter the dynamics in a fundamentally different way. All the effects related to the poloidal magnetic field can be encapsulated in the effective potential together with the gravitational contribution, whereas this is never the case for the toroidal part. As we detail below, this crucial difference makes it possible for the toroidal magnetic field to provide stability in regions of parameter space where a purely poloidal field cannot. This is a manifestation of the phenomenon known as gyroscopic stability, which is usually associated with the Coriolis force.

The rest of the paper is organized as follows. In section 2 we state the equations of motion for a test particle in reduced phase space by introducing the Routhian and an effective potential that depends only on the poloidal flux function. In section 3 we analyze the stability of circular orbits and find the conditions for gyroscopic stabilization via a toroidal magnetic field. In particular, we show that in source-free regions, gyroscopic stabilization via a toroidal magnetic field is impossible. In section 4 we illustrate some of the implications of our findings by analysing the problem of a rotating magnetosphere. We conclude by discussing our results in section 5.

2. EQUATIONS OF MOTION

The motion of a classical, non-relativistic¹ particle is governed by the Lagrangian per unit particle mass

$$L = \frac{1}{2} \dot{\mathbf{r}}^2 + \mathbf{A} \cdot \dot{\mathbf{r}} - \Phi, \quad (1)$$

where $\dot{\mathbf{r}}$ denotes the time derivative of the particle's position vector. We have absorbed the coupling constants (i.e., charge and mass) in the scalar potential Φ and the vector potential \mathbf{A} , both of which are assumed time-independent.

¹ This is the most general single particle Lagrangian compatible with Galilean invariance (Jauch 1964, Roman and Leveille 1974).

In the following, we will refer to Φ and \mathbf{A} as electromagnetic potentials. Note, however, that Φ can include the gravitational potential and/or the centrifugal potential and \mathbf{A} can include a contribution accounting for the Coriolis force that arises in a rotating frame. The equation of motion derived from eq. (1) via the Euler-Lagrange equation is $\ddot{\mathbf{r}} = \mathbf{E} + \dot{\mathbf{r}} \times \mathbf{B}$, where the electric field is $\mathbf{E} = -\nabla\Phi$ and the magnetic field is $\mathbf{B} = \nabla \times \mathbf{A}$. We note that if only electromagnetic forces are present, then \mathbf{E} and \mathbf{B} differ from the true electromagnetic fields by a factor equal to the charge-to-mass ratio.

2.1. Motion in reduced phase space

We work in cylindrical coordinates (ρ, φ, z) and assume that the system is symmetric about the z -axis. This means that neither \mathbf{A} nor Φ depend on the cyclic coordinate φ . From the Euler-Lagrange equation $d/dt(\partial L/\partial \dot{\varphi}) = \partial L/\partial \dot{\varphi}$ it then follows that the generalized angular momentum $p_\varphi = \partial L/\partial \dot{\varphi}$ is an integral of motion. Substituting eq. (1) we obtain

$$p_\varphi = \rho^2 \dot{\varphi} + \psi, \quad (2)$$

where we have introduced the poloidal flux function

$$\psi = \rho A_\varphi, \quad (3)$$

in terms of which the magnetic field is given by $\mathbf{B} = \nabla\psi \times \nabla\varphi + \rho B_\varphi \nabla\varphi$.

Since p_φ is an integral of motion, the dimensionality of the problem may be reduced by one. For this we introduce the Routhian $R = L - \omega p_\varphi$, where

$$\omega = \frac{1}{\rho^2}(p_\varphi - \psi) \quad (4)$$

is equal to the angular velocity $\dot{\varphi}$ expressed through eq. (2) as a function of ρ and z . With eq. (4) the Routhian is given by

$$R = \frac{1}{2}(\dot{\rho}^2 + \dot{z}^2) + A_\rho \dot{\rho} + A_z \dot{z} - U, \quad (5)$$

where

$$U = \Phi + \frac{\rho^2 \omega^2}{2} \quad (6)$$

is the *effective* potential.

The equations of motion of the reduced system are $d/dt(\partial R/\partial \dot{q}^i) = \partial R/\partial q^i$ for $q^i = (\rho, z)$ or

$$\begin{aligned} \ddot{\rho} + B_\varphi \dot{z} + \partial U / \partial \rho &= 0, \\ \ddot{z} - B_\varphi \dot{\rho} + \partial U / \partial z &= 0. \end{aligned} \quad (7)$$

These equations possess the energy integral

$$H = \frac{1}{2}(\dot{\rho}^2 + \dot{z}^2) + U, \quad (8)$$

which evidently is independent of the *toroidal* magnetic field. This is because the force due to the toroidal magnetic field does not do work in the reduced configuration space. Such forces are called gyroscopic forces (Thomson and Tait 1883a,b). Very much in contrast to this, the force due to the poloidal magnetic field, which in fact is gyroscopic before reduction, has become a potential force in the reduced configuration space. A reduced system of the form (7) with $B_\varphi \neq 0$ is said to be gyroscopically constrained or coupled (see e.g. Rumiantsev 1966, Merkin 1996).

2.2. Hamiltonian formalism

Before moving on to study the stability of equilibrium solutions of the equations of motion, we note that the reduced system may also be described using a Hamiltonian formalism. In a gyroscopically coupled system this is best done by working in non-canonical phase space coordinates

$$w^\alpha = (\rho, z, \dot{\rho}, \dot{z}), \quad (9)$$

see e.g. Littlejohn (1979, 1982) or Bolotin and Negrini (1995). In these coordinates, the equations of motion are

$$\dot{w}^\alpha = J^{\alpha\beta} \frac{\partial H}{\partial w^\beta}, \quad (10)$$

where the Hamiltonian H is defined in eq. (8) and the Poisson matrix $J^{\alpha\beta}$ is given by

$$J^{\alpha\beta} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & -B_\varphi \\ 0 & -1 & B_\varphi & 0 \end{pmatrix}. \quad (11)$$

It is straightforward to verify that given the definitions in eqs. (10) and (11), Hamilton's equations in the form of eq. (10) are equivalent to the equations of motion (7). We also note that the Poisson bracket defined by $\{f, g\} = (\partial f / \partial w^\alpha) J^{\alpha\beta} (\partial g / \partial w^\beta)$ satisfies the Jacobi identity $\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0$ for any $B_\varphi(\rho, z)$, as it should.

3. STABILITY OF CIRCULAR ORBITS

In this section we discuss the stability of equilibrium solutions to the equations of motion (7). We note that all that is said here in fact holds for arbitrary forms of the effective potential. This means that for instance the restricted three-body problem is within the scope of our discussion. Only in section 4 will we specialize to effective potentials of the form given in eq. (6), in which the scalar potential Φ is independent of the canonical angular momentum p_φ .

3.1. Stability criteria

Equilibria of the reduced system described by eq. (5) are solutions with $\dot{\rho} = \dot{z} = 0$. The angular velocity $\dot{\varphi}$ is, however, in general non-zero. Equilibria of the reduced system thus correspond to uniformly rotating solutions of the original system. In the classic literature (e.g. Routh 1884), such solutions are known as steady motions. A more modern term is relative equilibria (e.g. Marsden and Weinstein 1974).² In the following we will refer to these solutions simply as *circular orbits*.

Inspection of the equations of motion (7) reveals that circular orbits are stationary points of the effective potential, i.e. points at which $\delta U = 0$ for arbitrary variations $\delta\rho$ and δz . Their location is evidently independent of the toroidal magnetic field B_φ . This agrees with the expectation that because the magnetic force is perpendicular to the velocity, the toroidal magnetic field should of course have no effect on strictly circular orbits.

Whether or not circular orbits can be expected to actually occur in nature (along with nearly circular orbits in their vicinity) depends on their stability. Various notions of stability exist in the literature (see e.g. Holm *et al.* 1985). Arguably the most important one for practical purposes is due to Lyapunov: an equilibrium w_e^α , with w^α defined in eq. (9), is stable for every $\varepsilon > 0$ if there is a $\delta > 0$ such that if $|w^\alpha(0) - w_e^\alpha| < \delta$ then $|w^\alpha(t) - w_e^\alpha| < \varepsilon$ for $t > 0$. It is important to note that $w^\alpha(t)$ in this definition evolves according to the nonlinear equations of motion (10). By contrast, spectral stability is concerned with the spectrum of the Hamiltonian matrix $J^{\alpha\beta} \partial^2 H / \partial w^\beta \partial w^\gamma$, obtained by linearizing eq. (10). A Hamiltonian system is spectrally stable if all eigenvalues of this matrix lie on the imaginary axis. Lyapunov stability implies spectral stability but not vice versa (Holm *et al.* 1985). In the following, stability will generally be synonymous with Lyapunov stability — the stronger of the two notions — unless specified otherwise.

The discussion in the above paragraph has an important caveat: the canonical angular momentum p_φ (the cyclic integral) is assumed fixed. Truly arbitrary perturbations would also allow p_φ to be varied. It is often argued that in practice the restriction of fixed cyclic integrals is unimportant because, as Pars (1965) writes, “if we do allow small changes in the [cyclic integrals] we are merely transferring our attention to oscillations about a neighboring state of steady motion.” This argument is originally due to Lyapunov (see Rumiantsev 1966). It may of course be that there is no neighboring state of steady motion, in which case this argument fails and a more elaborate approach becomes necessary (Salvadori 1953). A comprehensive discussion is given by Hagedorn (1971). In the following we will ignore these subtleties and take p_φ as fixed.

² It should be noted that relative equilibria potentially encompass a much wider class of solutions than just steady motions. Like steady motions, relative equilibria are obtained by reduction through symmetry, but unlike steady motions, relative equilibria allow for the underlying symmetry group to be non-Abelian.

According to Routh's theorem (Routh 1884), a circular orbit is stable if it corresponds to an isolated minimum of the effective potential. This is the case if the second variation $\delta^2 U > 0$ for any non-zero $\delta\rho$ and/or δz . Expressed in terms of the Hessian \mathcal{Q} of the effective potential, whose trace and determinant are given by

$$\text{tr } \mathcal{Q} = \frac{\partial^2 U}{\partial \rho^2} + \frac{\partial^2 U}{\partial z^2} \quad \text{and} \quad \det \mathcal{Q} = \frac{\partial^2 U}{\partial \rho^2} \frac{\partial^2 U}{\partial z^2} - \left(\frac{\partial^2 U}{\partial \rho \partial z} \right)^2, \quad (12)$$

a critical point of the effective potential is an isolated minimum if and only if

$$\text{tr } \mathcal{Q} > 0 \quad \text{and} \quad \det \mathcal{Q} > 0. \quad (13)$$

If the inequalities in eq. (13) are satisfied, then the total energy at equilibrium is positive definite (i.e. $\delta^2 H > 0$ for any non-zero δw^α) and can thus be used as a Lyapunov function to prove Routh's theorem using Lyapunov's direct method (Merkin 1996). In the absence of a toroidal magnetic field ($B_\varphi = 0$), in which case the system is gyroscopically decoupled, the converse is also true (Hagedorn 1971, Lyapunov 1907, Malkin 1959, Chetaev 1961, Rumyantsev and Sosnitskii 1993): circular orbits are unstable if they do *not* minimize U locally.

3.2. Gyroscopic stabilization

In the presence of a toroidal magnetic field, the system is gyroscopically coupled. In this case, all circular orbits located at isolated minima of U are still stable. However, there may now also exist stable circular orbits located at isolated *maxima* of the effective potential, where $\delta^2 U < 0$. This is known as gyroscopic stabilization (Thomson and Tait 1883a,b, Merkin 1996, Chetaev 1961). All orbits are spectrally unstable at *saddle points* ($\det \mathcal{Q} < 0$).

A note about energetics is in order here. Isolated minima of the effective potential correspond to isolated minima of the total energy because the kinetic energy is positive definite. In other words $\delta^2 H > 0$. Definiteness of $\delta^2 H$ is referred to as formal or energetic stability (Holm *et al.* 1985, Scheeres 2006). It is a sufficient but not necessary condition for Lyapunov stability. The toroidal magnetic field does not affect the energetic stability of circular orbits. It can, however, stabilize energetically unstable orbits, namely isolated maxima of U , for which $\delta^2 H$ is indefinite because $\delta^2 U < 0$.

In order to determine the conditions for gyroscopic stabilization to occur, we first carry out a spectral stability analysis. Considering infinitesimal perturbations of the linearized equations of motion eq. (7) leads to the characteristic polynomial

$$\sigma^4 + (B_\varphi^2 + \text{tr } \mathcal{Q})\sigma^2 + \det \mathcal{Q} = 0. \quad (14)$$

The roots of this equation comprise the spectrum of the Hamiltonian matrix $J^{\alpha\beta} \partial^2 H / \partial w^\beta \partial w^\gamma$ mentioned above. A comprehensive discussion of eq. (14) is given in Bloch *et al.* (1994), see also Chetaev (1961) and Haller (1992). Isolated maxima of the effective potential are critical points where $\delta^2 U < 0$ or, equivalently, $\det \mathcal{Q} > 0$ and $\text{tr } \mathcal{Q} < 0$. Depending on the strength of the toroidal field, only a subset of these maxima are gyroscopically stabilized. The precise conditions are

$$-B_\varphi^2 < \text{tr } \mathcal{Q} < 0 \quad \text{and} \quad 0 < 4 \det \mathcal{Q} < (\text{tr } \mathcal{Q} + B_\varphi^2)^2. \quad (15)$$

Note that for a given maximum of the effective potential, it is always possible to satisfy these inequalities for large enough B_φ . A visual representation of the inequalities in eqs. (13) and (15) is given in fig. 1.

If the inequalities in eq. (15) are satisfied, then the circular orbit is gyroscopically stabilized in the spectral sense: all eigenvalues σ as given by the roots of eq. (14) lie on the imaginary axis. In order to show that circular orbits can be gyroscopically stabilized in the Lyapunov sense is more challenging. Since U is negative definite at an isolated maximum, the total energy H defined in eq. (8) is indefinite and thus cannot be used as a Lyapunov function to prove stability using the direct method.

Instead, Lyapunov stability at isolated maxima of the effective potential can be demonstrated with the help of the Kolmogorov-Arnold-Moser (KAM) theorem. At a spectrally stable isolated maximum, where eqs. (15) are satisfied, the eigenvalues $\pm\sigma_1$ and $\pm\sigma_2$ of the linearized system, given by the roots of the characteristic polynomial (14), are purely imaginary. If these eigenvalues are non-resonant ($\sigma_1^2 \neq n^2 \sigma_2^2$ for $n = 1, 2, 3$), then the nonlinear dynamics of the system close to equilibrium is nearly integrable. The integrable, linear dynamics takes place on two-dimensional tori in phase space and the KAM theorem ensures these tori persist under nonlinear perturbations, provided certain non-degeneracy conditions are met (Haller 1992, Arnold 1963). Note that since p_φ is assumed fixed, these considerations strictly speaking only demonstrate *conditional* Lyapunov stability for the reduced system.

It is important to stress that both resonance and degeneracy occur with probability zero in continuous (real valued) parameter space. Moreover, if the system is resonant it may still be stable (Sokolskii 1974). Likewise, non-degeneracy is sufficient but not necessary for stability: in systems with two degrees of freedom (such as the present one) degenerate equilibria are, as a rule, stable (Arnold *et al.* 2006, sec. 6.3.6.B).

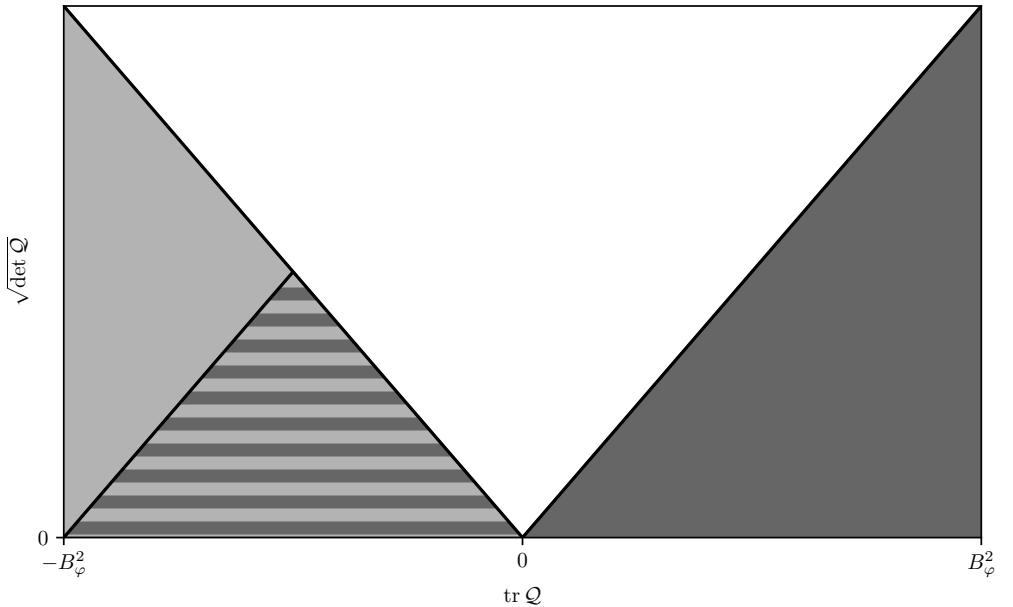


Figure 1. Orbits in the dark-grey region are formally stable. Orbits in the stripped region are gyroscopically stabilized by a toroidal magnetic field B_φ . Orbits in the light-grey region outside the stripped region are unstable. The fraction of this parameter space that is actually populated with orbits depends on the specific structure of the electromagnetic potentials. Note that the trace and the determinant of a real symmetric 2×2 -matrix satisfy $2\sqrt{\det \mathcal{Q}} \leq |\text{tr } \mathcal{Q}|$, thus the white area is inaccessible.

3.3. The effects of dissipation

The above considerations need to be amended if the system is dissipative. Minima of the effective potential remain stable when dissipation is added to the system, however, all other equilibria are unstable no matter how small (but finite) the dissipation is. In particular, gyroscopically stabilized equilibria, which correspond to maxima of the effective potential, lose their stability if dissipation is added (Thomson and Tait 1883a,b, Rumiantsev 1966, Chetaev 1961, Haller 1992). Loss of gyroscopic stabilization due to dissipation is an instance of a wider class of phenomena known as dissipation-induced instabilities (Bloch *et al.* 1994, Krechetnikov and Marsden 2007).

In practice gyroscopic stabilization is thus only a transient phenomenon. The growth rate of these instabilities, which are generally proportional to the dissipation rate (Bloch *et al.* 1994, MacKay 1991), is much smaller than the growth rates in the absence of gyroscopic forces. Thomson and Tait (1883a,b) refer to this as temporary, as opposed to secular, stability.

The fact that in realistic systems, gyroscopic forces are not able to truly stabilize an otherwise unstable equilibrium does not diminish the significance of gyroscopic stabilization by very much. This can be seen in the restricted three-body problem applied to the Sun-Jupiter system. The critical points of the effective potential in this problem are either saddle points (L_1 , L_2 , and L_3) or local maxima (L_4 and L_5). Gyroscopic stability at L_4 and L_5 is provided by the Coriolis force. The Trojan asteroids are found to cluster around L_4 and L_5 even though they are subject to a dissipative force due to nebular drag and hence to dissipation induced instability.

3.4. Impossibility of gyroscopic stabilization in source-free regions

In the previous section we have seen that whether gyroscopic stabilization is possible depends on the sign of $\text{tr } \mathcal{Q}$. In order to calculate the trace Hessian, we need to compute its diagonal elements. These are given by

$$\frac{\partial^2 U}{\partial \rho^2} = \frac{\partial^2 \Phi}{\partial \rho^2} - \omega \frac{\partial^2 \psi}{\partial \rho^2} + 3\omega^2 + 4\omega B_z + B_z^2 \quad \text{and} \quad \frac{\partial^2 U}{\partial z^2} = \frac{\partial^2 \Phi}{\partial z^2} - \omega \frac{\partial^2 \psi}{\partial z^2} + B_\rho^2, \quad (16)$$

where the poloidal magnetic field components are $B_\rho = -\rho^{-1}\partial\psi/\partial z$ and $B_z = \rho^{-1}\partial\psi/\partial\rho$. Adding up eq. (16) yields the trace of the Hessian. The resulting expression can be simplified with the help of Gauss' law $\nabla \cdot \mathbf{E} = \varrho^s$, i.e.

$$\frac{\partial^2\Phi}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial\Phi}{\partial\rho} + \frac{\partial^2\Phi}{\partial z^2} = -\varrho^s \quad (17)$$

and Ampère's law $\nabla \times \mathbf{B} = \mathcal{J}^s$, whose toroidal component is given by

$$\frac{\partial^2\psi}{\partial\rho^2} - \frac{1}{\rho}\frac{\partial\psi}{\partial\rho} + \frac{\partial^2\psi}{\partial z^2} = -\rho\mathcal{J}_\varphi^s. \quad (18)$$

In the plasma physics literature, the differential operator acting on ψ in eq. (18) is known as the Grad-Shafranov operator (see e.g. Almaguer *et al.* 1988). With eqs. (17) and (18), evaluating the trace at equilibrium yields

$$\text{tr } \mathcal{Q} = \omega^2 + (\omega + B_z)^2 + B_\rho^2 - \varrho^s + \rho\omega\mathcal{J}_\varphi^s, \quad (19)$$

where we have used

$$\frac{\partial\Phi}{\partial\rho} - \omega\frac{\partial\psi}{\partial\rho} = \rho\omega^2 \quad \text{and} \quad \frac{\partial\Phi}{\partial z} - \omega\frac{\partial\psi}{\partial z} = 0. \quad (20)$$

The first three terms on the right hand side of eq. (19) are all non-negative. Thus, in source-free regions, where $\varrho^s = 0$ and $\mathcal{J}_\varphi^s = 0$, the effective potential has no local maxima and gyroscopic stabilisation cannot occur. We note that matter distributions make a *negative* contribution to ϱ^s and hence a positive contribution to $\text{tr } \mathcal{Q}$. We also note that poloidal currents, i.e. sources of the toroidal magnetic field, do not enter eq. (19) at all.

In addition to physical sources in the form or matter and charge distributions there may also be fictitious sources that arise in a rotating frame. For instance, associated with the centrifugal force is a fictitious charge distribution $\varrho^s = 2\Omega^2$. That being said, in appendix A we show that the stability of circular orbits is not affected by a transformation to a rotating frame of reference.

4. MOTION IN A ROTATING MAGNETOSPHERE

As an example of astrophysical relevance where gyroscopic stability by a toroidal magnetic field is possible we analyse the stability of circular orbits in the classical problem of a rotating magnetosphere. We consider the same model studied by Howard *et al.* (1999, 2000) and Dullin *et al.* (2002), where the poloidal magnetic field and the planetary gravitational potential are due to a point dipole and a point mass, respectively.

The poloidal flux function is given by

$$\psi = \frac{\gamma\rho^2}{r^3}, \quad (21)$$

where r is the spherical radius. The parameter γ is proportional to the product of the charge to mass ratio and the dipole strength. Without loss of generality we take the magnetic dipole to point along the z -axis. With this convention, $\gamma > 0$ for positive charges. The poloidal magnetic field components $B_\rho = -\rho^{-1}\partial\psi/\partial z$ and $B_z = \rho^{-1}\partial\psi/\partial\rho$ derived from eq. (21) are

$$B_\rho = \frac{3\gamma\rho z}{r^5} \quad \text{and} \quad B_z = \frac{\gamma(2z^2 - \rho^2)}{r^5}. \quad (22)$$

The scalar potential is

$$\Phi = -\frac{\mu}{r} + \Omega\psi. \quad (23)$$

The first term is the gravitational potential. The second term arises from the requirement that the electric field vanishes in a frame rotating with the planetary rotation rate Ω .

The current density that derives from eq. (21) via Ampère's law (18) vanishes away from the origin. However, the charge density that derives from eq. (23) via Gauss' law (17) does not vanish and is given by

$$\varrho^s = -2\Omega B_z \quad (24)$$

for $r > 0$. This charge density, known in the astrophysical literature as the Goldreich-Julian charge density (Goldreich and Julian 1969), is distributed continuously throughout space. Because of this, positivity of the trace in eq. (19) is no longer ensured.

In order to assess what orbits can be subject to gyroscopic stabilisation, we analyse the characteristics of equilibria, focusing our attention on the regions of parameter space associated with negative $\text{tr } \mathcal{Q}$. The calculations involved in obtaining $\det \mathcal{Q}$ and $\text{tr } \mathcal{Q}$ in eq. (12) have already been carried out in Dullin *et al.* (2002). For the reader's convenience, and because we use different notation, we restate the results that are relevant to the discussion here.

Using eq. (21) and eq. (23) we obtain the location of the equilibrium circular orbits by requiring that the first variation of the effective potential U in eq. (6) vanishes, i.e. by requiring

$$\frac{\partial U}{\partial \rho} = \frac{\rho}{r^5} \left[\mu r^2 + 3\gamma\rho^2(\omega - \Omega) - 2\gamma(\omega - \Omega)r^2 - \omega^2 r^5 \right] = 0, \quad (25)$$

$$\frac{\partial U}{\partial z} = \frac{z}{r^5} \left[\mu r^2 + 3\gamma\rho^2(\omega - \Omega) \right] = 0. \quad (26)$$

Setting to zero each of the two factors on eq. (26) leads to equatorial and halo orbits respectively. We analyse each of these cases separately.

4.1. Equatorial orbits

Equatorial orbits lie within the plane $z = 0$. Their radial location is obtained from eq. (25) with $r = \rho$. The result is

$$\rho^3 = \frac{\mu + \gamma(\omega - \Omega)}{\omega^2}. \quad (27)$$

The determinant and trace of the Hessian evaluated at equatorial equilibria are given by

$$\det \mathcal{Q} = -\frac{1}{2\omega^2\rho^9} \left[3\gamma(\omega - \Omega) + \mu \right] \left[(2\omega\gamma - \Omega\gamma + \mu)^2 - 3(\mu - \gamma\Omega)^2 \right] \quad (28)$$

and

$$\text{tr } \mathcal{Q} = \frac{1}{\omega^2\rho^6} \left[\gamma^2(\omega - 2\Omega)^2 + 2\gamma\mu(\omega - 3\Omega) + 2\mu^2 \right]. \quad (29)$$

From these expressions we can easily compute the zeros of both $\det \mathcal{Q}$ and $\text{tr } \mathcal{Q}$ and characterise the various regions of parameter space according to their stability properties. The results are illustrated in fig. 2, which is almost identical to Figure 7 in Dullin *et al.* (2002), except that we also plot the curve along which the trace vanishes.

In addition to the two energetically stable regions identified already by Dullin *et al.* (2002), there is now a region spanned by orbits corresponding to particles with positive charge ($\gamma > 0$) in prograde rotation ($\omega > 0$). These orbits can be stabilized via gyroscopic effects provided a sufficiently strong toroidal field is present. We remark, however, that this is unlikely to be the case as the toroidal field is typically anti-symmetric about the equator in realistic magnetospheres (see e.g. Bunce and Cowley 2001).

4.2. Halo orbits

The coordinates for circular halo orbits are obtained by setting to zero the second term in eq. (26) and using this result in eq. (25). This leads to

$$r^3 = -\frac{2\gamma(\omega - \Omega)}{\omega^2} \quad \text{and} \quad \sin^2 \vartheta = -\frac{\mu}{3\gamma(\omega - \Omega)}, \quad (30)$$

where ϑ is the angle subtended between the radius vector \mathbf{r} and the z -direction such that $\rho = r \sin \vartheta$ and $z = r \cos \vartheta$.

The determinant and trace of the Hessian evaluated at halo equilibria are given by

$$\det \mathcal{Q} = -\frac{16\gamma^2\mu}{3\omega^4r^{12}} [\omega^2 - 4\Omega\omega + \Omega^2] [3\gamma(\omega - \Omega) + \mu] \quad (31)$$

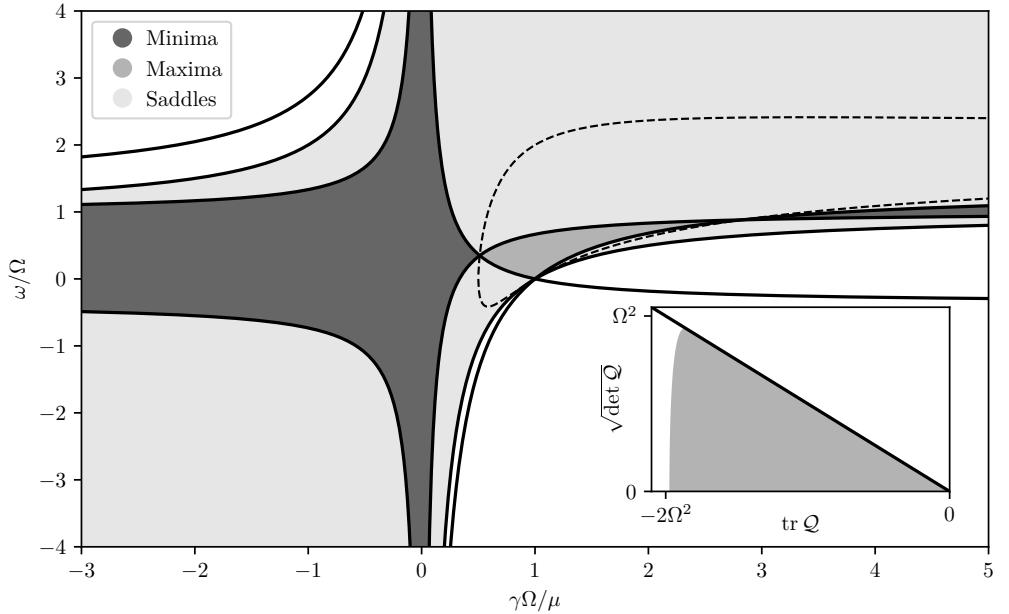


Figure 2. Stability diagram for equatorial orbits. The region where equatorial orbits exist are shaded light gray. Regions of formal stability are shaded dark gray. The only difference with respect to Dullin *et al.* (2002) is that we also indicate (by the intermediate shade of gray) the region where gyroscopic stabilization can in principle occur. This figure is consistent with eqs. (19) and (24) in that, at the equator, $B_z < 0$ for $\gamma > 0$ and thus the term $+2\Omega B_z$ makes a negative contribution to the trace. Note that $B_z < 0$ at the equator corresponds to a rotating dipole for which the magnetic moment and the angular velocity are aligned. The inset shows the region spanned by the orbits for which gyroscopic stabilization is possible in the parameter space defined in fig. 1. Note that a toroidal magnetic field $B_\varphi \geq \sqrt{2}\Omega$ provides gyroscopic stabilization for all the orbits for which this is possible.

and

$$\text{tr } \mathcal{Q} = \frac{2\gamma^2}{\omega^4 r^9} [\mu(3\omega^2 - 4\Omega^2) - 4\gamma(\omega - \Omega)(\omega - 2\Omega)^2]. \quad (32)$$

The regions of stability that derive from determining the signs of both $\det \mathcal{Q}$ and $\text{tr } \mathcal{Q}$ are illustrated in fig. 3. This figure is similar to Figure 9 in Dullin *et al.* (2002), except that we also plot the curve along which the trace vanishes. Close inspection of this figure, see in particular the inset, reveal that there are no regions where $\det \mathcal{Q} > 0$ and $\text{tr } \mathcal{Q} < 0$. We thus conclude that gyroscopic stabilization of halo orbits via a toroidal magnetic field is impossible in an aligned dipolar magnetosphere.

5. SUMMARY AND DISCUSSION

Howard (1999) presented an overview of the stability of circular orbits in combined axisymmetric gravitational and poloidal magnetic fields. In this case, orbital stability is completely determined by the effective potential, which contains all the dynamical effects arising from the magnetic field. Under these assumptions, because the kinetic energy is positive definite, equilibria are stable if and only if they minimize the effective potential.

In this paper, we have generalized this problem by including an axisymmetric toroidal magnetic field, which cannot be included in the effective potential. We pointed out that, unlike the poloidal field component, the toroidal magnetic field does no work in the reduced phase space, i.e., the magnetic force associated with it is gyroscopic³. Thus even

³ We point out that forces that are gyroscopic in three-dimensional space do not necessarily act as gyroscopic forces in the reduced space. This is indeed the case for a purely poloidal axisymmetric magnetic field. We also note the Lagrangian can acquire gyroscopic terms in the reduced space even in the absence of gyroscopic forces in three-dimensional space.

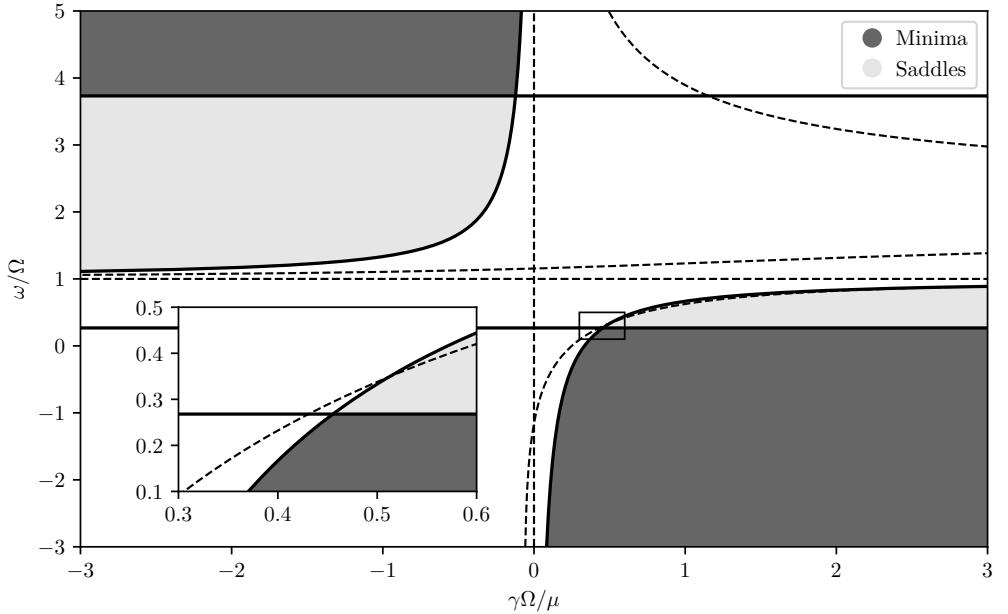


Figure 3. Stability diagram for halo orbits. This is analogous to Figure 9 in Dullin *et al.* (2002). The dark regions are stable with $\det Q > 0$ and $\text{tr } Q > 0$. The sign of the determinant changes across solid lines and that of the trace across dashed lines. The solid hyperbola $\omega = \Omega - \mu/(3\gamma)$ is also the boundary of the region where circular halo orbits exist. Close inspection of the figure — especially the inset — reveals that there is a small region where (a) halo orbits exist and (b) the trace is negative. But the determinant is also negative in that region. The solid horizontal lines $\omega/\Omega = 2 \pm \sqrt{3}$ thus mark the transition from local minima to saddle points. Gyroscopic stabilization is impossible because no halo orbits exist that correspond to local maxima of the effective potential.

thought the toroidal field does not influence the location of the circular orbits it can alter their stability properties by enabling gyroscopic stability.

Absent dissipation, we carried out a spectral stability analysis and determined the conditions for gyroscopic stabilization by an axisymmetric toroidal magnetic field and summarized our results in fig. 1. We showed that, given a circular orbit in an isolated local maxima of the effective potential, it is always possible to find a sufficiently strong axisymmetric toroidal magnetic field to gyroscopically stabilize it. Making use of the KAM theorem we concluded that gyroscopic stability holds in the Lyapunov sense. In real systems, dissipative processes prevent gyroscopic stability from being truly realized. Nevertheless, this type of dissipation induced instabilities evolve on timescales that are much longer than the growth rates of instabilities that would operate in the absence of gyroscopic stabilization. We thus argue that gyroscopic stabilization should be relevant.

We showed that the effective potential associated with combined axisymmetric gravitational and poloidal magnetic fields does not present isolated local maxima in source free regions, thus implying that gyroscopic stabilisation is impossible in this case. As an example where sources are present, we considered a rotating, aligned magnetosphere and investigated the effects of a toroidal magnetic field. This is a generalization of the problem investigated by Howard, who provided a detailed account of the stability properties of equatorial (Howard *et al.* 1999) and halo (Howard *et al.* 2000) orbits of charged dust-particles for an aligned, rotating dipole. We found that there are no equilibrium halo orbits that can be subject to gyroscopic stabilisation. We also found, however, that there do exist prograde equatorial orbits for positive charges for which a toroidal magnetic field that does not vanish at the equator provides gyroscopic stabilization.

Appendix A: Circular orbits in a rotating frame

Let us carry out a coordinate transformation to a frame rotating with a constant frequency Ω around the z -axis. The angular velocity in the rotating frame is

$$\omega' = \omega - \Omega. \quad (\text{A1})$$

The transformed scalar potential Φ' and the poloidal flux function ψ' are respectively given by

$$\Phi' = \Phi - \Omega\psi - \frac{1}{2}\rho^2\Omega^2, \quad (\text{A2})$$

and

$$\psi' = \psi + \rho^2\Omega. \quad (\text{A3})$$

The second term on the right hand side of eq. (A2) arises because the electrostatic potential transforms as the temporal component of an ultra space-like four-vector ($A'_0 = A_0 - \mathbf{v} \cdot \mathbf{A}/c$). The last terms in eqs. (A2) and (A3) account for the centrifugal force and the Coriolis force, respectively.

From eq. (A3) it follows that the generalized angular momentum, defined in eq. (2), is invariant, i.e.

$$p'_\varphi = p_\varphi. \quad (\text{A4})$$

Given the transformations in eqs. (A2) to (A4), the effective potential transforms according to

$$U' = U - \Omega p_\varphi. \quad (\text{A5})$$

This is obviously consistent with eq. (A1) since $\omega = \partial U / \partial p_\varphi$. From eq. (A5) it follows immediately that

$$\delta U' = \delta U \quad \text{and} \quad \delta^2 U' = \delta^2 U \quad (\text{A6})$$

for arbitrary variations $\delta\rho$ and δz . This means that neither the location of circular orbits nor their energetic stability is affected by a transformation to a rotating frame of reference.

In the rotating frame, the components of the poloidal electric field are given by

$$\begin{aligned} E'_\rho &= E_\rho + \rho\Omega B_z + \rho\Omega^2 \\ E'_z &= E_z - \rho\Omega B_\rho. \end{aligned} \quad (\text{A7})$$

The second term on each right hand side arise simply from a Galilean transformation with relative velocity $\rho^2\Omega\nabla\varphi$. The last term in eq. (A7) is the centrifugal force. The components of the transformed poloidal magnetic field are

$$\begin{aligned} B'_\rho &= B_\rho \\ B'_z &= B_z + 2\Omega, \end{aligned} \quad (\text{A8})$$

which include the Coriolis force. The toroidal component of the magnetic field B_φ is invariant. In light of eqs. (14) and (A6) this means that the spectral stability of circular orbits is not affected either by a transformation to rotating frame of reference.

From Gauss' law (17) and Ampère's law (18) it follows that

$$\varrho^{s'} = \varrho^s - \rho\Omega\mathcal{J}_\varphi^s + 2\Omega B_z + 2\Omega^2 \quad (\text{A9})$$

and $\mathcal{J}_\varphi^{s'} = \mathcal{J}_\varphi^s$. It is easy to check that these transformations together with eqs. (A1) and (A8) indeed leave the trace of the Hessian as given in eq. (19) invariant.

Equation (A9) shows that regions of space that are source-free in an inertial frame are not source-free in a rotating frame. We stress, however, that the last term in eq. (A9) is not a physical source of either the electromagnetic or gravitational fields, but is a fictitious charge density that derives from the centrifugal potential via Gauss' law (17).

ACKNOWLEDGMENTS

We thank Pablo Benítez-Llambay, Luis García-Naranjo and Jihad Touma for insightful comments. We are grateful for the hospitality of the Institute for Advanced Study where part of this work was carried out. The research leading to these results has received funding from the European Research Council (ERC) under the European Union's Seventh Framework programme (FP/2007–2013) under ERC grant agreement No 306614.

- J. E. Howard, Celestial Mechanics and Dynamical Astronomy **74**, 19 (1999).
- J. M. Jauch, Helvetica Physica Acta **37**, 284 (1964).
- P. Roman and J. P. Leveille, Journal of Mathematical Physics **15**, 1760 (1974).
- W. Thomson and P. G. Tait, *Treatise on Natural Philosophy (Vol 1)* (Cambridge University Press, Cambridge, 1883).
- W. Thomson and P. G. Tait, *Treatise on Natural Philosophy (Vol 2)* (Cambridge University Press, Cambridge, 1883).
- V. Rumiantsev, Journal of Applied Mathematics and Mechanics **30**, 1090 (1966).
- D. R. Merkin, *Introduction to the Theory of Stability*, Texts in Applied Mathematics, Vol. 24 (Springer New York, New York, NY, 1996).
- R. G. Littlejohn, Journal of Mathematical Physics **20**, 2445 (1979).
- R. G. Littlejohn, Journal of Mathematical Physics **23**, 742 (1982).
- S. Bolotin and P. Negrini, Nonlinear Differential Equations and Applications NoDEA Springer Monographs in Mathematics, **2**, 417 (1995).
- E. J. Routh, *The Advanced Part of a Treatise on the Dynamics of a System of Rigid Bodies* (Cambridge University Press, Cambridge, 1884).
- J. E. Marsden and A. Weinstein, Reports on Mathematical Physics **5**, 121 (1974).
- D. D. Holm, J. E. Marsden, T. Ratiu, and A. Weinstein, Physics Reports **123**, 1 (1985).
- L. A. Pars, *A Treatise on Analytical Dynamics* (Heinemann, London, 1965).
- L. Salvadori, Rend. Accad. Sci. fis. e math. Soc. nza. lett. ed arti. Napoli **20**, 269 (1953).
- F. Hagedorn, Archive for Rational Mechanics and Analysis **42**, 281 (1971).
- A. Lyapunov, Annales de la Faculté des sciences de Toulouse: Mathématiques **9**, 203 (1907).
- J. G. Malkin, *Theorie der Stabilität einer Bewegung* (Akademie-Verlag, Berlin, 1959).
- N. G. Chetaev, *The Stability of Motion* (Pergamon Press, Oxford, 1961).
- V. V. Rumyantsev and S. P. Sosnitskii, Journal of Applied Mathematics and Mechanics **57**, 1101 (1993).
- D. J. Scheeres, Celestial Mechanics and Dynamical Astronomy **94**, 317 (2006).
- A. Bloch, P. S. Krishnaprasad, J. E. Marsden, and T. Ratiu, Annales de l'Institut Henri Poincaré (C) Analyse Non Linéaire **11**, 37 (1994).
- G. Haller, International Journal of Non-Linear Mechanics **27**, 113 (1992).
- V. I. Arnold, Russian Mathematical Surveys **18**, 85 (1963).
- A. G. Sokolskii, Journal of Applied Mathematics and Mechanics **38**, 741 (1974).
- V. I. Arnold, V. V. Kozlov, and A. I. Neishtadt, *Mathematical Aspects of Classical and Celestial Mechanics*, Encyclopaedia of Mathematical Sciences, Vol. 3 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2006).
- R. Krechetnikov and J. E. Marsden, Reviews of Modern Physics **79**, 519 (2007).
- R. S. MacKay, Physics Letters A **155**, 266 (1991).
- J. A. Almaguer, E. Hameiri, J. Herrera, and D. D. Holm, Physics of Fluids **31**, 1930 (1988).
- J. E. Howard, M. Horányi, and G. R. Stewart, Physical Review Letters **83**, 3993 (1999).
- J. E. Howard, H. R. Dullin, and M. Horányi, Physical review letters **84**, 3244 (2000).
- H. R. Dullin, M. Horányi, and J. E. Howard, Physica D: Nonlinear Phenomena **171**, 178 (2002).
- P. Goldreich and W. H. Julian, The Astrophysical Journal **157**, 869 (1969).
- E. J. Bunce and S. W. H. Cowley, Planetary and Space Science **49**, 1089 (2001).

Paper II

Spectral Analysis of Non-ideal MRI Modes: The Effect of Hall Diffusion

Reprinted from

The Astrophysical Journal, Volume 838, Issue 1, March 2017

doi:10.3847/1538-4357/aa6118

with the permission of IOP Publishing.



Spectral Analysis of Non-ideal MRI Modes: The Effect of Hall Diffusion

Gopakumar Mohandas and Martin E. Pessah

Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, DK-2100, Copenhagen Ø, Denmark,
gopakumar@nbi.ku.dk, mpessah@nbi.ku.dk

Received 2016 December 2; revised 2017 February 11; accepted 2017 February 14; published 2017 March 22

Abstract

The effect of magnetic field diffusion on the stability of accretion disks is a problem that has attracted considerable interest of late. In particular, the Hall effect has the potential to bring about remarkable changes in the dynamical behavior of disks that are without parallel. In this paper, we conduct a systematic examination of the linear eigenmodes in a weakly magnetized differentially rotating gas with a special focus on Hall diffusion. We first develop a geometrical representation of the eigenmodes and provide a detailed quantitative description of the polarization properties of the oscillatory modes under the combined influence of the Coriolis and Hall effects. We also analyze the effects of magnetic diffusion on the structure of the unstable modes and derive analytical expressions for the kinetic and magnetic stresses and energy densities associated with the non-ideal magnetorotational instability (MRI). Our analysis explicitly demonstrates that, if the dissipative effects are relatively weak, the kinetic stresses and energies make up the dominant contribution to the total stress and energy density when the equilibrium angular momentum and magnetic field vectors are anti-parallel. This is in sharp contrast to what is observed in the case of the ideal or dissipative MRI. We conduct shearing box simulations and find very good agreement with the results derived from linear theory. Because the modes under consideration are also exact solutions of the nonlinear equations, the unconventional nature of the kinetic and magnetic stresses may have significant implications for the nonlinear evolution in some regions of protoplanetary disks.

Key words: accretion, accretion disks – instabilities – magnetohydrodynamics (MHD)

1. Introduction

The magnetorotational instability (MRI; Balbus & Hawley 1998), driven by differential rotation and weak magnetic fields, is considered to be the foremost mechanism of linear destabilization in astrophysical disk systems. There has been substantial ongoing interest in studying the effect of magnetic field diffusion on the MRI primarily with a view to understanding protoplanetary disk evolution (Turner et al. 2014). In particular, diffusion mediated by Hall currents has commanded a great deal of attention by virtue of its capacity to pave the way to new avenues of destabilization (Wardle 1999; Balbus & Terquem 2001). Local linear analysis has helped reveal the markedly different character of the unstable dynamics (Wardle 1999; Balbus & Terquem 2001; Wardle & Salmeron 2012) and their fundamental dependence on disk conditions, namely, the relative orientation of the net equilibrium angular momentum and magnetic field vectors and the strength of the Hall currents.

One expects to find vast swathes within a protoplanetary disk that are conducive to the prevalence of significant Hall currents as a result of ion–neutral collisions (Kunz & Balbus 2004; Pandey & Wardle 2008; Armitage 2011). This has provided great impetus in driving efforts to understand the nonlinear evolution of disks influenced by non-ideal effects. A number of local shearing box simulations with Hall diffusion either in isolation or in unison with other non-ideal effects (viz. ohmic and ambipolar diffusion) have been carried out in the recent past (Sano & Stone 2002a, 2002b; Bejaramo 2011; Kunz & Lesur 2013; Bai 2014, 2015; Lesur et al. 2014; Simon et al. 2015). Efforts are currently underway to perform global simulations including the Hall effect and the first among them has already been reported by Béthune et al. (2016).

While the march to conduct ever more sophisticated numerical experiments of a non-ideal MHD disk system strides

onward, certain fundamental aspects, especially those pertaining to the question of angular momentum transport may be beneficially served by a systematic examination of the non-ideal MRI eigenmodes. With this goal in mind, we revisit the local linear analysis of a uniformly magnetized disk with Hall diffusion in the shearing sheet approximation. We adopt the approach of Pessah et al. (2006) and Pessah & Chan (2008) that has previously been employed to thoroughly examine the ideal and dissipative MRI eigenmodes. Here, we carry out an exhaustive analysis of the detailed eigenmode structure of the unstable and oscillatory modes affected primarily by Hall diffusion. As part of our analysis, we determine the mean kinetic and magnetic stresses and energy densities of the non-ideal MRI mode across parameter space. Our work reveals that the relative dominance of the Reynolds and Maxwell stresses as well as the ratio of magnetic to kinetic energy can deviate from that of ideal or dissipative MRI when the background field and angular momentum vector are anti-parallel. These departures depend intimately on the range of length scales involved and may have significant implications for the ensuing turbulence. A detailed analysis of the linear eigenmodes may also find utility in testing and benchmarking numerical algorithms designed to include Hall diffusion.

This paper is organized as follows. In Section 2, we outline the fundamental assumptions and equations. In Section 3, we layout the basic groundwork for our analysis and solve the eigenvalue problem. We then examine the mode properties in detail and provide a physical picture of mode behavior in Section 4. In Section 5, we discuss the properties of the kinetic and magnetic stresses and energy densities for the unstable mode. We present the results of numerical simulations in Section 6 to test the validity of our analytical results and conclude with a summary and discussion of the potential implications in Section 7.

2. Basic Equations and Assumptions

We consider a partially ionized, weakly magnetized, incompressible gas subject to ohmic, Hall, and ambipolar diffusion in the presence of a gravitational field due to a central point mass. While we shall strive to retain generality wherever possible, our primary focus will nevertheless be on characterizing the effect of Hall diffusion on the linear modes.

We work in the shearing sheet approximation (Goldreich & Lynden-Bell 1965) and therefore adopt a frame of reference that co-rotates at a fiducial radius, r_0 , in the midplane of the disk. The shearing sheet frame is defined by the set of cartesian coordinates

$$x = r - r_0, \quad y = r_0(\phi - \Omega_0 t), \quad z = z,$$

where $x/r_0 \sim \varepsilon \ll 1$ and is based on a local expansion of the combined gravitational and centrifugal potentials to first order in ε around the fiducial radius. The angular frequency at the fiducial radius is denoted by Ω_0 and the disk is assumed to be in dominant centrifugal balance with the radial gravitational force. Consequently, all other dynamical state variables are taken to be uniform to lowest order in ε . Ignoring vertical stratification, the incompressible shearing sheet equations are given by

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = 2\mathbf{u} \times \boldsymbol{\Omega}_0 + q\Omega_0^2 \nabla x^2 - \frac{1}{\rho} \nabla \left(P + \frac{B^2}{8\pi} \right) + \frac{(\mathbf{B} \cdot \nabla) \mathbf{B}}{4\pi\rho} + \nu \nabla^2 \mathbf{u}, \quad (1)$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u}_e \times \mathbf{B}) - \frac{c}{\sigma} \nabla \times \mathbf{J}, \quad (2)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (3)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (4)$$

where ρ is the gas density, P is the gas pressure, \mathbf{B} is the magnetic field, σ is the constant electrical conductivity, c is the speed of light, and ν is the constant fluid viscosity. The shear rate q evaluated at the fiducial radius is defined as

$$q = -\frac{\partial \ln \Omega}{\partial \ln r} \Big|_{r=r_0}.$$

Here, \mathbf{u} is the velocity of the neutrals and the electron velocity, \mathbf{u}_e , may be expressed as (Balbus & Terquem 2001)

$$\mathbf{u}_e = \mathbf{u} + (\mathbf{u}_e - \mathbf{u}_i) + (\mathbf{u}_i - \mathbf{u}) = \mathbf{u} - \frac{\mathbf{J}}{en_e} + \frac{\mathbf{J} \times \mathbf{B}}{\gamma_d \rho \rho_i c}, \quad (5)$$

where e is the electron charge, n_e is the electron number density, γ_d is the drag coefficient, and ρ_i is the ion mass density. The current density is given by

$$\mathbf{J} = \frac{c}{4\pi} (\nabla \times \mathbf{B}). \quad (6)$$

Equations (1)–(4) admit $\mathbf{u} = -q\Omega_0 x \hat{\mathbf{y}}$ and $\mathbf{B} = B_0 \hat{\mathbf{z}}$ as a steady-state solution for the velocity and magnetic field.¹ We consider Eulerian perturbations ($\delta\mathbf{u}$, $\delta\mathbf{B}$) to all the fluid variables, which are assumed to depend only on the vertical coordinate and time. Rescaling the Eulerian magnetic field

¹ Note that Equations (1)–(4) are insensitive to the presence of a uniform background toroidal field under axial symmetry.

perturbations to have dimensions of velocity, $\delta\mathbf{b} \equiv \delta\mathbf{B}/\sqrt{4\pi\rho}$, we obtain the following set of linearized equations

$$\frac{\partial \delta u_x}{\partial t} = 2\Omega_0 \delta u_y + v_A \frac{\partial \delta b_x}{\partial z} + \nu \frac{\partial^2 \delta u_x}{\partial z^2}, \quad (7)$$

$$\frac{\partial \delta u_y}{\partial t} = (q - 2)\Omega_0 \delta u_x + v_A \frac{\partial \delta b_y}{\partial z} + \nu \frac{\partial^2 \delta u_y}{\partial z^2}, \quad (8)$$

$$\begin{aligned} \frac{\partial \delta b_x}{\partial t} = v_A \frac{\partial \delta u_x}{\partial z} + \frac{cB_0}{4\pi en_e} \frac{\partial^2 \delta b_y}{\partial z^2} \\ + \left(\frac{c^2}{4\pi\sigma} + \frac{B_0^2}{4\pi\rho\gamma\rho_i} \right) \frac{\partial^2 \delta b_x}{\partial z^2}, \end{aligned} \quad (9)$$

$$\begin{aligned} \frac{\partial \delta b_y}{\partial t} = v_A \frac{\partial \delta u_y}{\partial z} - \frac{cB_0}{4\pi en_e} \frac{\partial^2 \delta b_x}{\partial z^2} - q\Omega_0 \delta b_x \\ + \left(\frac{c^2}{4\pi\sigma} + \frac{B_0^2}{4\pi\rho\gamma\rho_i} \right) \frac{\partial^2 \delta b_y}{\partial z^2}. \end{aligned} \quad (10)$$

We have also defined the equilibrium Alfvén speed as

$$v_A \equiv \frac{B_0}{\sqrt{4\pi\rho_0}}. \quad (11)$$

The constraints of incompressibility, Equation (3), and solenoidality, Equation (4), require that $\delta u_z = \delta b_z = \text{const}$, and we may thus set $\delta u_z = \delta b_z = 0$ without loss of generality. Furthermore, restricting the spatial dependence of the perturbations to the vertical dimension implies that nonlinear terms vanish exactly from Equations (7)–(10). Therefore, even though we refer to the problem at hand as a linear mode analysis, the modes under consideration are expected to be long-lived (Goodman & Xu 1994).

3. Eigenvalue Problem

We conduct the linear analysis by solving the eigenvalue problem defined in the shearing sheet frame. The basic analysis in this setting has been carried out in a number of previous studies (Wardle 1999; Balbus & Terquem 2001; Kunz 2008; Wardle & Salmon 2012). We shall however, closely inspect the characteristics of the linear eigenmodes that will enable us to establish fundamental properties of the mean kinetic and magnetic stresses and energy densities.

Assuming vertically periodic boundary conditions over the domain $[-H, H]$, where $2H$ may be taken to be the vertical extent of the disk, we express the perturbed variables as a Fourier series in z , such that

$$\delta f(z, t) = \sum_{n=-\infty}^{\infty} \hat{\delta f}(k_n, t) \exp(ik_n z), \quad (12)$$

where $k_n = n\pi/H$, n is an integer number and δf represents any of the given Eulerian perturbations.² In what follows, we shall omit the subscript n for the wavenumber as well as the subscript 0 for the equilibrium variables for brevity and convenience.

² For weak magnetic fields, we may approximate $\Delta k = k_{n+1} - k_n \propto \beta^{-1/2}$ and thus consider the distribution of wavenumbers to be approximately continuum even for moderate values of the plasma $\beta \sim O(10^{2-3})$.

The set of Equations (7)–(10) can be expressed more compactly as

$$\frac{\partial}{\partial t} \hat{\delta}(k, t) = \mathbf{L} \hat{\delta}(k, t), \quad (13)$$

where

$$\hat{\delta}(k, t) = [\delta\hat{u}_x \quad \delta\hat{u}_y \quad \delta\hat{b}_x \quad \delta\hat{b}_y]^T, \quad (14)$$

and the linear operator \mathbf{L} is

$$\mathbf{L} = \begin{bmatrix} -\omega_\nu & 2\Omega & i\omega_A & 0 \\ (q-2)\Omega & -\omega_\nu & 0 & i\omega_A \\ i\omega_A & 0 & -\omega_P & -\omega_H \\ 0 & i\omega_A & \omega_H - q\Omega & -\omega_P \end{bmatrix}, \quad (15)$$

which we have expressed entirely in terms of the frequencies defined below

$$\omega_A \equiv kv_A, \quad \text{Alfvén frequency} \quad (16)$$

$$\omega_\nu \equiv k^2\nu, \quad \text{Viscous frequency} \quad (17)$$

$$\omega_P \equiv k^2\eta_P, \quad \text{Pedersen frequency} \quad (18)$$

$$\omega_H \equiv k^2\eta_H, \quad \text{Hall frequency.} \quad (19)$$

Here we have also introduced the Pedersen diffusivity

$$\eta_P = \eta_O + \eta_A \equiv \frac{c^2}{4\pi\sigma} + \frac{B^2}{4\pi\rho\gamma_d\rho_i}, \quad (20)$$

with η_O and η_A denoting the ohmic and ambipolar diffusivities, respectively, as well as the Hall diffusivity

$$\eta_H \equiv \hat{\Omega} \cdot \hat{\mathbf{B}} \frac{cB}{4\pi en_e} = s|\eta_H|. \quad (21)$$

The parameter s assumes the value of ± 1 depending on the value of the scalar product $\hat{\Omega} \cdot \hat{\mathbf{B}}$ in Equation (21).³

The linear operator \mathbf{L} has four eigenvalues, σ_j , and associated eigenvectors, \mathbf{e}_j , that satisfies the eigenvalue equation

$$\mathbf{L}\mathbf{e}_j = \sigma_j \mathbf{e}_j \quad \text{for } j = 1, \dots, 4. \quad (22)$$

\mathbf{L} is a normal operator and therefore its eigenvectors are orthogonal if the associated eigenvalues are non-degenerate. In this case, the eigenvectors of \mathbf{L} constitute a linearly independent basis set and thus any given arbitrary vector $\hat{\delta}$ can be represented as the linear combination

$$\hat{\delta} = \sum_{j=1}^4 a_j \mathbf{e}_j, \quad (23)$$

where a_j are in general complex valued time dependent quantities and may be thought of as the coordinates in the \mathbb{C}_4 space defined by the eigenvectors. Substituting Equation (23) in Equation (13), we obtain

$$a_j(t) = a_j(0)e^{\sigma_j t}. \quad (24)$$

³ With more general wavevectors and angular frequency profiles, the sign of η_H is determined by the quantity $(\mathbf{k} \cdot \boldsymbol{\omega})(\mathbf{k} \cdot \mathbf{B})$, where $\boldsymbol{\omega} = \nabla \times \mathbf{u}$ is the equilibrium vorticity (Kunz 2008).

Therefore

$$\hat{\delta}(k, t) = \sum_{j=1}^4 a_j(0)e^{\sigma_j t} \mathbf{e}_j. \quad (25)$$

3.1. Dispersion Relation and Eigenvalues

The characteristic polynomial derived from the matrix operator \mathbf{L} , given by Equation (15) yields the dispersion relation

$$(\sigma_\nu\sigma_P + \omega_A^2)^2 - 2q\Omega^2(\sigma_P^2 + \omega_A^2) + 4\Omega^2\sigma_P^2 + (\sigma_\nu^2 + \kappa^2)\kappa_H^2 + (4-q)\Omega\omega_H\omega_A^2 = 0. \quad (26)$$

where

$$\kappa = \sqrt{2(2-q)}\Omega \quad \text{and} \quad \kappa_H = \sqrt{\omega_H(\omega_H - q\Omega)}, \quad (27)$$

are the epicyclic and the Hall-epicyclic frequency respectively. Defining κ_H makes it easier to recognize the parallel between the Hall-Shear Instability (Rüdiger & Kitchatinov 2005; Kunz 2008) that occurs when $\kappa_H^2 < 0$ and the well-known Rayleigh instability that is present when $\kappa^2 < 0$. We also use the shorthands,

$$\sigma_\nu = \sigma + \omega_\nu \quad \text{and} \quad \sigma_P = \sigma + \omega_P. \quad (28)$$

The dispersion relation Equation (26) is rather cumbersome to solve analytically when dissipative effects are included. Nevertheless, we sketch the procedure for obtaining the roots below. We begin by converting Equation (26) to depressed form

$$\sigma_\alpha^4 + L\sigma_\alpha^2 + M\sigma_\alpha + N = 0, \quad (29)$$

with the coefficients

$$L = 2(\omega_A^2 - \alpha^2) + \kappa^2 + \kappa_H^2, \quad (30)$$

$$M = -2\alpha(\kappa^2 - \kappa_H^2), \quad (31)$$

$$N = (\omega_A^2 - \alpha^2)^2 + \kappa^2(\omega_A^2 + \alpha^2) - 4\omega_A^2\Omega^2 + \kappa_H^2(\kappa^2 + \alpha^2) + (4-q)\Omega\omega_H\omega_A^2, \quad (32)$$

where $\sigma_\alpha = (\sigma_\nu + \sigma_P)/2$ and $\alpha = (\omega_\nu - \omega_P)/2$.

The solutions of Equation (29) are given by

$$\sigma_\alpha = \pm_a \sqrt{-\Lambda \mp_b \sqrt{\Delta}} \pm_b \frac{M}{4\sqrt{\Delta}}, \quad (33)$$

with

$$\Lambda = \frac{3L}{4} + \frac{Y}{2}, \quad \text{and} \quad \Delta = (Y + L)^2 - N, \quad (34)$$

where a and b in Equation (33) mark the four possible combinations of the \pm signs and y is the solution of the cubic equation

$$\left(Y + \frac{L}{2}\right)[(Y + L)^2 - N] = \frac{M^2}{8}. \quad (35)$$

Provided $Y \neq -L/2$, we may recast Equation (35) as

$$\sqrt{(Y + L)^2 - N} = \frac{M/4}{\sqrt{L/4 + Y/2}}, \quad (36)$$

and substituting in Equation (33), we obtain

$$\sigma_\alpha = \pm_a \sqrt{-\left(\frac{3L}{4} + \frac{Y}{2} \pm_b \frac{M/4}{\sqrt{L/4 + y/2}}\right)} \pm_b \sqrt{\frac{L}{4} + \frac{Y}{2}}. \quad (37)$$

Finally, using the shorthands defined in Equation (28), we obtain the eigenvalues,

$$\sigma = \sigma_\alpha - \frac{\omega_\nu + \omega_p}{2}. \quad (38)$$

Two of the four possible solutions given by Equation (38) are of an oscillatory nature and two are exponentially varying. We derive asymptotic expressions for the eigenvalues in the dissipationless limit ($\omega_\nu = \omega_p = 0$) in Appendix A.

4. The Eigenmodes

The set of normalized eigenvectors of the operator L , Equation (15), can be expressed as

$$\hat{e}_j = \frac{e_j}{|e_j|} \quad \text{for } j = 1, 2, 3, 4, \quad (39)$$

where

$$e_j = \begin{bmatrix} 2\Omega\sigma_p^2 + 2\Omega\kappa_H^2 + \omega_H\omega_A^2 \\ \sigma_p(\sigma_p\omega_\nu + \omega_A^2) + \sigma_\nu\kappa_H^2 \\ i\omega_A(2\Omega\sigma_p - \omega_H\sigma_\nu) \\ i\omega_A(\sigma_p\omega_\nu + \omega_A^2 - 2q\Omega^2 + 2\Omega\omega_H) \end{bmatrix}. \quad (40)$$

The eigenvector components satisfy the following relationship

$$-\frac{e_j^4}{e_j^1} = \frac{e_j^3}{e_j^2} + \frac{\omega_A q \Omega (\omega_p - \omega_\nu)}{e_j^2}, \quad (41)$$

where the superscripts denote the corresponding eigenvector component. In the absence of Hall diffusion $\omega_H \rightarrow 0$, Equation (40) reduces to Equation (48) of Pessah & Chan (2008) and to Equation (32) of Pessah et al. (2006) in the ideal limit, $\omega_\nu = \omega_p = \omega_H = 0$.

In the dissipationless limit, but including Hall diffusion, by multiplying Equation (40) with

$$\frac{i}{\omega_A [2\Omega(q\Omega - \omega_H) - (\sigma_j^2 + \omega_A^2)]}, \quad (42)$$

and using the identity (derived from the dispersion relation)

$$\frac{2\Omega\sigma_j^2 + 2\Omega\kappa_H^2 + \omega_H\omega_A^2}{2\Omega(q\Omega - \omega_H) - (\sigma_j^2 + \omega_A^2)} = \frac{\sigma_j^2 + \omega_A^2 + \kappa_H^2}{2\Omega - \omega_H}, \quad (43)$$

we may recast Equation (40) in the more useful form of

$$e_j = \begin{bmatrix} iF & i\sigma_j F & -\frac{\sigma_j}{G} & 1 \end{bmatrix}^T, \quad (44)$$

where

$$F = (\sigma_j^2 + \omega_A^2 + \kappa_H^2)(2\Omega - \omega_H)^{-1}, \quad (45)$$

$$G = [2\Omega(q\Omega - \omega_H) - (\sigma_j^2 + \omega_A^2)](2\Omega - \omega_H)^{-1}. \quad (46)$$

The physically meaningful perturbation components are then obtained from the real part of the eigenvector as

$$\delta_j(z, t) = \Re[\hat{\delta}(k, t)\exp(ikz)]. \quad (47)$$

Since δ_j is a function of the real spatial variable z and time t , we can draw geometrical meaning from the eigenvector, Equation (47), and construct a physical picture of the mode evolution.

A defining property is the relative orientation of the velocity and magnetic field components associated with the perturbations by taking the scalar product of the two-dimensional vectors defined by $\delta_u = [e_j^1 \ e_j^2]$ and $\delta_b = [e_j^3 \ e_j^4]$, i.e., $\delta_u \cdot \delta_b = u_0 b_0 \cos \theta_j$, where

$$u_0 = \sqrt{|e_j^1|^2 + |e_j^2|^2} \text{ and } b_0 = \sqrt{|e_j^3|^2 + |e_j^4|^2}. \quad (48)$$

4.1. The Oscillatory Eigenmode

The Hall effect is distinct from the other kinds of magnetic diffusion in that the electromotive forces it induces act as a “magnetic-Coriolis” force (Balbus & Terquem 2001). This property leads to the polarization of the oscillatory eigenmodes in a manner akin to that rendered by the kinetic Coriolis force. The only effect that ohmic and ambipolar diffusion has on these modes is to damp the wave amplitude over time. Since the effect of dissipation on the eigenmodes has been studied extensively in Pessah & Chan (2008), we shall focus exclusively on the geometric aspects of the oscillatory modes due to Hall diffusion alone and set $\omega_\nu = \omega_p = 0$ here.

In order to provide a geometrical representation of the modes in physical space, it is useful to first consider the norm of the ratios constructed using the components of Equation (44) below:

$$\left| \frac{e_j^2}{e_j^1} \right|^2 = \left| \frac{e_j^3}{e_j^4} \right|^2 = \frac{|\sigma_j|^2 (2\Omega - \omega_H)^2}{[2\Omega(q\Omega - \omega_H) - (\sigma_j^2 + \omega_A^2)]^2}. \quad (49)$$

Note that we retain the label j to denote the eigenmode here because the unstable modes may also become oscillatory beyond a cut-off wavenumber for certain values of the Hall parameter. Using the dispersion relation, Equation (26), the ratio defined in Equation (49) becomes

$$\left| \frac{e_j^2}{e_j^1} \right|^2 = \left| \frac{e_j^3}{e_j^4} \right|^2 = -\frac{|\sigma_j|^2 \sigma_j^{-2}}{1 + \mu}, \quad (50)$$

where we have defined the quantity

$$\mu = q\Omega \frac{[\sigma_j^2 + \omega_A^2 + 2\Omega(\omega_H - q\Omega)]}{\sigma_j^2 (2\Omega - \omega_H)}. \quad (51)$$

When the mode is purely oscillatory, $-|\sigma_j|^2 \sigma_j^{-2} = 1$ and Equation (50) simply describes an ellipse where both the perturbed velocity and magnetic field components of the eigenvector Equation (44) represent their respective semimajor and minor axes. The eccentricity of the ellipse, ϵ , is related to μ as

$$\epsilon^2 = \begin{cases} |\mu| & \text{if } \mu < 0, \\ \mu/(1 + \mu) & \text{if } \mu > 0. \end{cases} \quad (52)$$

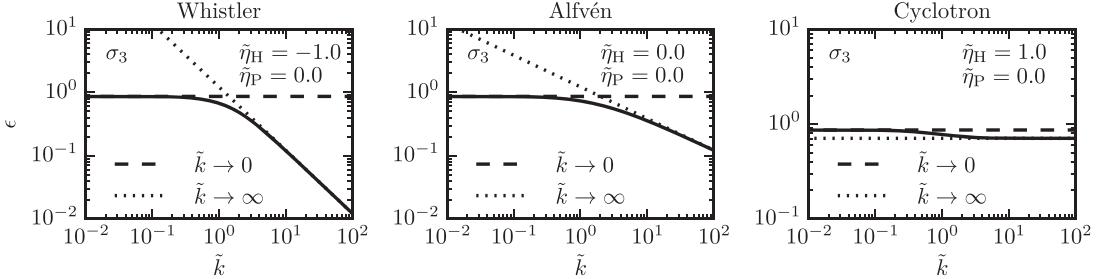


Figure 1. Eccentricity ϵ of the polarized oscillatory mode, $\sigma_3 = i\omega$, as a function of wavenumber for $q = 1.5$ and $\tilde{\eta}_H = -1.0, 0.0, 1.0$. Asymptotic approximations in the low wavenumber limit ($\tilde{k} \rightarrow 0$) and the high wavenumber limit ($\tilde{k} \rightarrow \infty$) are represented by the dashed and dotted lines respectively. The left panel corresponds to a Whistler mode, the central panel corresponds to an Alfvén mode, and the right panel corresponds to a cyclotron mode; all three are subject to the combined influence of rotation and shear (see Appendix A).

With the aid of the asymptotic forms for the eigenvalues, Equations (97)–(98), we can determine the asymptotic behavior of the eccentricity as given below

$$-2 < \tilde{\eta}_H \leq -1/2:$$

$$\sigma_3: \lim_{\tilde{k} \rightarrow 0} \epsilon^2 \sim \frac{q}{2}, \quad \lim_{\tilde{k} \rightarrow \infty} \epsilon^2 \sim \frac{q\Omega}{|\omega_H|}, \quad (53)$$

$$-1/2 < \tilde{\eta}_H < 0:$$

$$\begin{aligned} \sigma_1: \lim_{\tilde{k} \rightarrow \infty} \epsilon^2 &\sim \frac{q\Omega|\omega_H|}{\omega_A^2 + 2\Omega|\omega_H|}, \\ \sigma_3: \lim_{\tilde{k} \rightarrow 0} \epsilon^2 &\sim \frac{q}{2}, \quad \lim_{\tilde{k} \rightarrow \infty} \epsilon^2 \sim \frac{q\Omega}{|\omega_H|}, \end{aligned} \quad (54)$$

$$\tilde{\eta}_H = 0:$$

$$\begin{aligned} \sigma_1: \lim_{\tilde{k} \rightarrow \infty} \epsilon^2 &\sim \frac{q\Omega}{\omega_A}, \\ \sigma_3: \lim_{\tilde{k} \rightarrow 0} \epsilon^2 &\sim \frac{q}{2}, \quad \lim_{\tilde{k} \rightarrow \infty} \epsilon^2 \sim \frac{q\Omega}{\omega_A}, \end{aligned} \quad (55)$$

$$\tilde{\eta}_H > 0:$$

$$\begin{aligned} \sigma_1: \lim_{\tilde{k} \rightarrow \infty} \epsilon^2 &\sim \frac{q\Omega}{\omega_H}, \\ \sigma_3: \lim_{\tilde{k} \rightarrow 0} \epsilon^2 &\sim \frac{q}{2}, \quad \lim_{\tilde{k} \rightarrow \infty} \epsilon^2 \sim \frac{q\Omega|\omega_H|}{\omega_A^2 + 2\Omega\omega_H}. \end{aligned} \quad (56)$$

The eccentricity of the Alfvén and Whistler modes (see Appendix A for mode nomenclature) decreases with increasing wavenumber and the polarization becomes increasingly circular. The eccentricity is generally maximum in the limit of $\tilde{k} \rightarrow 0$, and has the value of $\epsilon_{\max} = \sqrt{q}/2$, which incidentally shares the value of the Oort constant for a differentially rotating disk. The eccentricity of the cyclotron mode (see Appendix A) is only marginally lower than the maximum ϵ_{\max} at large wavenumbers because its frequency is bounded at ω_G , see Appendix A. In Figure 1, we show the three distinct ways in which the eccentricity of the oscillatory mode can vary as a function of the wavenumber with the asymptotic forms derived above to match.

Using Equations (41) and (50), the relative orientation of δ_u and δ_b for the oscillatory modes can be described by the angle

$$\cos \theta_\omega = -\sqrt{\frac{1 - \epsilon^2(k)}{[1 - \epsilon^2(k)\cos^2 \varphi][1 - \epsilon^2(k)\sin^2 \varphi]}}, \quad (57)$$

where $\varphi = kz + \omega t$. In general, θ_ω oscillates in time, so δ_u and δ_b move in and out of phase as φ changes by a factor of $\pi/2$.

Figure 2 charts the evolution of the net velocity vector of the positive branch eigensolutions, σ_1 and σ_3 , over a half period for a fixed wavenumber and two different values of the Hall parameter. Notice that the polarization of σ_1 for $\tilde{\eta}_H = 1$ as well as σ_3 for $\tilde{\eta}_H = -1.5$ is very nearly circular whereas the polarization of σ_3 for $\tilde{\eta}_H = 1$ is visibly elliptical. We also remind the reader that any determination of the direction of polarization (right or left) is to be made by examining the eigenvector, Equation (47). For instance, σ_1 associated with $\tilde{\eta}_H = 1$ is right elliptically polarized whereas σ_3 associated with $\tilde{\eta}_H = -1.5$ is left elliptically polarized even though both behave like a Whistler mode at large wavenumbers.

4.2. The Non-ideal MRI Eigenmode

Here, we examine the properties of the eigenvector corresponding to the non-ideal MRI mode. Closed form expressions are easily derived in the absence of viscous effects and so we set $\omega_v = 0$ hereafter. This would correspond to considering the very low magnetic Prandtl number limit $Pm = \nu/\eta_p \rightarrow 0$, which is also the relevant regime of parameter space with regard to protoplanetary disks.

We express below the main characteristic scales associated with the unstable mode obtained from the dispersion relation, Equation (26) in the inviscid limit (Wardle & Salmeron 2012) and applicable in the parameter space defined by $(\tilde{\eta}_H, \tilde{\eta}_P)$. In what follows, it shall be expedient, on occasion, to use the dimensionless variables

$$\tilde{k} = \frac{k v_A}{\Omega}, \quad \tilde{\eta}_H = \frac{\eta_H \Omega}{v_A^2}, \quad \tilde{\eta}_P = \frac{\eta_P \Omega}{v_A^2}. \quad (58)$$

The critical wavenumber beyond which the non-ideal MRI is cut-off is

$$\tilde{k}_c^2 = \frac{2q[1 + (2 - q)\tilde{\eta}_H]}{1 + (4 - q)\tilde{\eta}_H + 2(2 - q)(\tilde{\eta}_H^2 + \tilde{\eta}_P^2)}. \quad (59)$$

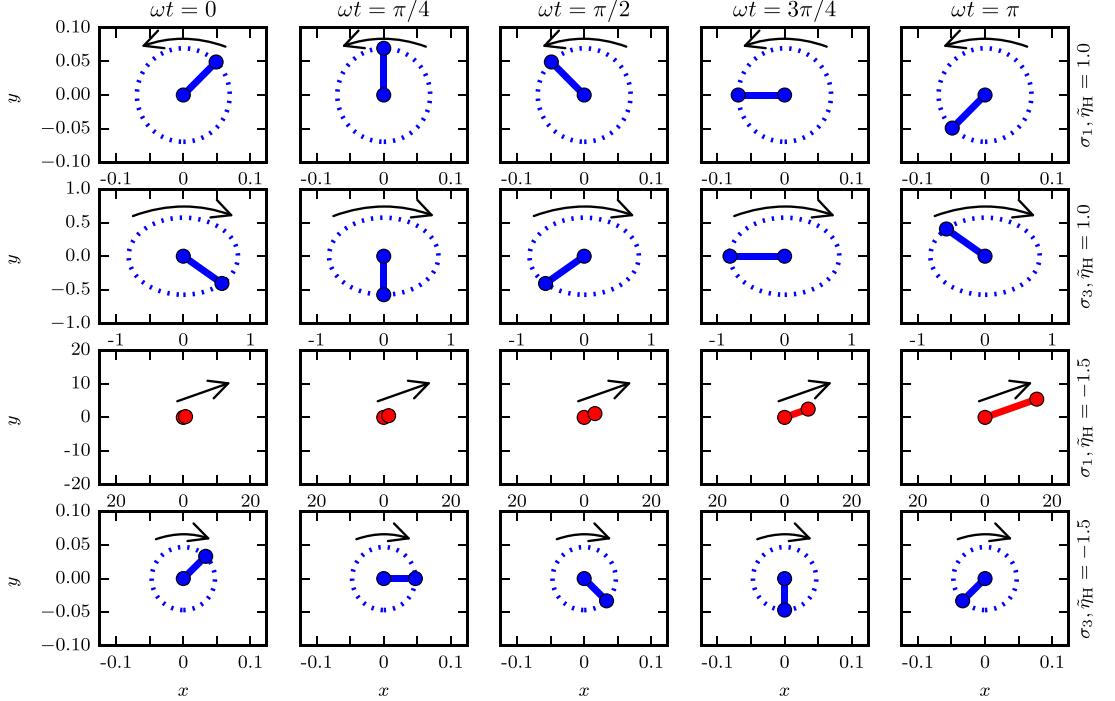


Figure 2. Visualization of the eigenmodes σ_1 and σ_3 in the horizontal plane at $kz = \pi/4$ for $\tilde{k} = 10$, $q = 1.5$ and two different values of the Hall diffusivity, $\eta_H = 1.0, -1.5$. The thick solid line denotes the velocity vector and the arrows indicate the direction of polarization (right or left, see Appendix A) of the corresponding eigenmodes as seen by an observer looking down at the midplane from above. Each row of plots depicts the mode evolution over one half period in time.

A suitable combination of the Pedersen and Hall diffusivities can lead to $\tilde{k}_c \rightarrow \infty$. This occurs when the denominator in Equation (59) vanishes

$$1 + (4 - q)\tilde{\eta}_H + 2(2 - q)(\tilde{\eta}_H^2 + \tilde{\eta}_P^2) = 0. \quad (60)$$

The wavenumber at which the growth rate is maximum is

$$\tilde{k}_m^2 = \frac{-2\gamma_m^2[\gamma_m^2 + 2(2 - q)]}{2\gamma_m^2 - 2q - [\gamma_m^2 + 2(2 - q)](q\tilde{\eta}_H - 2\gamma_m\tilde{\eta}_P)}, \quad (61)$$

and the maximum growth rate γ_m , normalized by the angular frequency, satisfies

$$\tilde{\eta}_H = \frac{16q\tilde{\eta}_P\gamma_m}{4q^2 - 16\gamma_m^2} - \frac{2}{\gamma_m^2 + 2(2 - q)}. \quad (62)$$

In a portion of the parameter space defined by $(\tilde{\eta}_H, \tilde{\eta}_P)$, the maximum growth rate is reached asymptotically as the wavenumber approaches infinity and the denominator of Equation (61) vanishes. The growth rate in this region is obtained by solving

$$[\gamma_m^2 + 2(2 - q)](\tilde{\eta}_H^2 + \tilde{\eta}_P^2) + [2\tilde{\eta}_P\gamma_m + (4 - q)\tilde{\eta}_H] + 1 = 0. \quad (63)$$

This regime will be the subject of greater discussion in the following section.

Let us now examine how the planes containing the velocity and magnetic vectors δ_u and δ_b associated with the unstable mode are oriented relative to each other. Using Equation (41), we find

$$\cos \theta_\gamma \equiv \frac{e_\gamma^1 e_\gamma^3 + e_\gamma^2 e_\gamma^4}{b_0 u_0} = -\frac{\omega_A \omega_P q \Omega e_\gamma^1}{b_0 u_0}. \quad (64)$$

In the absence of dissipation, $\omega_P \rightarrow 0$, $\theta_\gamma = \pi/2, 3\pi/2$, and δ_u and δ_b are orthogonal to each other. Additionally, the angle ψ subtended by the velocity vector with respect to the x axis in the xy plane is simply given by $\tan \psi = |e_\gamma^2|/|e_\gamma^1|$.

Figure 3 illustrates δ_u and δ_b projected onto the midplane of the disk for four representative values of the Hall diffusivity $\tilde{\eta}_H$, for a fixed wavenumber k , with and without dissipation ω_P . The angle ψ becomes smaller with increasingly negative values of the Hall parameter, $\tilde{\eta}_H$. This is shown graphically in Figure 4 for the wavenumber \tilde{k}_m at which the growth rate of the ideal MRI is maximum. One can also see that the velocity and magnetic vectors are not quite orthogonal when $\omega_P \neq 0$ (Pessah & Chan 2008).

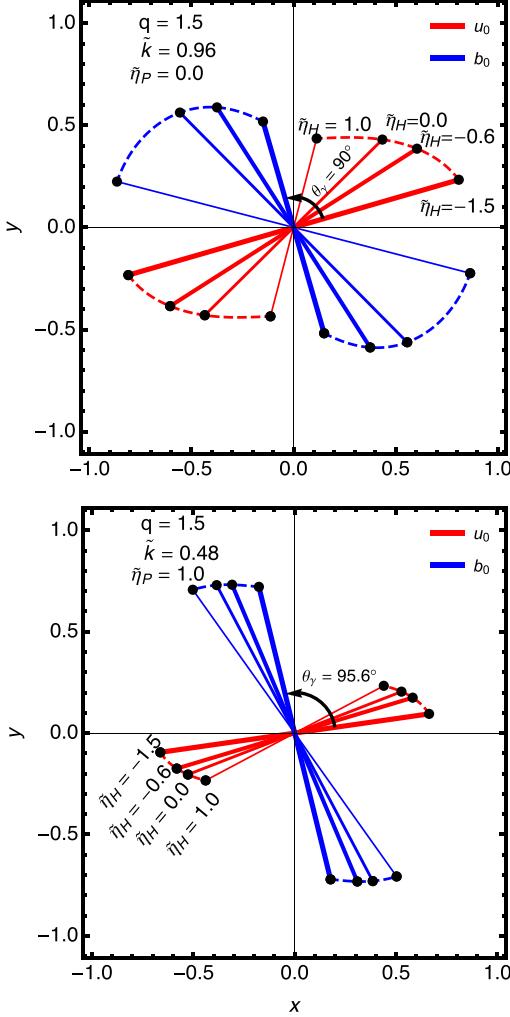


Figure 3. Geometrical representation of the velocity field δ_b and the magnetic field δ_b projected on the horizontal plane of the disk for different values of the Hall parameter, η_H . The top panel presents the case without dissipation $\eta_p = 0$, evaluated at $\tilde{k} = 0.96$ and the bottom panel illustrates the case with magnetic dissipation $\eta_p = 1.0$, evaluated at the wavenumber $\tilde{k} = 0.48$. A general trend that one observes is for the velocity vector to lean in toward the positive x axis and for the magnetic vector to lean in toward the positive y axis with an increasingly negative Hall parameter. The magnetic and velocity vector are, however, only orthogonal to each other in the dissipationless limit and when $Pm = 0$.

Finally, the ratio of the magnitudes of the magnetic vector to the velocity vector, b_0/u_0 , can also be computed from the eigenvector components Equation (40). Figure 5 measures this ratio as a function of wavenumber for different values of the Hall parameter. We find that this ratio becomes lower than unity, implying that the magnetic perturbation is weaker in comparison to the velocity perturbation when $\eta_H < 0$ and for a very large range of wavenumbers with $\eta_p < 1$. This feature will

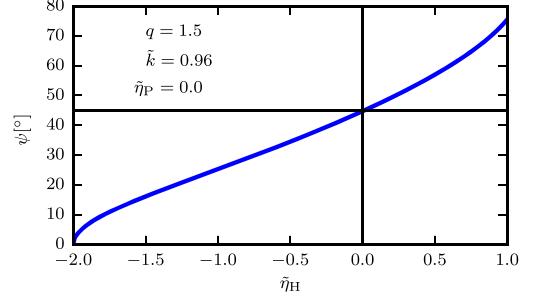


Figure 4. Angle ψ between the velocity vector δ_b and the x axis plotted as a function of the Hall parameter η_H . The angle is calculated from the eigenvector components evaluated at the wavenumber, \tilde{k}_m for the ideal MRI.

be of particular interest with regard to the transport stresses of the non-ideal MRI unstable mode and will be explored further in the following section.

5. Kinetic and Magnetic Stresses and Energy Densities

We now use the results of the eigenmode analysis to ascertain the properties of the mean kinetic and magnetic stresses and energy densities. In particular, we focus on the xy component of the Reynolds and Maxwell stresses of the MRI mode. We define the mean Reynolds and Maxwell stresses as

$$R_{ij}(t) = \overline{\delta u_i(z, t) \delta u_j(z, t)} \text{ and } M_{ij}(t) = \overline{\delta b_i(z, t) \delta b_j(z, t)}, \quad (65)$$

where the over-line denotes the vertical average over the domain $[-H, H]$. In terms of their Fourier components, the stress components are given by (see Pessah et al. 2006 for the derivation)⁴

$$R_{ij}(t) \equiv 2 \sum_{n=1}^{\infty} \Re[\hat{\delta u}_i(k_n, t) \hat{\delta u}_j^*(k_n, t)], \quad (66)$$

$$M_{ij}(t) \equiv 2 \sum_{n=1}^{\infty} \Re[\hat{\delta b}_i(k_n, t) \hat{\delta b}_j^*(k_n, t)]. \quad (67)$$

The xy component of the Reynolds and Maxwell stress tensor associated with the MRI unstable eigenmode are

$$R_{xy}(t) = 2 \sum_{n=1}^{\infty} \mathcal{R}_{xy}(k_n) e^{2\sigma(k_n)t}, \quad (68)$$

$$M_{xy}(t) = 2 \sum_{n=1}^{\infty} \mathcal{M}_{xy}(k_n) e^{2\sigma(k_n)t}, \quad (69)$$

where

$$\mathcal{R}_{xy}(k_n) = \frac{\Re[e_{\gamma}^1 e_{\gamma}^{2*}]}{\|e_{\gamma}\|^2}, \quad (70)$$

$$\mathcal{M}_{xy}(k_n) = \frac{\Re[e_{\gamma}^3 e_{\gamma}^{4*}]}{\|e_{\gamma}\|^2}. \quad (71)$$

⁴ In order to keep track of the various modes contributing to the mean values, we restore the index wavenumber n throughout this section.

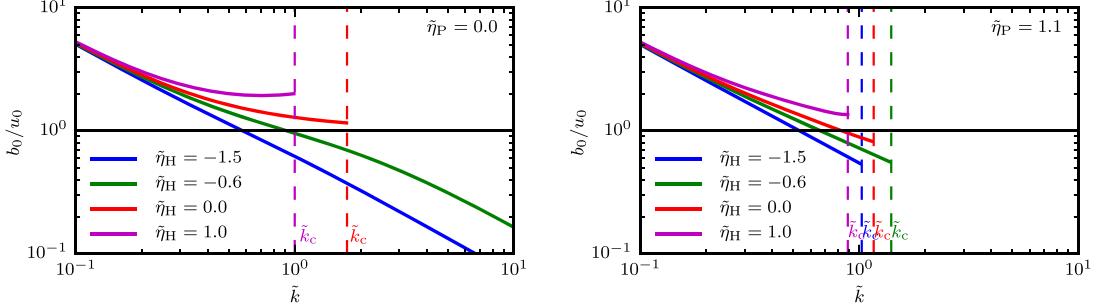


Figure 5. Ratio b_0/u_0 of the MRI unstable eigenmode for different values of the Hall parameter $\tilde{\eta}_H$ without dissipation (left) and with dissipation (right). This ratio becomes lesser than unity for $\tilde{\eta}_H < 0$ and it has implications for the relative strengths of the magnetic and kinetic stresses pertinent to angular momentum transport.

The trace of the tensors R_{ij} and M_{ij} gives us the mean kinetic and magnetic energy densities, respectively,

$$E_K(t) = 2 \sum_{n=1}^{\infty} \mathcal{E}_K(k_n) e^{2\sigma(k_n)t}, \quad (72)$$

$$E_M(t) = 2 \sum_{n=1}^{\infty} \mathcal{E}_M(k_n) e^{2\sigma(k_n)t}, \quad (73)$$

where

$$\mathcal{E}_K(k_n) = \frac{\mathcal{R}_{xx}(k_n) + \mathcal{R}_{yy}(k_n)}{2}, \quad (74)$$

$$\mathcal{E}_M(k_n) = \frac{\mathcal{M}_{xx}(k_n) + \mathcal{M}_{yy}(k_n)}{2}. \quad (75)$$

The quantities \mathcal{R}_{xy} , \mathcal{M}_{xy} , \mathcal{E}_K , and \mathcal{E}_M represent the contribution of each mode k to the mean values of the corresponding functions (Pessah et al. 2006).

The ratio of the xy components of Maxwell stress to the Reynolds stress is a non-trivial function of k_n . In the ideal limit (with $\omega_\nu = \omega_p = \omega_h = 0$), using the dispersion relation, one can easily see that $\mathcal{M}_{xy} > \mathcal{R}_{xy}$ for the full range of unstable modes, k_n . In the dissipationless limit, where $\omega_\nu = \omega_p = 0$ but $\omega_h \neq 0$, this ratio reduces to

$$\frac{-\mathcal{M}_{xy}(k_n)}{\mathcal{R}_{xy}(k_n)} = \frac{\omega_A^2 (2\Omega - \omega_H)^2}{[\gamma^2(k_n) + \omega_A^2 + \kappa_H^2]^2}. \quad (76)$$

Interestingly, the ratio defined in Equation (76) is only greater than unity if

$$\tilde{k}_n^2 < \tilde{k}_i^2 = \frac{(q-2)}{\tilde{\eta}_H}. \quad (77)$$

The wavenumber \tilde{k}_i is purely imaginary if $\tilde{\eta}_H > 0$ and is infinite if $\tilde{\eta}_H = 0$. However, when $\tilde{\eta}_H < 0$ and $q < 2$, \tilde{k}_i is finite and real valued. This implies that there is a range of unstable wavenumbers for which $\mathcal{R}_{xy} > \mathcal{M}_{xy}$. It is rather difficult to derive an equivalent expression for \tilde{k}_i in closed form with $\omega_p \neq 0$ since this would require solving a quartic equation in k . However, numerical calculations hint at the presence of such a scale with dissipative effects present as well and we comment further on this in the following section. As we shall discuss below, the potential for a role-reversal of the dominant stress components are directly tied to the exact nature of the unstable mode in different parts of the parameter space.

The characteristic variables that specify the wavenumber at which the growth rate is quenched \tilde{k}_c , and the wavenumber at which the growth rate is maximum \tilde{k}_m , divides the parameter space defined by $(\tilde{\eta}_H, \tilde{\eta}_P)$ into three regions I, II, and III as described in Wardle & Salmon (2012). Region I is defined by the space outside of a semi-circle in the coordinates $(\tilde{\eta}_H, \tilde{\eta}_P)$ spanning from $(-1/2, 0)$ to $(-2, 0)$. Here the unstable mode has a finite \tilde{k}_c and \tilde{k}_m . The space contained within the aforementioned semi-circular locus and an arc extending from $(\tilde{\eta}_H, \tilde{\eta}_P) = (-4/5, 0)$ to $(-2, 0)$ is designated Region II. Here the unstable mode has a finite \tilde{k}_m but \tilde{k}_c is infinite. Finally, the area enclosing the lower boundary of Region II and the horizontal axis $\tilde{\eta}_H$ is designated as Region III. In this region, both \tilde{k}_c and \tilde{k}_m are infinite. The region $\tilde{\eta}_H < -2$ is stable to the MRI for all values of $\tilde{\eta}_P$. This classification will be useful in specifying the dominant stresses in parameter space as we discuss below.

5.1. Stresses and Energies in Region I

As mentioned above, the MRI growth is cut off at a finite wavenumber in Region I. This implies that the major contributions to Equations (68), (69), (72), and (73) come from a finite range of unstable wavenumbers $n = 1$ to $n = N_c$, where N_c labels the cut-off wavenumber k_c . At late times, the mean stresses and energy densities may then be expressed as

$$R_{xy}(t) = 2 \sum_{n=1}^{N_c} \mathcal{R}_{xy}(k_n) e^{2\gamma(k_n)t} + \dots, \quad (78)$$

$$M_{xy}(t) = 2 \sum_{n=1}^{N_c} \mathcal{M}_{xy}(k_n) e^{2\gamma(k_n)t} + \dots, \quad (79)$$

$$E_K(t) = 2 \sum_{n=1}^{N_c} \mathcal{E}_K(k_n) e^{2\gamma(k_n)t} + \dots, \quad (80)$$

$$E_M(t) = 2 \sum_{n=1}^{N_c} \mathcal{E}_M(k_n) e^{2\gamma(k_n)t} + \dots, \quad (81)$$

with the dots representing oscillatory contributions that we may safely neglect. Within this region of parameter space, it is reasonable to expect that at late times during the linear evolution, the kinetic and magnetic stresses are dominated by contributions linked to the scale \tilde{k}_m . In the dissipationless limit,

we can thus expect

$$\lim_{t\Omega \gg 1} \frac{-M_{xy}}{R_{xy}} \sim \frac{-\mathcal{M}_{xy}}{\mathcal{R}_{xy}} \Big|_{k_m} = \frac{(4-q)[(4-q)\tilde{\eta}_H + 2]}{2q}. \quad (82)$$

Equation (82) trivially reduces to Equation (65) of Pessah et al. (2006) in the ideal MHD limit. Deriving an equivalent analytical expression for the late time stress ratios in the presence of dissipation is tedious but can easily be computed numerically. However, numerical calculations also reveal that a real valued \tilde{k}_i may be present for certain values of $\tilde{\eta}_p$ in Region I and the scales are arranged in the order of $\tilde{k}_m < \tilde{k}_i \lesssim \tilde{k}_c$. Nevertheless, the ratio of the stress components will be dominated by the fastest growing mode, at which one always finds $-\mathcal{M}_{xy} > \mathcal{R}_{xy}$. In the dissipationless limit, \tilde{k}_i is never real valued.

5.2. Stresses and Energies in Regions II and III

The unstable mode grows at a uniform rate for a wide range of wavenumbers that extend infinitely in both Regions II and III. One can therefore derive asymptotic forms of the *per-k* kinetic and magnetic stress energy densities, Equations (70), (71), (74), and (75) as given below

$$\lim_{\tilde{k} \rightarrow \infty} \mathcal{R}_{xy} \sim \frac{\eta_T^2 (2\Omega + v_A^2 \eta_H / \eta_T^2) (\gamma_\infty + v_A^2 \eta_p / \eta_T^2)}{v_A^4 + 2v_A^2 (\gamma_\infty \eta_p + 2\Omega \eta_H) + \eta_T^2 (\gamma_\infty^2 + 4\Omega^2)}, \quad (83)$$

$$\lim_{\tilde{k} \rightarrow \infty} \mathcal{M}_{xy} \sim 0, \quad (84)$$

$$\lim_{\tilde{k} \rightarrow \infty} \mathcal{E}_K \sim \frac{1}{2}, \quad (85)$$

$$\lim_{\tilde{k} \rightarrow \infty} \mathcal{E}_M \sim 0, \quad (86)$$

where $\eta_T^2 = \eta_H^2 + \eta_p^2$ and γ_∞ is the solution to Equation (62) for Region II and Equation (63) for Region III. Using Equations (83) and (85) in Equations (68) and (72), we may then approximate the time dependent *xy* Reynolds stress tensor and kinetic energy density as

$$R_{xy}(t) \approx 2e^{2\gamma_\infty t} \mathcal{R}_{xy}(k \rightarrow \infty) \sum_n l^n, \quad (87)$$

$$E_K(t) \approx e^{2\gamma_\infty t} \sum_n l^n. \quad (88)$$

While the infinite sum in Equations (87) and (88) appear to be a divergent series, it is in fact the Riemann zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s},$$

with $s = 0$ and possesses a finite valued sum of $\zeta(0) = -1/2$ (Hardy 1956). We shall not endeavour to speculate on the implications of this curious feature since an infinite range of scales will never come to pass as the fluid approximation inevitably breaks down. The alternative is no less dramatic in that a finite series would have the sum = N , where N can be rather large.

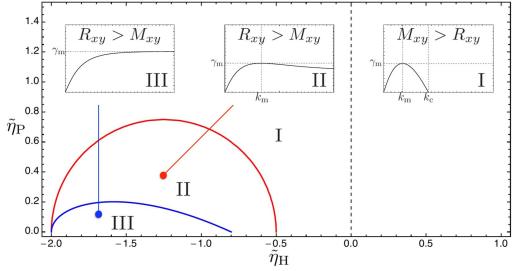


Figure 6. Parameter space defined by $\tilde{\eta}_H$ and $\tilde{\eta}_p$ demarcated into three regions I, II, and III based on the distinct characteristic traits of the MRI for the said range of parameter values. The figure is identical to Figure 5 of Wardle & Salmeron (2012) with the relative strengths of the *xy* kinetic and magnetic stress components additionally specified.

We are thus led to expect

$$\lim_{t\Omega \gg 1} \frac{-M_{xy}}{R_{xy}} \ll 1, \quad (89)$$

with the ratio becoming increasingly smaller the greater the unstable range of wavelengths accounted for. In a real astrophysical system such as a protoplanetary disk, dissipation due to ohmic and ambipolar diffusion may be large enough in some parts of the disk to keep the kinetic stress R_{xy} and energy density E_K , bounded, by suppressing the unstable growth at smaller length scales. Therefore, the dominance of kinetic stresses may go unchallenged unless dissipation forces the instability to operate within Region I, see Figure 6. On the other hand, if one can find parts of the disk where the diffusivities fall within Regions II and III, one should expect the Reynolds stress to dominate. Figure 7 shows the *per-k* kinetic and magnetic *xy* stress component and energy densities in the dissipationless limit for different values of the Hall parameter, $\tilde{\eta}_H$.

6. Comparing Analytical Results with Numerical Simulations

In this section, we present the results of unstratified shearing box simulations with a uniform net vertical field including Hall and ohmic diffusion, performed using the grid-based higher order Godunov MHD code ATHENA (Stone et al. 2008). The Hall effect is implemented in Athena using an operator-split technique (Bai 2014) that is similar to the dimensionally split scheme proposed by O'Sullivan & Downes (2006, 2007). We use the HLLD Riemann solver and a CTU unsplit integrator with third-order reconstruction. The simulations we performed are identical to the test runs reported in Appendix B of Bai (2014).

We adopt an isothermal equation of state and the initial conditions constitute random velocity perturbations of strength, $\delta u/c_s = 10^{-6}$. The default boundary conditions are periodic in y and z and shearing periodic in x . Our simulations were performed with a plasma beta, defined as the ratio of thermal to magnetic pressure $\beta = 800$, background angular frequency $\Omega = 1$, equilibrium density $\rho_0 = 1$, isothermal sound speed $c_s = 1$, and dimensionless shear rate $q = 3/2$. The computational domain has an extent of $L_x \times L_y \times L_z = 0.1H \times 0.1H \times 2H$. We work with the default grid resolution $N_x \times N_y \times N_z = 4 \times 4 \times 256$.

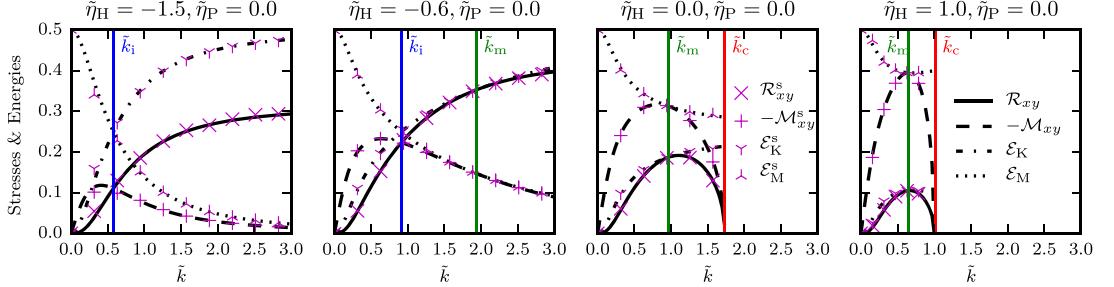


Figure 7. xy components of the per- k Reynolds and Maxwell's stress tensor and the kinetic and magnetic energy densities of the MRI unstable mode, Equations (70), (71), (74), and (75), for different values of the Hall parameter, $\tilde{\eta}_H = -1.5, -0.6, 0.0, 1.0$ and $q = 1.5$ are plotted by the line curves. The discrete markers denote the corresponding values of the said quantities derived from shearing box simulations. Legends with the superscript “ s ” label the corresponding quantity derived from simulation data.

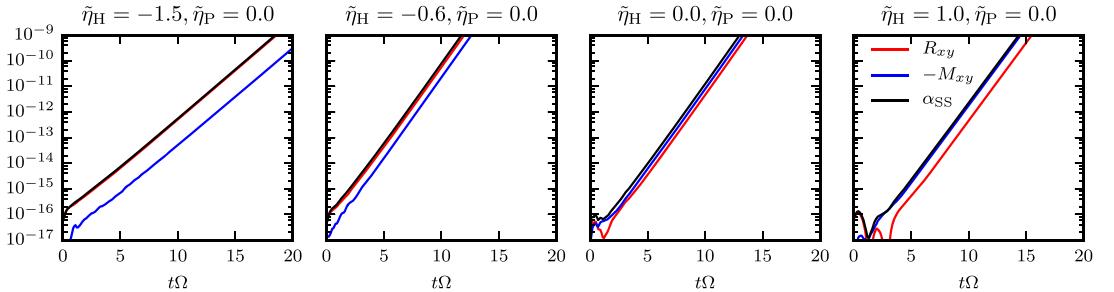


Figure 8. xy components of the Reynolds stress tensor, Maxwell's stress tensor and the Shakura-Sunyaev α_{SS} parameter of the MRI unstable mode for different values of the Hall parameter, $\tilde{\eta}_H = -1.5, -0.6, 0.0, 1.0$ and $q = 1.5$ obtained from shearing box simulations with $N_z = 256$. In accordance with the results of the linear theory, we find that the Reynolds stress dominates over the Maxwell's stress when $\tilde{\eta}_H < 0$.

In order to directly test and compare against the predictions of analytical theory, we run the code by varying the Hall parameter over the different values, $\tilde{\eta}_H = -1.5, -0.6, 0.0, 1.0$ and $\tilde{\eta}_P = 0.0$.⁵ We also perform one additional simulation with the parameters $\tilde{\eta}_H = -1.5$ and $\tilde{\eta}_P = 1.0$. The simulations were run for up to 20 orbits with orbital advection via Fargo enabled. Such short run times suffice for the task at hand since the aim is to test the agreement between our analytical results and the linear evolution of the simulations. We obtain the perturbations, $\delta u_x, \delta u_y, \delta b_x, \delta b_y$, from the Athena output and compute their Fourier transform at time, $t = 11\Omega^{-1}$. We then combine these variables as given by Equations (70), (71), (74), and (75) to obtain the kinetic and magnetic stress components and energies at a given scale.

We have found the simulation and the theoretical results to be in excellent agreement for as many vertical modes, \tilde{k} , as can be reliably resolved. The output of the shearing box simulation conducted with a vertical grid resolution, $N_z = 256$, is overplotted against the values of the corresponding stresses and energy densities obtained from linear theory in Figure 7. Figure 8 plots the growth in the xy time dependent Reynolds and Maxwell's stress as well as the Shakura-Sunyaev alpha parameter defined as

$$\alpha_{SS} \equiv \frac{\int (\rho \delta u_x \delta u_y - \delta b_x \delta b_y) dz}{c_s^2 \int \rho dz}, \quad (90)$$

⁵ Note that the dimensionless Hall parameter in Athena, Q_H , is related to the Hall parameter in our work as $Q_H = \sqrt{2/\beta} \tilde{\eta}_H$.

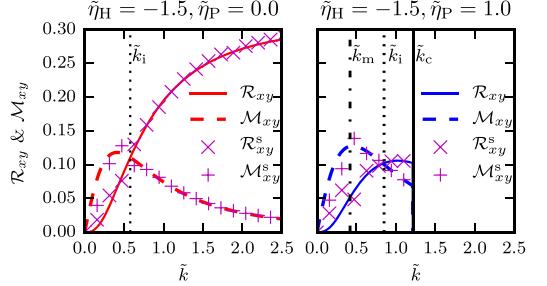


Figure 9. Reynolds and Maxwell's stress components \mathcal{R}_{xy} and M_{xy} of the MRI unstable mode with $\tilde{\eta}_H = -1.5$ and $q = 1.5$ without dissipation $\tilde{\eta}_P = 0$ (left panel) and with dissipation $\tilde{\eta}_P = 1.0$ (right panel). The discrete markers denote the corresponding values of the stresses derived from shearing box simulations. Legends with the superscript “ s ” label the corresponding quantity derived from simulation data.

for the same set of parameters $\tilde{\eta}_H = -1.5, -0.6, 0.0, 1.0$, $\tilde{\eta}_P = 0.0$, and $N_z = 256$ and where the over-lines denote horizontal averages. In accordance with the implications that followed from Equations (87) and (88), we find that even for such moderate resolutions, the Reynolds stress noticeably dominates the Maxwell's stress during the linear growth of the instability. For a fixed value of $\tilde{\eta}_H = -1.5$, we compare the kinetic and magnetic stress and energy densities with two different values of $\tilde{\eta}_P = 0.0, 1.0$ in Figure 9. Although a finite value of k_i appears to

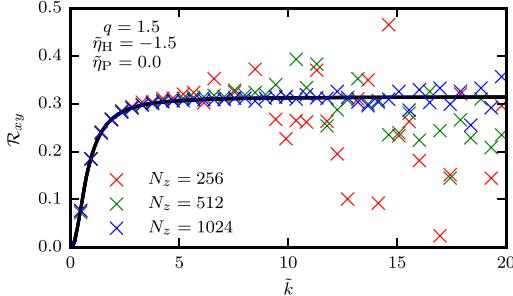


Figure 10. Per- k Reynolds stress component \mathcal{R}_{xy} of the Hall-MRI unstable mode with $\tilde{\eta}_H = -1.5$ and $q = 1.5$. The crosses denote the corresponding values of the per- k stress component derived from shearing box simulations with three different vertical grid resolutions, $N_z = 256, 512, 1024$. The agreement between analytical and numerical results improves at higher wavenumbers as the resolution increases.

be present with $\tilde{\eta}_P = 1.0$, $-\mathcal{M}_{xy} \gg \mathcal{R}_{xy}$ at \tilde{k}_m and so Maxwell's stress maintains its hegemony over its kinetic counterpart.

Figure 10 compares the values of the xy component of the per- k Reynolds stress tensor obtained from simulations with three different vertical grid resolutions. It is quite apparent that with increasing resolution, the agreement between theory and simulation improves substantially as many more smaller scale modes are reliably resolved. This places a stringent requirement upon the resolution demands, while performing simulations of a weakly magnetized shearing system when Hall diffusion is present and dissipation is comparatively weak, if one is to obtain accurate results in accordance with theoretical expectations. In the simulations conducted by Sano & Stone (2002b), the vertical resolution was generally low ($N_z = 32$, $L_z = H$). However, one can already see in their results that the volume averaged Reynolds and Maxwell's stresses at saturation were the same order of magnitude when $\tilde{\eta}_H < 0$ and $\tilde{\eta}_P < 1$. This is not so for comparable simulations performed with resistivity but without Hall diffusion (Sano et al. 2004), where the xy Maxwell's stress at saturation was larger than the corresponding Reynolds stress. While we have not explored the nonlinear regime in our work, we anticipate that with higher grid resolution, one might find stronger mean Reynolds stress perpetuating even at late times. This could be confirmed with dedicated numerical studies.

7. Summary and Discussion

In this paper, we have carried out a detailed examination of the linear eigenmodes in the shearing sheet framework for a weakly magnetized system subject to non-ideal effects with a special focus on Hall diffusion. Although our analysis invoked simplifying assumptions, we have nonetheless been able to go a step further from similar analysis performed in the past and glean certain key attributes governing these modes. A careful examination of the eigenvectors has enabled us to provide a detailed description of the polarization properties and to sketch a visual representation of the eigenmodes as they evolve in space and time. By employing the formalism of Pessah et al. (2006), we have also derived expressions for the kinetic and magnetic stresses and energy densities in terms of the complex eigenvector components. This has enabled us to generalize the ratio of the magnetic to kinetic stresses applicable to the later stages of linear

evolution of the MRI when subject to Hall diffusion. Our central result is the identification of regimes in the parameter space defined by $(\tilde{\eta}_H, \tilde{\eta}_P)$, wherein the kinetic stresses and energies are found to dominate their magnetic equivalents. This property is in sharp contrast with what one expects of the ideal MRI or the MRI subject to dissipative effects alone.

Since the non-ideal MRI unstable eigenmodes studied here are also exact nonlinear solutions of the shearing sheet equations (Goodman & Xu 1994; Kunz & Lesur 2013), the unique traits associated with these modes may carry through or influence the subsequent nonlinear evolution of the system. In ideal as well as dissipative MHD (Latter et al. 2009; Pessah & Goodman 2009; Pessah 2010), these so-called channel modes have been shown to be unstable to parasitic instabilities, which may result in their ultimate saturation. Kunz & Lesur (2013) is the only work we are aware of that has explored the stability of the Hall-MRI modes to parasitic instabilities. In light of the findings presented here, it would be worthwhile to revisit the question of saturation via parasitic modes, particularly for the case with negative Hall diffusivities ($\tilde{\eta}_H < 0$) and weak dissipation.

There have been a number of recent numerical studies of a weakly magnetized system subject to Hall diffusion (Kunz & Lesur 2013; Bai 2014, 2015; Lesur et al. 2014; Simon et al. 2015) in the shearing box framework. To our knowledge, none of these studies have reported anything resembling the behavior of stresses with $\tilde{\eta}_H < 0$, that we have presented in this paper. We surmise that this may be due to the insufficient vertical grid resolution and comparably strong ohmic and ambipolar diffusion present in virtually all of these simulations. Most of these studies have been performed with primary applications to protoplanetary disks and, among them, simulations exploring the system with anti-parallel angular momentum and magnetic field vectors have been comparatively few. However, Simon et al. (2015) did report the appearance of transient turbulent bursts in their shearing box simulations with all non-ideal effects and anti-parallel angular momentum and magnetic field vectors. However, they attribute this behavior to a non-axisymmetric version of the Hall-Shear instability (Rüdiger & Kitchatinov 2005; Kunz 2008).

Conventional wisdom dictates that the ensuing turbulence in a magnetorotationally unstable system is one that is dominated by magnetic stresses and energies. Astrophysical disks such as those around young stellar objects are thought to harbor regions within them where Hall diffusion is the dominant non-ideal effect (Balbus & Terquem 2001; Kunz & Balbus 2004; Wardle 2007; Bai 2011; Wardle & Salmeron 2012; Xu & Bai 2016). These regions may also be subject to diffusion by ohmic and ambipolar diffusion to varying extents. If the dissipative effects are sufficiently strong, they can act to cut down the range of scales unstable to the MRI and thereby curtail the dominance of kinetic stresses if $\tilde{\eta}_H < 0$. However, there is no definitive estimate at the moment of how prevalent the different non-ideal effects are and to what degree. Therefore, it is still too early to judge whether factors that favor the conditions leading to predominant kinetic stresses may or may not be found. The implications that this role-reversal might have upon the ensuing turbulence warrants further study.

We are grateful to the referee whose comments led to an improved version of the paper. We acknowledge useful discussions with Tobias Heinemann, Oliver Gressel, and

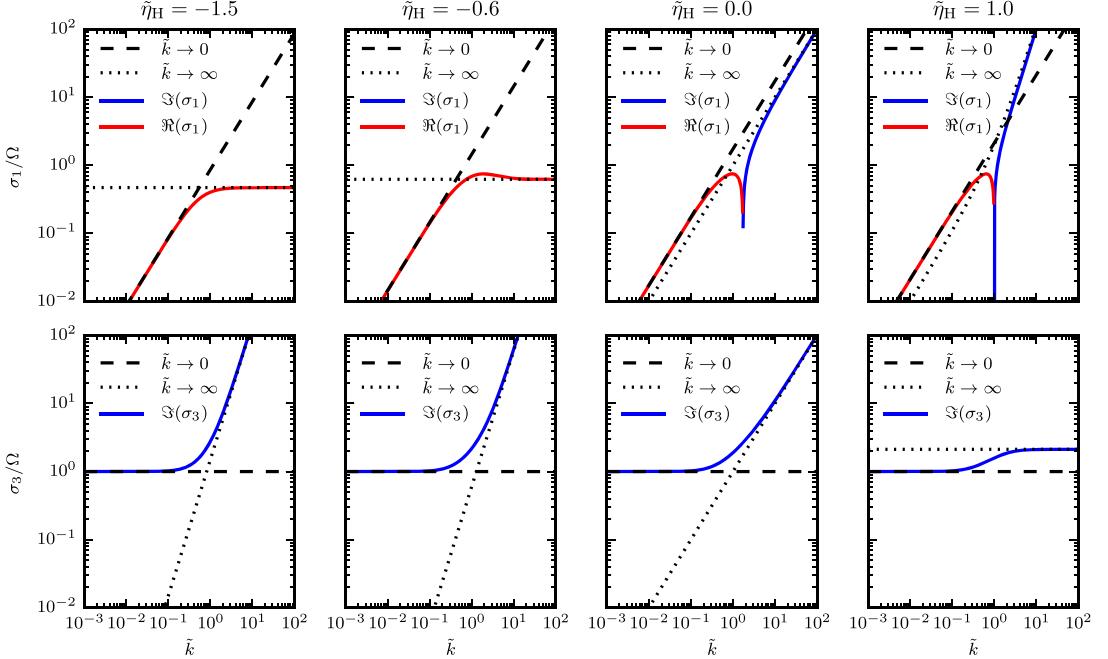


Figure 11. Positive branch solutions, σ_1 and σ_3 , of the eigenvalue problem for four representative values of the Hall parameter, $\tilde{\eta}_H = -1.5, -0.6, 0.0, 1.0$ and $q = 1.5$ in the dissipationless limit. Solid lines represent the numerically computed eigenvalues (the real part in red and the imaginary part in blue). Asymptotic approximations in the low wavenumber limit ($\lim \tilde{k} \rightarrow 0$) and the high wavenumber limit ($\lim \tilde{k} \rightarrow \infty$) are represented by the dashed and dotted lines, respectively.

Leonardo Krapp. We are grateful to Thomas Berlok for help with the simulations and for useful comments on the manuscript. The research leading to these results has received funding from the European Research Council under the European Unions Seventh Framework Programme (FP/2007-2013) under ERC grant agreement 306614.

Appendix A Classification of the Eigenmodes in the Dissipationless Limit

Here, we solve the dispersion relation Equation (26) in the dissipationless limit $\omega_\nu = \omega_p = 0$ and describe the nature of the different solutions in some detail. In the limit $M \rightarrow 0$ and choosing the root such that $y = -L/2$ in Equation (35), we find that the roots of Equation (29) given by Equation (37) reduces to

$$\lim_{M \rightarrow 0} \sigma = \pm \sqrt{-\Lambda_0 \mp \sqrt{\Delta_0}}, \quad (91)$$

where

$$\Lambda_0 = \frac{L_0}{2} \quad \text{and} \quad \Delta_0 = \frac{L_0^2}{4} - N_0, \quad (92)$$

$$\lim_{\tilde{k} \rightarrow 0} \sigma \sim \begin{cases} \omega_A \sqrt{q(2-q)^{-1} + q\tilde{\eta}_H} & \text{R.E.P if } \tilde{\eta}_H \geq 0 \text{ and L.E.P if } \tilde{\eta}_H < 0, \\ i\kappa, & \text{L.E.P} \end{cases} \quad (97)$$

and

$$\Lambda_0 = \omega_A^2 + \frac{\kappa^2}{2} + \frac{\kappa_H^2}{2}, \quad (93)$$

$$\Delta_0 = \left(\Omega - \frac{\omega_H}{2} \right)^2 \left[\left(\omega_H + \frac{\kappa^2}{2\Omega} \right)^2 + 4\omega_A^2 \right]. \quad (94)$$

Setting $\omega_H \rightarrow 0$ in Equations (93) and (94), we recover the ideal MRI solutions (Pessah et al. 2006). For the purpose of identification, we shall designate the four eigenvalues as

$$\sigma_1 = \gamma, \quad \sigma_2 = -\gamma, \quad \sigma_3 = i\omega, \quad \sigma_4 = -i\omega, \quad (95)$$

where

$$\gamma = \sqrt{-\Lambda_0 + \sqrt{\Delta_0}} \quad \text{and} \quad \omega = \sqrt{\Lambda_0 + \sqrt{\Delta_0}}. \quad (96)$$

The notation γ and ω has been chosen to be redolent of the unstable and oscillatory nature of the corresponding eigenmodes. The positive branch eigensolutions, σ_1 and σ_3 , have the following asymptotic forms, at very low and high wavenumbers

$$\lim_{k \rightarrow \infty} \sigma \sim \begin{cases} i\omega_H, & \text{R.E.P if } \tilde{\eta}_H \geq 0 \text{ and L.E.P if } \tilde{\eta}_H < 0, \\ i\omega_G, & \text{L.E.P} \end{cases} \quad (98)$$

where ω_G is the so-called gyration frequency (Heinemann & Quataert 2014)

$$\omega_G = \sqrt{\left[2\Omega + \frac{\omega_A^2}{\omega_H}\right]\left[(2-q)\Omega + \frac{\omega_A^2}{\omega_H}\right]}. \quad (99)$$

In the absence of rotation and shear, ω_G is approximately equal to the ion-cyclotron frequency, $\omega_{ci} = eB/m_i c$ reduced by the ionization fraction n_e/n . The acronyms R.E.P and L.E.P stand for Right and Left Elliptically Polarized, respectively, and indicates the direction of polarization of the oscillatory eigenmodes as seen by an observer looking down perched above the disk midplane.

The Coriolis force and the Hall effect endow the oscillatory modes with a circular polarization or helicity. The effect of shear is to make the polarization elliptical. Hall diffusion has the added effect of bringing about divergent behavior of the oscillatory modes at large wavenumbers. One of the otherwise Alfvénic branches breaks out into what is commonly referred to as the Whistler mode where the frequency varies quadratically with wavenumber. The other Alfvén branch asymptotes to a maximum frequency corresponding to the reduced ion-cyclotron frequency for smaller wavelengths.

Under ideal MHD conditions, an infinitesimal perturbation executes a circular trajectory due to the Coriolis force. The shear eccentrically stretches this motion toward positive azimuth inward from the point of origin and toward negative azimuth outward. The Lorentz tension is activated and tries to restore the fluid element, thereby transferring angular momentum from the inward moving fluid element to the tethered element moving outward. The respective fluid elements fall further inward and outward to compensate and the egression is greater at intermediate length scales where tension is weakest. This is the standard physical picture of the MRI (Balbus & Hawley 1998). When $\tilde{\eta}_H > 0$, the Hall effect introduces an “epicyclic motion” of its own (Balbus & Terquem 2001) that has the opposite sense of the Coriolis induced epicycles. At smaller length scales, this push-back is intensified and together with tension, suppresses any unstable motion. When $\tilde{\eta}_H < 0$, the Hall effect induced epicycles have the same sense as the Coriolis motion and moreover acts to negate the restoring magnetic tension forces at the smaller length scales. These epicycles respond at the frequency ω_G , which is also now purely imaginary and leads to continued exponential growth at ever smaller length scales. Wardle & Salmeron (2012) refer to

the instability as operating in the “cyclotron limit” at the high wavenumber end.

Figure 11 shows the positive eigensolutions, σ_1 and σ_3 as a function of wavenumber for four representative values of $\tilde{\eta}_H$. The asymptotic forms given by Equations (97) and (98) are plotted over the exact solutions for comparison. Notice the eigensolutions σ_1 and σ_3 , splitting into separate branches with $\tilde{\eta}_H = 1$ in Figure 11, at high wavenumbers. For the sake of identification, we shall refer to modes that asymptote to the frequency ω_G , as simply the cyclotron mode. Bear in mind, however, that when $-1/2 < \tilde{\eta}_H < \infty$, σ_1 becomes oscillatory beyond the cut-off wavenumber \tilde{k}_c . The change in sign of $\tilde{\eta}_H$ effects an interchange of the Whistler and cyclotron behavior on the modes, σ_1 and σ_3 , at high wavenumbers. Furthermore, when $-2 < \tilde{\eta}_H < -1/2$, ω_G is purely imaginary and corresponds to the large wavenumber growth rate of the unstable mode, σ_1 .

References

- Armitage, P. J. 2011, *ARA&A*, **49**, 195
- Bai, X.-N. 2011, *ApJ*, **739**, 51
- Bai, X.-N. 2014, *ApJ*, **791**, 137
- Bai, X.-N. 2015, *ApJ*, **798**, 84
- Balbus, S. A., & Hawley, J. F. 1998, *RvMP*, **70**, 1
- Balbus, S. A., & Terquem, C. 2001, *ApJ*, **552**, 235
- Bejarno, C., Gómez, D. O., & Brandenburg, A. 2011, *ApJ*, **737**, 62
- Béthune, W., Lesur, G., & Ferreira, J. 2016, *A&A*, **589**, A87
- Goldreich, P., & Lynden-Bell, D. 1965, *MNRAS*, **130**, 125
- Goodman, J., & Xu, G. 1994, *ApJ*, **432**, 213
- Hardy, G. 1956, *Divergent Series* (Oxford: Clarendon)
- Heinemann, T., & Quataert, E. 2014, *ApJ*, **792**, 70
- Kunz, M. W. 2008, *MNRAS*, **385**, 1494
- Kunz, M. W., & Balbus, S. A. 2004, *MNRAS*, **348**, 355
- Kunz, M. W., & Balbus, S. A. 2013, *MNRAS*, **434**, 2295
- Latter, H. N., Lesaffre, P., & Balbus, S. A. 2009, *MNRAS*, **394**, 715
- Lesur, G., Kunz, M. W., & Fromang, S. 2014, *A&A*, **566**, A56
- O’Sullivan, S., & Downes, T. P. 2006, *MNRAS*, **366**, 1329
- O’Sullivan, S., & Downes, T. P. 2007, *MNRAS*, **376**, 1648
- Pandey, B. P., & Wardle, M. 2008, *MNRAS*, **385**, 2269
- Pessah, M. E. 2010, *ApJ*, **716**, 1012
- Pessah, M. E., & Chan, C.-K. 2008, *ApJ*, **684**, 498
- Pessah, M. E., Chan, C.-K., & Psaltis, D. 2006, *MNRAS*, **372**, 183
- Pessah, M. E., & Goodman, J. 2009, *ApJL*, **698**, L72
- Rüdiger, G., & Kitchatinov, L. L. 2005, *A&A*, **434**, 629
- Sano, T., Inutsuka, S.-i., Turner, N. J., & Stone, J. M. 2004, *ApJ*, **605**, 321
- Sano, T., & Stone, J. M. 2002a, *ApJ*, **570**, 314
- Sano, T., & Stone, J. M. 2002b, *ApJ*, **577**, 534
- Simon, J. B., Lesur, G., Kunz, M. W., & Armitage, P. J. 2015, arXiv:[1508.00904](#)
- Stone, J. M., Gardiner, T. A., Teuben, P., Hawley, J. F., & Simon, J. B. 2008, *ApJS*, **178**, 137
- Turner, N. J., Fromang, S., Gammie, C., et al. 2014, in *Protostars and Planets VI*, ed. H. Beuther et al. (Tucson, AZ: Univ. of Arizona Press), **411**
- Wardle, M. 1999, *MNRAS*, **307**, 849
- Wardle, M. 2007, *Ap&SS*, **311**, 35
- Wardle, M., & Salmeron, R. 2012, *MNRAS*, **422**, 2737
- Xu, R., & Bai, X.-N. 2016, *ApJ*, **819**, 68

Paper III

Analytical Models of Exoplanetary Atmospheres. V. Non-gray Thermal Structure with Coherent Scattering

Reprinted from

The Astrophysical Journal, Volume 858, Issue 1, May 2018

doi:10.3847/1538-4357/aab35d

with the permission of IOP Publishing.



Analytical Models of Exoplanetary Atmospheres. V. Non-gray Thermal Structure with Coherent Scattering

Gopakumar Mohandas^{1,2} , Martin E. Pessah¹ , and Kevin Heng³

¹ Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark; gopakumar@nbi.ku.dk

² Kavli Institute for Theoretical Physics, University of California, Santa Barbara 93106, USA

³ University of Bern, Center for Space and Habitability, Gesellschaftsstrasse 6, CH-3012, Bern, Switzerland

Received 2018 February 2; accepted 2018 February 26; published 2018 April 26

Abstract

We apply the picket fence treatment to model the effects brought about by spectral lines on the thermal structure of irradiated atmospheres. The lines may be due to pure absorption processes, pure coherent scattering processes, or some combination of absorption and scattering. If the lines arise as a pure absorption process, the surface layers of the atmosphere are cooler, whereas this surface cooling is completely absent if the lines are due to pure coherent isotropic scattering. The lines also lead to a warming of the deeper atmosphere. The warming of the deeper layers is, however, independent of the nature of line formation. Accounting for coherent isotropic scattering in the shortwave and longwave continuum results in anti-greenhouse cooling and greenhouse warming on an atmosphere-wide scale. The effects of coherent isotropic scattering in the line and continuum operate in tandem to determine the resulting thermal structure of the irradiated atmosphere.

Key words: methods: analytical – planets and satellites: atmospheres – radiative transfer

1. Introduction

Analytical radiative transfer models have proved to be of considerable utility in the study of stellar and planetary atmospheres since their development a century ago (Mihalas 1970). Despite the availability of sophisticated high-speed numerical techniques in the present day (Hubeny 2017), simplified analytical treatments continue to remain valuable primarily as a means to derive physical insight and understanding of atmospheric conditions.

The usefulness of analytical models to construct model atmospheres are afforded by the simplistic nature of the underlying assumptions, namely that of gray opacities, plane-parallel steady-state structure, hydrostatic and radiative equilibrium (Mihalas 1970). These assumptions are, however, also the source of their limitations. In particular, the assumption of frequency independent (gray) mean opacities are far from a realistic representation of true atmospheric opacities (Heng 2017). Nevertheless, such simplified 1D analytical models have been useful in providing an exact solution that predicts basic atmospheric trends and one that may serve as a good initial approximation for numerical schemes (Hubeny & Mihalas 2014).

Analytical or semi-analytical models with small departures from grayness have been derived over the years (Hubeny & Mihalas 2014). In the context of irradiated atmospheres, such as that of close-in extrasolar planets, the most elementary extension constitutes what is referred to as the two-step gray or semi-gray or dual-band transfer models (Hansen 2008; Guillot 2010; Heng et al. 2012, 2014). These models are predicated on the approximate division of radiant energy into two distinct and nearly nonoverlapping bands; the shortwave associated with external stellar irradiation and the longwave associated with internal planetary thermal emission. The transfer equations or their moments, with different mean opacities for the shortwave and longwave components, are then solved separately and linked together by the principle of energy conservation.

Recently, Parmentier & Guillot (2014) derived an analytical model by applying the picket fence method of Chandrasekhar (1935) to irradiated atmospheres. The picket fence treatment was originally developed to model line blanketing effects in nonirradiated stellar atmospheres and has since been refined by a number of authors (Münch 1946; Athay & Skumanich 1969; Mihalas & Luebke 1971). Spectral line blanketing leads to two major effects that introduces departures from a gray atmosphere: surface cooling and backwarming (Mihalas & Luebke 1971; Athay 2012). The former refers to the reduction of temperature in the upper layers of the atmosphere due to the added emissivity of the lines whereas the latter effect describes the temperature enhancement in the deeper atmosphere as a result of the flux redistribution within the continuum due to the lines. Both of these effects were present in the model derived by Parmentier & Guillot (2014). However, their analysis did not account for the influence of scattering in both the shortwave and the longwave. The surface cooling effect has been seen to be dependent on the nature of line formation in nonirradiated atmospheres (Chandrasekhar 1935). The degree of cooling is lower if scattering contributes to the line in some measure and is completely absent when the lines are entirely due to a pure scattering process. One would expect this feature to be present in irradiated atmospheres as well. Furthermore, continuum scattering is known to induce an atmosphere-wide shift in temperatures (Heng et al. 2014; Heng 2017). This shift is toward hotter temperatures if the isotropic scattering contributes to the longwave continuum and is toward lower temperatures if isotropic scattering contributes to the shortwave continuum. Therefore, one must account for scattering in the line and continuum in order to derive a closer approximation to actual atmospheric thermal structures.

In this paper, we generalize the picket fence treatment to irradiated atmospheres to include coherent scattering effects. We derive solutions that accommodate for coherent isotropic scattering in the lines as well as the continuum, in the longwave and shortwave frequency bands. Our model therefore provides a fuller picture of the possible atmospheric thermal structure

while preserving the advantages and utility rendered by tractable nongray analytical models.

The paper is organized as follows. In Section 2, we present the picket fence model equations. In Section 3, we connect the longwave picket fence model with the shortwave equations and solve for the resulting temperature profile. We present a discussion of the results in Section 4 and conclude with a summary in Section 5.

2. The Picket Fence Model

We consider a plane-parallel atmosphere and model the transport of radiation by solving the moments of the steady-state radiative transfer equation. We apply the dual-band approximation and derive separate moment equations in the longwave and shortwave frequency bands.

We begin with the longwave band where we make use of the picket fence method (Chandrasekhar 1935). The radiative transfer equation in a plane-parallel atmosphere has the basic form

$$\mu \frac{\partial I_\nu}{\partial m} = k(\nu)(I_\nu - S_\nu), \quad (1)$$

where μ is the cosine of the zenith angle, m is the column mass, k is the frequency dependent extinction opacity, I_ν is the specific intensity, and S_ν is the source function. The intensity and source function are, in general, functions of μ , m , and ν .

We define the moments of the intensity, the mean intensity, the Eddington flux, and the K -integral, respectively, as follows (Mihalas 1970)

$$J_\nu \equiv \frac{1}{2} \int_{-1}^1 I_\nu(\tau, \mu) d\mu, \quad (2)$$

$$H_\nu \equiv \frac{1}{2} \int_{-1}^1 I_\nu(\tau, \mu) \mu d\mu, \quad (3)$$

$$K_\nu \equiv \frac{1}{2} \int_{-1}^1 I_\nu(\tau, \mu) \mu^2 d\mu. \quad (4)$$

In the parts of the frequency interval containing lines, the equation of transfer is (Athey 2012; Mihalas 1970)

$$\mu \frac{\partial I_\nu}{\partial m} = (k_c + k_l) I_\nu - (k_c S_{c\nu} + k_l S_{l\nu}), \quad (5)$$

where the subscripts c and l refer to the continuum and line respectively. Considering coherent isotropic scattering, the continuum and line source functions have the form

$$S_{c\nu} = \varepsilon_c B_\nu + (1 - \varepsilon_c) J_\nu, \quad \varepsilon_c = \frac{\kappa_c}{k_c}, \quad (6)$$

$$S_{l\nu} = \varepsilon_l B_\nu + (1 - \varepsilon_l) J_\nu, \quad \varepsilon_l = \frac{\kappa_l}{k_l}, \quad (7)$$

where κ denotes the absorption opacity that together with the corresponding scattering opacity assume constant but separate values in the line and continuum. The parameter ε is a measure of the fraction of photons lost to pure absorption (Mihalas 1970). It is in fact the complementary parameter to the single scattering albedo but we shall refer to it here as the scattering parameter regardless.

Integrating over the frequency interval containing only the lines leads to the transfer equation

$$\mu \frac{\partial I_l}{\partial \tau} = \left(\frac{1}{\varepsilon_c} + \xi \right) I_l - \left[\frac{1}{\varepsilon_c} - 1 + (1 - \varepsilon_l)\xi \right] J_l - (1 + \varepsilon_l\xi)\beta \mathcal{B}, \quad (8)$$

whereas integrating over the remainder of the frequency interval representing the continuum yields

$$\varepsilon_c \mu \frac{\partial I_2}{\partial \tau} = I_2 - (1 - \varepsilon_c) J_2 - (1 - \beta)\varepsilon_c \mathcal{B}. \quad (9)$$

Here the subscripts 1 and 2 represent the integrated variables in the line and continuum respectively. We have also defined the line to continuum opacity ratio

$$\xi = \frac{k_l}{\kappa_c}, \quad (10)$$

the integrated blackbody function

$$\mathcal{B} = \int B_\nu d\nu, \quad (11)$$

and the frequency independent optical depth

$$d\tau = \kappa_c dm. \quad (12)$$

Finally, we introduce the parameter β , which gives the relative probability of finding a line in the frequency interval and may in general be a function of frequency. However, we take β to be a constant in Equations (8) and (9) assuming that the lines have uniform width and are uniformly spread across the spectral range.

The first and second moments of the radiative transfer equation in the line and continuum are, respectively,

$$\frac{dH_l}{d\tau} = \lambda(J_l - \beta \mathcal{B}), \quad (13)$$

$$\frac{dH_2}{d\tau} = J_2 - (1 - \beta)\mathcal{B}, \quad (14)$$

and

$$\frac{dK_l}{d\tau} = \frac{\eta}{\varepsilon_c} H_l, \quad (15)$$

$$\frac{dK_2}{d\tau} = \frac{1}{\varepsilon_c} H_2, \quad (16)$$

where we have defined

$$\lambda = 1 + \varepsilon_l\xi, \quad \eta = 1 + \varepsilon_c\xi. \quad (17)$$

In the limit $\varepsilon_c = 1$, we recover the moment equations used by Chandrasekhar (1935).

3. Irradiated Atmospheres

We now extend the original picket fence treatment to irradiated atmospheres, like that of close-in extrasolar giant planets. This is achieved by linking the shortwave transfer solution to the longwave picket fence model solution via the radiative equilibrium condition.

We first consider the radiative transfer equation as it applies to the shortwave frequencies. The frequency integrated transfer equation for the shortwave band, where the source function contains only a non-negligible contribution due to coherent

isotropic scattering, is given by

$$\mu \frac{dI_s}{d\tau} = \frac{\gamma}{\varepsilon_s} [I_s - (1 - \varepsilon_s)J_s], \quad (18)$$

with

$$\gamma = \frac{\kappa_s}{\kappa_c}, \quad \varepsilon_s = \frac{\kappa_s}{k_s}, \quad (19)$$

where κ_s and k_s are the shortwave absorption and extinction opacities respectively. The parameter γ quantifies the strength of the shortwave opacity to its longwave continuum counterpart and ε_s measures the fraction of shortwave photons lost to absorption. The moment equations for the shortwave are

$$\frac{dH_s}{d\tau} = \gamma J_s, \quad (20)$$

and

$$\frac{dK_s}{d\tau} = \frac{\gamma}{\varepsilon_s} H_s, \quad (21)$$

Using the closure relation $K_s = \bar{\mu}^2 J_s$ (Guillot 2010; Heng et al. 2012, 2014), where $\bar{\mu}$ is the cosine of the angle of the collimated stellar beam with respect to the vertical, we obtain the second-order ordinary differential equation

$$\frac{d^2J_s}{d\tau^2} = \frac{\gamma_\mu^2}{\varepsilon_s} J_s, \quad (22)$$

which has the simple exponential solution

$$J_s = J_s(0) \exp(-\gamma_{\mu\varepsilon}\tau), \quad (23)$$

where $\gamma_\mu = \gamma/|\bar{\mu}|$ and $\gamma_{\mu\varepsilon} = \gamma_\mu/\sqrt{\varepsilon_s}$ are assumed constant. Consistency with Equation (20) implies

$$H_s = H_s(0) \exp(-\gamma_{\mu\varepsilon}\tau) \quad (24)$$

with $H_s(0) = -|\bar{\mu}|\sqrt{\varepsilon_s}J_s(0)$ Heng et al. (2014).

The radiative equilibrium condition, which is given by $d(H_1 + H_2 + H_s)/d\tau = 0$ implies

$$\lambda J_1 + J_2 + \gamma J_s = [\lambda\beta + 1 - \beta]\mathcal{B}, \quad (25)$$

Adding Equations (13) and (14), we have

$$\frac{d}{d\tau}(H_1 + H_2) = -\gamma J_s = \gamma_{\mu\varepsilon} H_s, \quad (26)$$

which has the full solution

$$H_1 + H_2 = \mathcal{H} - H_s. \quad (27)$$

where \mathcal{H} is the total integrated longwave Eddington flux. Combining Equations (15) and (16), we obtain

$$\frac{d}{d\tau}\left(\frac{K_1}{\eta} + K_2\right) = \frac{\mathcal{H}}{\varepsilon_c} - \frac{1}{\varepsilon_c} H_s, \quad (28)$$

which has the full solution

$$\frac{K_1}{\eta} + K_2 = \frac{\mathcal{H}}{\varepsilon_c}\tau + c + \frac{1}{\varepsilon_c\gamma_{\mu\varepsilon}} H_s. \quad (29)$$

With the Eddington approximation $J_{1,2} = 3K_{1,2}$, we may express the integrated Planck function as

$$\mathcal{B} = \frac{3}{\bar{\lambda}} \left[\lambda K_1 + K_2 - \frac{\gamma_{\mu\varepsilon}}{3} H_s \right]. \quad (30)$$

Using Equation (29), this may be written in either of the two forms given below

$$\mathcal{B} = \frac{3}{\bar{\lambda}} \left[\frac{\mathcal{H}}{\varepsilon_c} \tau + c + \left(\lambda - \frac{1}{\eta} \right) K_1 + \left(\frac{1}{\varepsilon_c\gamma_{\mu\varepsilon}} - \frac{\gamma_{\mu\varepsilon}}{3} \right) H_s \right], \quad (31)$$

$$\mathcal{B} = \frac{3}{\bar{\lambda}} \left[\lambda\eta \left(\frac{\mathcal{H}}{\varepsilon_c} \tau + c \right) + (1 - \lambda\eta) K_2 + \left(\frac{\lambda\eta}{\varepsilon_c\gamma_{\mu\varepsilon}} - \frac{\gamma_{\mu\varepsilon}}{3} \right) H_s \right]. \quad (32)$$

Combining Equation (15) with Equation (13) and Equation (16) with Equation (14) by using the Eddington approximation, we obtain the pair of inhomogeneous second-order ordinary differential equations

$$\frac{\varepsilon_c d^2 K_1}{d\tau^2} = \eta\lambda(3K_1 - \beta\mathcal{B}), \quad (33)$$

$$\frac{\varepsilon_c d^2 K_2}{d\tau^2} = 3K_2 - (1 - \beta)\mathcal{B}. \quad (34)$$

Substituting Equations (31) and (32) in Equations (33) and (34) yields

$$\begin{aligned} \frac{d^2 K_1}{d\tau^2} &= \frac{3\lambda\tilde{\eta}}{\varepsilon_c\tilde{\lambda}} K_1 - \frac{3\eta\lambda\beta}{\varepsilon_c^2\tilde{\lambda}} (\mathcal{H}\tau + c) \\ &\quad - \frac{3\eta\lambda\beta}{\varepsilon_c\tilde{\lambda}} \left(\frac{1}{\varepsilon_c\gamma_{\mu\varepsilon}} - \frac{\gamma_{\mu\varepsilon}}{3} \right) H_s, \end{aligned} \quad (35)$$

$$\begin{aligned} \frac{d^2 K_2}{d\tau^2} &= \frac{3\lambda\tilde{\eta}}{\varepsilon_c\tilde{\lambda}} K_2 - \frac{3\eta\lambda(1 - \beta)}{\varepsilon_c^2\tilde{\lambda}} (\mathcal{H}\tau + c) \\ &\quad - \frac{3(1 - \beta)}{\varepsilon_c\tilde{\lambda}} \left(\frac{\lambda\eta}{\varepsilon_c\gamma_{\mu\varepsilon}} - \frac{\gamma_{\mu\varepsilon}}{3} \right) H_s, \end{aligned} \quad (36)$$

where we have defined the convenient shorthands

$$\tilde{\lambda} = \lambda\beta + 1 - \beta, \quad (37)$$

$$\tilde{\eta} = \beta + \eta(1 - \beta). \quad (38)$$

Bounded solutions to Equations (35) and (36) are given by

$$\begin{aligned} K_1 &= a \exp(-q\tau) + \frac{\eta\beta}{\varepsilon_c\tilde{\eta}} (\mathcal{H}\tau + c) \\ &\quad - \frac{3\lambda\eta\beta}{\gamma_{\mu\varepsilon}^2\varepsilon_c\tilde{\lambda} - 3\lambda\tilde{\eta}} \left(\frac{1}{\varepsilon_c\gamma_{\mu\varepsilon}} - \frac{\gamma_{\mu\varepsilon}}{3} \right) H_s, \end{aligned} \quad (39)$$

$$\begin{aligned} K_2 &= b \exp(-q\tau) + \frac{\eta(1 - \beta)}{\varepsilon_c\tilde{\eta}} (\mathcal{H}\tau + c) \\ &\quad - \frac{3(1 - \beta)}{\gamma_{\mu\varepsilon}^2\varepsilon_c\tilde{\lambda} - 3\lambda\tilde{\eta}} \left(\frac{\lambda\eta}{\varepsilon_c\gamma_{\mu\varepsilon}} - \frac{\gamma_{\mu\varepsilon}}{3} \right) H_s, \end{aligned} \quad (40)$$

where q , its inverse rather, is a characteristic optical depth that is given by

$$q = \sqrt{\frac{3\lambda\tilde{\eta}}{\varepsilon_c\tilde{\lambda}}}. \quad (41)$$

In order for Equations (39) and (40) to add up to Equation (29), we require $a = -b\eta$. Finally, the integrated Planck function

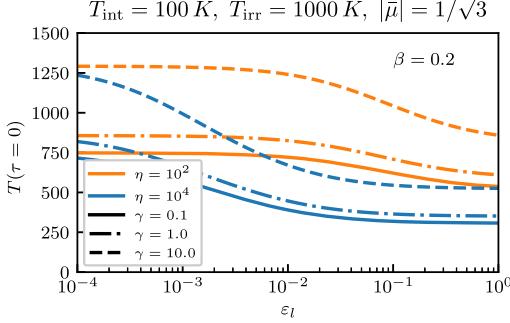


Figure 1. Boundary temperature $T(\tau = 0)$ of an irradiated atmosphere as a function of the line scattering parameter ε_l at a fixed line width β but different values of the line strength η and shortwave absorption opacity parameter γ .

may be expressed as

$$\begin{aligned} \mathcal{B} = & \frac{3\eta}{\varepsilon_c\tilde{\eta}}(\mathcal{H}\tau + c) - \frac{3(\lambda\eta - 1)}{\tilde{\lambda}}b \exp(-q\tau) \\ & - \frac{(\gamma_{\mu\varepsilon}^2\varepsilon_c - 3)(\gamma_{\mu\varepsilon}^2\varepsilon_c - 3\eta\lambda)}{\gamma_{\mu\varepsilon}\varepsilon_c(\gamma_{\mu\varepsilon}^2\varepsilon_c\tilde{\lambda} - 3\lambda\tilde{\eta})}H_s(0)\exp(-\gamma_{\mu\varepsilon}\tau), \end{aligned} \quad (42)$$

Thus it remains only to determine the integration constants b and c , which we achieve by the application of suitable boundary conditions. For the sake of conformity with Chandrasekhar (1935), we use the relations $\mathcal{F} = 4\mathcal{H}$ and $\mathcal{F}_* = 4H_s(0)$. Assuming that at $\tau = 0$, we have $4H_{1,2}(0) = 6K_{1,2}(0)$ (Chandrasekhar 1935) and by considering Equations (15) and (16) evaluated at $\tau = 0$, we obtain

$$\begin{aligned} (6\eta + 4q)b - \frac{6\eta\beta}{\varepsilon_c\tilde{\eta}}c + \frac{\beta}{\varepsilon_c\tilde{\eta}}\mathcal{F} \\ + \frac{3\lambda\beta(\gamma_{\mu\varepsilon} + \frac{3}{2}\eta)}{\gamma_{\mu\varepsilon}^2\varepsilon_c\tilde{\lambda} - 3\lambda\tilde{\eta}}\left(\frac{1}{\varepsilon_c\gamma_{\mu\varepsilon}} - \frac{\gamma_{\mu\varepsilon}}{3}\right)\mathcal{F}_* = 0, \end{aligned} \quad (43)$$

$$\begin{aligned} (6 + 4q)b + \frac{6\eta(1 - \beta)}{\varepsilon_c\tilde{\eta}}c - \frac{\eta(1 - \beta)}{\varepsilon_c\tilde{\eta}}\mathcal{F} \\ - \frac{3(1 - \beta)(\gamma_{\mu\varepsilon} + \frac{3}{2})}{\gamma_{\mu\varepsilon}^2\varepsilon_c\tilde{\lambda} - 3\lambda\tilde{\eta}}\left(\frac{\lambda\eta}{\varepsilon_c\gamma_{\mu\varepsilon}} - \frac{\gamma_{\mu\varepsilon}}{3}\right)\mathcal{F}_* = 0. \end{aligned} \quad (44)$$

The boundary condition used here corresponds to the application of the second and third choices listed in Parmentier & Guillot (2014, see their Section 2.3.2) with the second Eddington coefficient $f_H = 1/2$. Solving the system of Equations (43) and (44) results in

$$b = b_l\mathcal{F} + b_s\mathcal{F}_*, \quad c = c_l\mathcal{F} + c_s\mathcal{F}_*, \quad (45)$$

where

$$b_l = \frac{\beta(1 - \beta)(\eta - 1)}{\varepsilon_c\tilde{\eta}(6\tilde{\eta} + 4q)}, \quad (46)$$

$$\begin{aligned} b_s = & \frac{\beta(1 - \beta)[\gamma_{\mu\varepsilon}^2\varepsilon_c(\lambda - 1) + 3\lambda(\eta - 1)]}{\varepsilon_c(\gamma_{\mu\varepsilon}^2\varepsilon_c\tilde{\lambda} - 3\lambda\tilde{\eta})(6\tilde{\eta} + 4q)} \\ & + \frac{3}{2}\frac{\beta(1 - \beta)\gamma_{\mu\varepsilon}(\lambda\eta - 1)}{\gamma_{\mu\varepsilon}^2\varepsilon_c\tilde{\lambda} - 3\lambda\tilde{\eta})(6\tilde{\eta} + 4q)}, \end{aligned} \quad (47)$$

$$c_l = \frac{6[\beta + \eta^2(1 - \beta)] + 4q\tilde{\eta}}{6\eta(6\tilde{\eta} + 4q)}, \quad (48)$$

$$\begin{aligned} c_s = & \frac{\tilde{\eta}\left(\gamma_{\mu\varepsilon} + \frac{3}{2}\right)(6\eta + 4q)[\gamma_{\mu\varepsilon}^2\varepsilon_c(\beta - 1) + 3\lambda\eta]}{6\gamma_{\mu\varepsilon}\varepsilon_c\eta(\gamma_{\mu\varepsilon}^2\varepsilon_c\tilde{\lambda} - 3\lambda\tilde{\eta})(6\tilde{\eta} + 4q)} \\ & - \frac{3\tilde{\eta}\lambda\beta(\eta - 1)[9\eta + \gamma_{\mu\varepsilon}(6 + 4q + 6\eta)]}{6\gamma_{\mu\varepsilon}\varepsilon_c\eta(\gamma_{\mu\varepsilon}^2\varepsilon_c\tilde{\lambda} - 3\lambda\tilde{\eta})(6\tilde{\eta} + 4q)} \\ & - \frac{\tilde{\eta}\lambda\beta\gamma_{\mu\varepsilon}(6 + 4q)\left(\gamma_{\mu\varepsilon} + \frac{3}{2}\eta\right)}{\varepsilon_c(6\gamma_{\mu\varepsilon}^2\varepsilon_c\tilde{\lambda} - 3\lambda\tilde{\eta})(6\tilde{\eta} + 4q)}. \end{aligned} \quad (49)$$

We now express the fluxes in terms of their respective equilibrium temperatures as given by

$$\mathcal{B} = \frac{\sigma_{\text{SB}}T^4}{\pi}, \quad \mathcal{F} = \frac{\sigma_{\text{SB}}T_{\text{int}}^4}{\pi}, \quad \mathcal{F}_* = \tilde{\mu}\frac{\sigma_{\text{SB}}T_{\text{irr}}^4}{\pi}, \quad (50)$$

where σ_{SB} is the Stefan–Boltzmann constant, T_{int} is the internal temperature associated with the thermal flux at the bottom of the atmospheres and T_{irr} is the irradiation temperature associated with the flux at the top of the atmosphere. This substitution results in the temperature profile given by

$$\begin{aligned} T^4 = & \frac{3T_{\text{int}}^4}{4}\left[\frac{\eta}{\varepsilon_c\tilde{\eta}}\tau + \frac{4\eta}{\varepsilon_c\tilde{\eta}}c_l - \frac{4(\eta\lambda - 1)}{\tilde{\lambda}}b_l\exp(-q\tau)\right] \\ & - \frac{3|\tilde{\mu}|T_{\text{irr}}^4}{4}\left[\frac{4\eta}{\varepsilon_c\tilde{\eta}}c_s - \frac{4(\eta\lambda - 1)}{\tilde{\lambda}}b_s\exp(-q\tau)\right. \\ & \left. - \frac{(\gamma_{\mu\varepsilon}^2\varepsilon_c - 3)(\gamma_{\mu\varepsilon}^2\varepsilon_c - 3\eta\lambda)}{3\gamma_{\mu\varepsilon}\varepsilon_s(\gamma_{\mu\varepsilon}^2\varepsilon_c\tilde{\lambda} - 3\lambda\tilde{\eta})}\exp(-\gamma_{\mu\varepsilon}\tau)\right]. \end{aligned} \quad (51)$$

In the limit of $\beta \rightarrow 0$ and no external irradiation $T_{\text{irr}} = 0$, we have

$$\tilde{\eta} = \eta, \quad \tilde{\lambda} = 1, \quad b_l = 0, \quad c_l = \frac{1}{6}, \quad (52)$$

and Equation (51) reduces to the classic Milne's solution

$$T^4 = \frac{3}{4}T_{\text{int}}^4\left(\tau + \frac{2}{3}\right). \quad (53)$$

With only $T_{\text{irr}} = 0$, we recover the nonirradiated solution of Chandrasekhar (1935, see Equation (52))

4. Results

The picket fence model of Chandrasekhar (1935) was originally developed to capture the effects that arise due to line blanketing in stellar atmosphere models. The initial treatment was based on the ideal case of lines with uniform width, strength, and separation. A discussion of the specific limitations that result from these assumptions is presented in Athay & Skumanich (1969) and Athay (2012). Nevertheless, the original model did succeed in illustrating the basic effects due to line blanketing, namely that of cooler surface temperatures and warmer deeper temperatures. We have, therefore, retained the same simplified treatment of the lines in order to construct a general nongray model that includes coherent isotropic scattering. All of the results presented here are derived assuming fiducial values for the

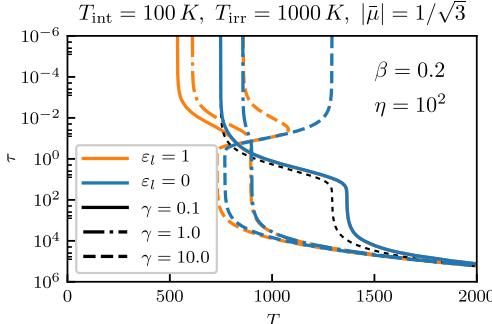


Figure 2. Temperature profiles for a given line width β and two different values of the line strength η . The temperature profile in each case is derived for three different values of the shortwave opacity parameter γ represented by the solid, dashed, and dashed-dotted curves. The blue curves represent the profile with lines that are due to pure scattering, whereas the orange curves represent lines due to pure absorption. $\gamma > 1$ results in an anti-greenhouse effect leading to a relatively hotter upper atmosphere and a cooler lower atmosphere. The dotted black line is the gray temperature profile computed for $\gamma = 0.1$ and is plotted here to illustrate the backwarming effect seen in the solid (blue and orange) curves for the same γ .

internal temperature $T_{\text{int}} = 100$ K, the irradiation temperature $T_{\text{irr}} = 1000$ K and irradiation angle given by $|\bar{\mu}| = 1/\sqrt{3}$.

4.1. Surface Cooling and Backwarming

Line blanketing introduces two main effects in the thermal structure and spectrum of atmospheres.⁴ If line formation can be attributed solely to absorption processes, the temperature of the upper layer is lower compared to what it would be in the absence of lines (the gray limit), an effect that is referred to as surface cooling (Athey 2012). The degree of surface cooling is lowered if the lines are partly due to scattering processes parameterized here by ε_l . The surface cooling is completely absent if the lines are entirely due to scattering $\varepsilon_l \rightarrow 0$. Figure 1 shows the change in boundary or skin temperature, here referred to as the temperature at zero optical depth, with the longwave scattering parameter ε_l for a fixed line width but different line strengths. In the case of irradiated atmospheres, the strength of the shortwave opacity also influences the boundary temperature due to greater or lower relative absorption of incident starlight in the upper layers. Higher values of γ , therefore, lead to higher upper layer temperatures.

The deeper layer warming observed in the nongray model that we have derived here is due to the effect of line blanketing. This backwarming results from the increase in temperature due to an attendant increase in the radiative flux per unit interval that is redistributed within the continuum band of frequencies as a result of its occlusion by the lines. Deeper layer backwarming is, however, insensitive to the nature of line formation. Figure 2 shows temperature profiles for two limiting values of the scattering parameter ε_l and two different values of the line strength η . The profiles are calculated for a fixed angle of the irradiation beam, a constant line width, fiducial values of the effective internal and irradiation temperatures, and three different values of the shortwave absorption opacity (excluding the effect of shortwave scattering here). We plot the gray temperature profile with $\gamma = 0.1$ as a dotted black line in Figure 2 to illustrate the backwarming effect. Notice that the corresponding nongray temperature profile with $\varepsilon_l = 1$ and $\varepsilon_l = 0$ in Figure 2 are both warmer by the same extent with

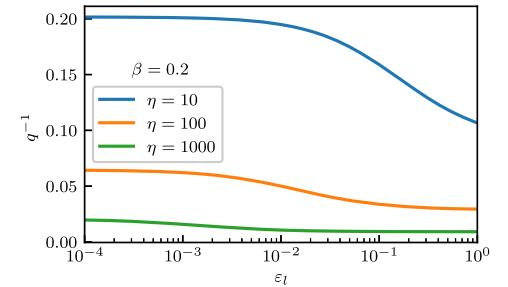
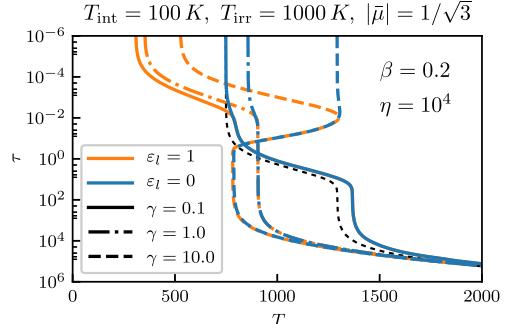


Figure 3. Characteristic optical depth q^{-1} , below which surface cooling effects are observed, as a function of the line scattering parameter ε_l for different values of the line strength η at a fixed line width β .

reference to the dotted line thereby illustrating the the deeper layer warming as well as its insensitivity to the nature of line formation. The degree of backwarming increases with the width of the line, represented here by β , and has been examined in detail by Parmentier & Guillot (2014). In the limit $\varepsilon_l = 1$ and $\gamma \gtrsim 1$, we see, as was also found in Parmentier & Guillot (2014), that the lower boundary temperatures confine the heating due to stronger shortwave absorption into a thin hot layer immediately below the surface. This thin hot layer is of course absent in the opposite limit of $\varepsilon_l = 0$ due to the lack of surface cooling.

4.2. Limit Optical Depth

A characteristic depth that emerges from the picket fence calculation is given in terms of q^{-1} as defined by Equation (41). The corresponding quantity in Parmentier & Guillot (2014) is referred to as τ_{lim} and it represents the depth above which the surface cooling effect may be present in the atmosphere provided that the lines are due in some part to absorption processes. This scale is a function of the line width β , the line opacity ratio ξ , as well as the scattering parameters ε_l and ε_c . However, as shown in Figure 3, the variation of q^{-1} with respect to ε_l is negligible and is largely decided by the line width and strength. Similarly, any meaningful change in ε_c brings about only a negligible modification to q^{-1} . Our results derived for $\varepsilon_l \rightarrow 1$ are

⁴ Some works include the additional effect of line blocking, which we do not consider here. Our usage of the term blanketing is in the sense of Athay (2012).

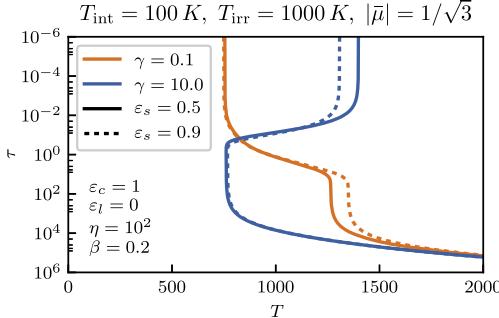


Figure 4. Temperature profiles for two different values of the shortwave continuum coherent isotropic scattering parameter ε_s , represented by the solid and dotted lines. The profiles are derived in the presence of spectral lines of a given width and strength as well as two different values of the shortwave opacity parameter γ represented by the orange and blue curves. The left panel illustrates the temperature profile when the lines are due to pure scattering and the right panel is for lines due to pure absorption. We consider the limit of pure absorption in the longwave continuum here. The inclusion of coherent shortwave continuum scattering results in a leftward shift in the temperature profile toward cooler temperatures.

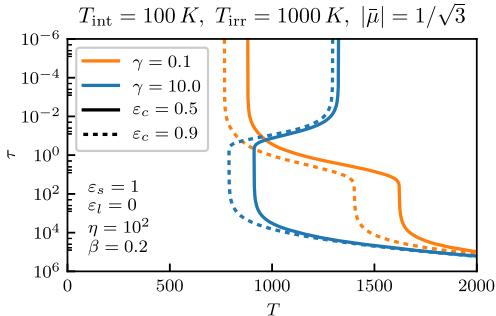


Figure 5. Temperature profiles for two different values of the longwave continuum coherent isotropic scattering parameter ε_c , represented by the solid and dotted lines. The profiles are derived in the presence of spectral lines of a given width and strength as well as two different values of the shortwave opacity parameter γ represented by the orange and blue curves. The left panel illustrates the temperature profile when the lines are due to pure scattering and the right panel displays the profiles when the lines are due to pure absorption. We consider the limit of pure absorption in the shortwave continuum here. The inclusion of coherent longwave continuum scattering results in a rightward shift in the temperature profile toward hotter temperatures.

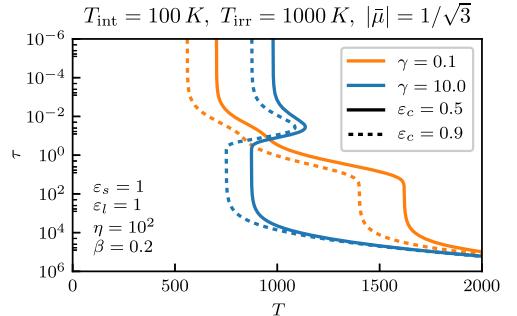
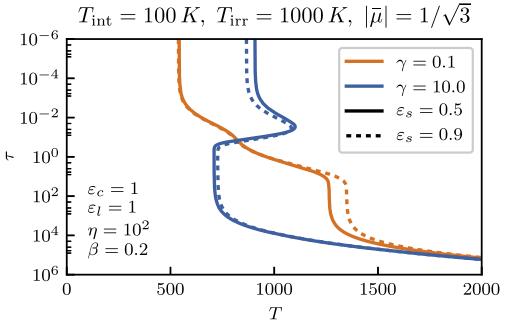
qualitatively similar to Parmentier & Guillot (2014) and differ quantitatively only by a factor of a Rosseland mean opacity ϱ , defined here in dimensionless form as (Chandrasekhar 1935, see Equation (56))

$$\frac{1}{\varrho} \equiv \frac{\beta}{\eta} + 1 - \beta. \quad (54)$$

4.3. Longwave and Shortwave Continuum Scattering

The effect of coherent scattering in the shortwave is to push the temperature profile to lower values on a near global scale (Heng 2017). Coherent isotropic continuum scattering in the shortwave is parameterized in terms of ε_s and its effect on the temperature profile is demonstrated in Figure 4. The global shift in the thermal profile toward lower temperatures adds to any surface cooling present that is due to absorption lines and also effectively offsets the backwarming in the mid to deeper layers. The lower temperatures result from a reduction in the total energy budget by a factor of $1-A_B$, where A_B is the Bond albedo, which may be expressed in terms of the scattering parameter ε_s (Heng et al. 2012).

Coherent isotropic scattering in the longwave continuum band of frequencies has a similar atmosphere-wide effect where



the temperatures are now shifted to higher values as illustrated in Figure 5. This is a manifestation of the classical greenhouse effect (Heng 2017) and is different from the lack of surface cooling due to lines formed by scattering. The former is an actual warming process and is present on a global scale, whereas the latter is the result of the lines being uncoupled from the thermal energy reservoir (Mihalas & Luebke 1971). Taken together, scattering processes therefore play an important role in determining the equilibrium temperature profile even in simple pseudo-nongray models.

5. Summary

We have derived an analytical model for irradiated atmospheres that combines the effect of spectral lines in the longwave band of frequencies, where the lines may be due to either pure absorption or pure coherent scattering processes or some combination of the two. To achieve this, we adapted the picket fence treatment of Chandrasekhar (1935) to model line blanketing effects. The picket fence treatment has been recently used to model irradiated atmospheres but without including the possibility of lines due to coherent scattering (Parmentier & Guillot 2014). Our results demonstrate that the cooling of the upper layers due to line blanketing depends on the nature of

line formation as was previously observed in the context of nonirradiated atmospheres. If scattering is solely responsible for the lines, then the surface temperatures retain their gray value, as the lines are not coupled to the thermal energy of the gas in this limit. Transit spectroscopy of exoplanets is generally most sensitive to very low pressure levels or equivalently the upper layers of the exoplanet's atmosphere (Madhusudhan et al. 2014). Given that the surface temperature is sensitive to the line formation process as revealed by the picket fence analysis, one must exercise caution in the interpretation of observations on the basis of atmospheric transfer models. Furthermore, the contribution of coherent scattering in the continuum can significantly alter global temperature levels depending on the wavelength band. If the planet reflects some fraction of its incident light, the deeper layer temperatures are lowered and negates the backwarming effect due to the lines. If coherent scattering is present in the longwave continuum, the greenhouse effect comes into play leading to greater warming throughout the atmosphere. Our analytical model therefore accommodates a greater range of possibilities over a larger parameter space and may be used to derive reasonable estimates of the thermal structure of irradiated atmospheres.

G.M. thanks the Center for Space and Habitability in Bern, where part of this work was carried out, for their hospitality. G.M. thanks the Kavli Institute for Theoretical Physics, where this work was completed, for their hospitality and support through the KITP graduate fellowship program. The research leading to these results has received funding from the European Research Council under the European Unions Seventh Framework Programme (FP/2007-2013) under ERC grant agreement 306614. K.H. thanks the Swiss National Science Foundation,

the Center for Space and Habitability, the European Research Council, and the Swiss-based MERAC Foundation for partial financial support. This research was supported in part by the National Science Foundation under Grant No. NSF PHY 17-48958.

ORCID iDs

Gopakumar Mohandas  <https://orcid.org/0000-0002-1212-3492>

Martin E. Pessah  <https://orcid.org/0000-0001-8716-3563>
Kevin Heng  <https://orcid.org/0000-0003-1907-5910>

References

- Athay, R. G., & Skumanich, A. 1969, *ApJ*, **155**, 273
- Athay, R. G. 2012, *Radiation Transport in Spectral Lines*, Vol. 1 (Berlin: Springer Science & Business Media)
- Chandrasekhar, S. 1935, *MNRAS*, **96**, 21
- Guillot, T. 2010, *A&A*, **520**, A27
- Hansen, B. M. S. 2008, *ApJS*, **179**, 484
- Heng, K. 2017, *Exoplanetary Atmospheres: Theoretical Concepts and Foundations* (Princeton, NJ: Princeton Univ. Press)
- Heng, K., Hayek, W., Pont, F., & Sing, D. K. 2012, *MNRAS*, **420**, 20
- Heng, K., Mendonça, J. M., & Lee, J.-M. 2014, *ApJS*, **215**, 4
- Hubeny, I. 2017, *MNRAS*, **469**, 841
- Hubeny, I., & Mihalas, D. 2014, *Theory of Stellar Atmospheres* (Princeton, NJ: Princeton Univ. Press)
- Madhusudhan, N., Knutson, H., Fortney, J. J., & Barman, T. 2014, in *Protostars and Planets VI*, ed. H. Beuther et al. (Tucson, AZ: Univ. of Arizona Press), 739
- Mihalas, D. 1970, *Stellar Atmospheres* (San Francisco, CA: Freeman)
- Mihalas, D., & Luebke, W. R. 1971, *MNRAS*, **153**, 229
- Münch, G. 1946, *ApJ*, **104**, 87
- Parmentier, V., & Guillot, T. 2014, *A&A*, **562**, A133