



Planet–Disk Interaction in Disks with Cooling: Basic Theory

Ryan Miranda¹ and Roman R. Rafikov^{1,2} ¹ Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA; miranda@ias.edu² Centre for Mathematical Sciences, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

Received 2019 November 8; revised 2020 February 10; accepted 2020 February 20; published 2020 March 30

Abstract

Gravitational coupling between young planets and their parent disks is often explored using numerical simulations, which typically treat the disk thermodynamics in a highly simplified manner. In particular, many studies adopt the locally isothermal approximation, in which the disk temperature is a fixed function of the stellocentric distance. We explore the dynamics of planet-driven density waves in disks with more general thermodynamics, in which the temperature is relaxed toward an equilibrium profile on a finite cooling timescale t_c . We use both linear perturbation theory and direct numerical simulations to examine the global structure of density waves launched by planets in such disks. A key diagnostic used in this study is the behavior of the wave angular momentum flux (AMF), which directly determines the evolution of the underlying disk. The AMF of free waves is constant for slowly cooling (adiabatic) disks but scales with the disk temperature for rapidly cooling (and locally isothermal) disks. However, cooling must be extremely fast, with $\beta = \Omega t_c \lesssim 10^{-3}$ for the locally isothermal approximation to provide a good description of density wave dynamics in the linear regime (relaxing to $\beta \lesssim 10^{-2}$ when nonlinear effects are important). For intermediate cooling timescales, density waves are subject to a strong linear damping. This modifies the appearance of planet-driven spiral arms and the characteristics of axisymmetric structures produced by massive planets: in disks with $\beta \approx 0.1$ – 1 , a near-thermal mass planet opens only a single wide gap around its orbit, in contrast to the several narrow gaps produced when cooling is either faster or slower.

Unified Astronomy Thesaurus concepts: [Protoplanetary disks \(1300\)](#); [Planet formation \(1241\)](#); [Hydrodynamics \(1963\)](#)

1. Introduction

The gravitational interaction of a gaseous disk with a massive orbital companion plays an important role in many astrophysical systems, including circumstellar (i.e., protoplanetary) disks, cataclysmic variables, and disk galaxies. The tidal gravitational potential of the companion excites density waves at Lindblad resonances—locations in the disk at which the natural frequency of the disk is commensurate with the forcing frequency of the companion (Goldreich & Tremaine 1980). These waves then travel across the disk carrying angular momentum and energy with them over large distances. Their dissipation, either due to linear damping (e.g., viscous damping, Takeuchi et al. 1996), or nonlinear dissipation (Goodman & Rafikov 2001; Rafikov 2002a), leads to the deposition of the wave angular momentum into the background disk fluid and completes the process of the global angular momentum transport in the disk (Lunine & Stevenson 1982; Goldreich & Nicholson 1989; Rafikov & Petrovich 2012). This transfer of angular momentum from the density waves after their dissipation can be a significant driver of disk evolution (Goodman & Rafikov 2001; Rafikov 2016; Arzamasskiy & Rafikov 2018), often resulting in the formation of axisymmetric features such as gaps and rings (Rafikov 2002b; Dong et al. 2017). The detailed outcome of such planet–disk coupling depends critically on the angular momentum flux (AMF) carried by the waves. As shown by Goldreich & Tremaine (1979), in adiabatic disks, the AMF of free waves (i.e., not subject to external torques) is conserved in the linear regime and in the absence of dissipation. However, for other thermodynamic assumptions, the AMF behavior may change.

In numerical simulations of protoplanetary disks, the disk thermodynamics are often treated in a highly simplified manner by using the so-called locally isothermal approximation. In this

approximation, the sound speed c_s , or equivalently, the disk temperature T , is assumed to be a prescribed function of the radial coordinate r only, dispensing with the need to solve an energy equation for the disk gas. In Miranda & Rafikov (2019b), we showed that in the locally isothermal disks, the AMF of free waves is *not conserved*, in contrast to the adiabatic disks studied in Goldreich & Tremaine (1979). Instead, AMF is proportional to c_s^2 . Since typically the disk temperature decreases with radius, this means that waves traveling inward *accumulate AMF* as they propagate. This occurs as a result of *extracting* angular momentum from the background disk flow (so that the total angular momentum of the disk-wave system is conserved), an effect previously pointed out by Lin & Papaloizou (2011) and Lin (2015). This has important consequences for wave-driven disk evolution, since the impact of a (dissipating) wave on the disk gets enhanced by this AMF amplification process as the wave propagates to smaller and smaller radii in locally isothermal disks (Miranda & Rafikov 2019b).

Adiabatic and locally isothermal disks represent the extreme limits of a more general thermodynamics, in which the disk temperature is relaxed toward an equilibrium profile on a finite timescale. Physically, the locally isothermal approximation corresponds to the scenario in which (1) the imposed temperature profile is maintained externally, e.g., by irradiation from the central star, and (2) deviations from the imposed temperature profile, associated with either compression/expansion of the gas or radial displacement of fluid elements, are quickly neutralized by the radiation or absorption of thermal energy. Here “quickly” means that the timescale for erasing temperature perturbations, which we refer to loosely as the cooling timescale t_c , is small compared to all other relevant

timescales. On the contrary, the adiabatic limit is expected to apply when t_c is very long. A common—but not rigorously motivated—assumption is that the relevant timescale separating these two limits and to which t_c should be compared is the orbital timescale.

In this paper, we carry out a linear perturbation analysis for disks with thermodynamics affected by thermal relaxation (cooling), in order to understand the behavior of planet-excited density waves and implications for wave-driven disk evolution. An important result of this analysis is the derivation of a “master equation” describing the global behavior of non-axisymmetric perturbations driven by an external gravitational potential in a two-dimensional (2D) disk. Such an equation was first presented by Goldreich & Tremaine (1979), for the case of adiabatic perturbations in disks with uniform entropy, and was later generalized to disks with general entropy profiles (Baruteau & Masset 2008; Tsang 2014). We present an even more general version of the equation for adiabatic perturbations in disks with thermal relaxation, which reduces to the locally isothermal and adiabatic regimes in the appropriate limits (short and long cooling timescales, respectively). By solving for the perturbations excited by an embedded planet, we determine the full, global structure of planet-driven density waves (e.g., Ogilvie & Lubow 2002; Miranda & Rafikov 2019a). These calculations are corroborated using fully nonlinear numerical simulations of low-mass planets in disks with cooling.

Another key result of this paper is the analysis of the behavior of the AMF for free waves in disks with cooling/thermal relaxation in the linear regime. We not only confirm that the AMF behavior reduces to the adiabatic limit (i.e., conserved) for sufficiently long t_c , and to the locally isothermal limit (i.e., proportional to c_s^2) for sufficiently short t_c , but we also *quantify* the conditions on the cooling timescale required for these regimes to be realized. In particular, we show that the condition required for the locally isothermal approximation to be valid is much more stringent than expected from simply requiring that the cooling timescale t_c is smaller than the orbital timescale. Instead, t_c must be a very small fraction ($\lesssim 10^{-3}$) of the orbital timescale. We also show that cooling leads to a strong linear damping of density waves for a range of cooling timescales. This has significant consequences for disk evolution driven by density waves.

The plan for this paper is as follows. In Section 2, we present the linear analysis for non-axisymmetric perturbations driven by an external potential in disks with different thermodynamic assumptions, deriving the master equation describing the global structure of perturbations in each case. We numerically validate the linear analysis using numerical simulations in Section 3. In Section 4, we analyze the behavior of the wave AMF under the different thermodynamic assumptions, including a discussion of the disk torques. In Section 5, we explore the role of cooling on disk evolution driven by a massive planet. We discuss our results in Section 6 and summarize our main conclusions in Section 7.

2. Linear Perturbation Theory with Different Thermodynamic Assumptions

In this section, we derive different versions of the master equation for non-axisymmetric linear perturbations driven by an external potential in two-dimensional disks. We consider three different thermodynamic assumptions: (i) adiabatic perturbations in disks with radially varying entropy profiles

(“adiabatic disks”), (ii) isothermal perturbations in disks with fixed radial temperature profiles (“locally isothermal disks”), and (iii) perturbations in disks in which the internal energy (or temperature) is relaxed toward a prescribed profile on a finite cooling timescale (“disks with cooling”). We progress in a pedagogical fashion to highlight the differences arising due to varying thermodynamic assumptions. In each case, we discuss the reduction to the master equation of Goldreich & Tremaine (1979) for adiabatic perturbations in uniform entropy disks.

2.1. Basic Assumptions and General Approach

We consider an inviscid two-dimensional gas disk that is subject to an external potential. The disk is described in polar coordinates (r, ϕ) by the surface density Σ , height-integrated pressure $P = c_{s,\text{iso}}^2 \Sigma$ (where $c_{s,\text{iso}} = (k_B T / \mu)^{1/2}$ is the isothermal sound speed), radial velocity u_r , and azimuthal velocity u_ϕ . The unperturbed disk is axisymmetric and described by the density $\Sigma_0(r)$, pressure $P_0(r)$, radial velocity $u_{r,0}(r) = 0$ and azimuthal velocity $u_{\phi,0}(r) = r\Omega(r)$, where $\Omega(r)$ is the rotation frequency. We consider perturbations to the background state, $\Sigma = \Sigma_0 + \delta\Sigma$, $P = P_0 + \delta P$, $u_r = u_{r,0} + \delta u_r$, and $u_\phi = u_{\phi,0} + \delta u_\phi$. For convenience, we will typically drop the subscripts from the unperturbed variables. The perturbed quantities are assumed to have the form of Fourier harmonics, i.e.,

$$\delta x(r, \phi, t) = \delta x(r) \exp[i m(\phi - \omega_p t)], \quad (1)$$

for any perturbed variable δx . Here, ω_p is the pattern frequency of the perturbation. The perturbed variables satisfy the following dynamical (mass and momentum conservation) equations:

$$-i\tilde{\omega}\delta\Sigma + \frac{1}{r}\frac{\partial}{\partial r}(r\Sigma\delta u_r) + \frac{im\Sigma}{r}\delta u_\phi = 0, \quad (2)$$

$$-i\tilde{\omega}\delta u_r - 2\Omega\delta u_\phi = -\frac{1}{\Sigma}\frac{\partial}{\partial r}\delta P + \frac{1}{\Sigma^2}\frac{dP}{dr}\delta\Sigma - \frac{\partial}{\partial r}\Phi_m, \quad (3)$$

$$-i\tilde{\omega}\delta u_\phi + \frac{\kappa^2}{2\Omega}\delta u_r = -\frac{im}{r}\left(\frac{\delta P}{\Sigma} + \Phi_m\right). \quad (4)$$

Here, $\tilde{\omega} = m(\omega_p - \Omega)$ is the Doppler-shifted frequency of the perturbation, Φ_m is the Fourier component of the external potential that has azimuthal number m and rotates at the rate ω_p , and $\kappa^2 = (2\Omega/r)(r^2\Omega)'$ is the squared radial epicyclic frequency (the prime denotes the radial derivative).

In order to provide a full description of behavior of the perturbations, Equations (2)–(4) must be supplemented with an equation of state (EoS), which relates P to Σ , as well as to other thermodynamic quantities. Specifying an EoS provides a fourth perturbation equation relating δP to $\delta\Sigma$ (and potentially other fluid variables). This equation, along with Equations (2)–(4), then form a closed system. Through algebraic substitution, these equations can be combined into a single equation, or master equation, for one variable only.

We choose as the preferred variable the “enthalpy” perturbation $\delta h = \delta P / \Sigma$. Note that δh is strictly equal to the true thermodynamic enthalpy perturbation only in isentropic disks. Nonetheless, the variable δh defined in this way serves as a convenient variable for which a master equation can be found.

Several useful intermediate results in the derivation of the final master equation for δh for each of the different thermodynamic assumptions are given in Appendices A and B. In Appendix A,

Equations (2)–(4) are expressed with $\delta\Sigma$ and δP eliminated in favor of δh only, and in Appendix B, the velocity perturbations δu_r and δu_ϕ are given in terms of the enthalpy perturbation δh and its radial derivative. The velocity perturbations expressed in this way are also useful for the analysis of AMF conservation presented in Section 4.

2.2. Adiabatic Disks

The master equation for planet-driven waves in an adiabatic, non-barotropic disk (i.e., with a radially varying entropy S) has been derived previously (Baruteau & Masset 2008; Tsang 2014). For completeness, we briefly restate its derivation here.

We assume an ideal equation of state

$$P = (\gamma - 1)e\Sigma, \quad (5)$$

where γ is the adiabatic index, and e is the specific internal energy, which is related to the adiabatic sound speed of the disk according to $c_{s,\text{adi}}^2 = \gamma(\gamma - 1)e = \gamma c_{s,\text{iso}}^2$. For adiabatic perturbations, the fluid entropy $S \propto \ln(P/\Sigma^\gamma)$ is conserved in the Lagrangian sense, i.e., $dS/dt = 0$. This results in the energy equation for the total (background + perturbation) P and Σ ,

$$\frac{de}{dt} + P \frac{d}{dt} \left(\frac{1}{\Sigma} \right) = 0. \quad (6)$$

The energy equation for the perturbed fluid variables is, therefore,

$$-i\tilde{\omega}(\delta P - c_{s,\text{adi}}^2 \delta\Sigma) = -\frac{\Sigma c_{s,\text{adi}}^2}{L_S} \delta u_r. \quad (7)$$

Here, we have defined

$$\frac{1}{L_S} = \frac{1}{\gamma} \frac{dS}{dr}, \quad (8)$$

the inverse length scale of entropy variation.

Equations (2)–(4) and (7) are combined into a single second-order equation for δh , which we represent in the form

$$\frac{d^2}{dr^2} \delta h + C_1 \frac{d}{dr} \delta h + C_0 \delta h = \Psi_m, \quad (9)$$

where

$$C_1 = \frac{d}{dr} \ln \left(\frac{r\Sigma}{D_S} \right), \quad (10)$$

$$C_0 = -\frac{2m\Omega}{r\tilde{\omega}} \left[\frac{2}{L_S} + \frac{d}{dr} \ln \left(\frac{\Sigma\Omega}{D_S} \right) \right] - \frac{1}{L_S^2} - \frac{1}{L_S} \frac{d}{dr} \ln \left(\frac{r\Sigma}{L_S D_S} \right) - \frac{m^2}{r^2} \left(1 - \frac{N_r^2}{\tilde{\omega}^2} \right) - \frac{D_S}{c_{s,\text{adi}}^2}, \quad (11)$$

and the forcing due to the planetary potential is

$$\begin{aligned} \Psi_m = & -\frac{d^2 \Phi_m}{dr^2} - \left[\frac{1}{L_S} + \frac{d}{dr} \ln \left(\frac{r\Sigma}{D_S} \right) \right] \frac{d\Phi_m}{dr} \\ & + \left\{ \frac{2m\Omega}{r\tilde{\omega}} \left[\frac{d}{dr} \ln \left(\frac{\Sigma\Omega}{D_S} \right) + \frac{1}{L_S} \right] \right. \\ & \left. + \frac{m^2}{r^2} \left(1 - \frac{N_r^2}{\tilde{\omega}^2} \right) \right\} \Phi_m. \end{aligned} \quad (12)$$

Here,

$$D_S = \kappa^2 - \tilde{\omega}^2 + N_r^2, \quad (13)$$

and

$$N_r^2 = -\frac{1}{\Sigma^2} \frac{dP}{dr} \left(\frac{1}{c_{s,\text{adi}}^2} \frac{dP}{dr} - \frac{d\Sigma}{dr} \right) \quad (14)$$

is the square of the Brunt–Väisälä frequency.

For barotropic disks ($S = \text{constant}$), $L_S \rightarrow \infty$ and $N_r \rightarrow 0$, resulting in the reduction of Equations (10)–(12) to the Goldreich & Tremaine (1979) equation.

2.3. Locally Isothermal Disks

In the locally isothermal treatment of the disk thermodynamics, a fixed temperature profile $T(r)$ is assumed. This assumption corresponds to the situation in which $T(r)$ is set by, e.g., irradiation by the central star, and in which radiative cooling eliminates any temperature variations very quickly. As a result, $c_{s,\text{iso}}^2$ is a fixed function of r , eliminating the need for an explicit energy equation. The locally isothermal approximation has been widely used in numerical simulations, but to the best of our knowledge, the full master equation for planet-driven waves has not been formulated before (Lee 2016 has previously derived only the homogeneous part of the equation).

As a result of making the locally isothermal assumption, the EoS (in terms of the total P and Σ) is expressed as

$$P = c_{s,\text{iso}}^2(r) \Sigma, \quad (15)$$

and corresponding equation for the perturbed variables is simply

$$\delta P = c_{s,\text{iso}}^2(r) \delta\Sigma. \quad (16)$$

Combining Equations (2)–(4) and (16), we find a master equation for δh (see Equation (9)):

$$C_1 = \frac{d}{dr} \ln \left(\frac{r\Sigma}{D} \right) - \frac{1}{L_T}, \quad (17)$$

$$\begin{aligned} C_0 = & -\frac{2m\Omega}{r\tilde{\omega}} \left[\frac{1}{L_T} + \frac{d}{dr} \ln \left(\frac{\Sigma\Omega}{D} \right) \right] \\ & - \frac{1}{L_T} \frac{d}{dr} \ln \left(\frac{r\Sigma}{L_T D} \right) - \frac{m^2}{r^2} - \frac{D}{c_{s,\text{iso}}^2}, \end{aligned} \quad (18)$$

and the forcing is

$$\Psi_m = -\frac{d^2\Phi_m}{dr^2} - \left[\frac{d}{dr} \ln \left(\frac{r\Sigma}{D} \right) \right] \frac{d\Phi_m}{dr} + \left\{ \frac{2m\Omega}{r\tilde{\omega}} \left[\frac{d}{dr} \ln \left(\frac{\Sigma\Omega}{D} \right) \right] + \frac{m^2}{r^2} \right\} \Phi_m. \quad (19)$$

Here, we have defined

$$D = \kappa^2 - \tilde{\omega}^2, \quad (20)$$

and

$$\frac{1}{L_T} = \frac{d \ln c_s^2}{dr}, \quad (21)$$

the length scale of the variation of $T \propto c_s^2$ (here c_s refers to either adiabatic or isothermal sound speed).

Note that if we take the limit $\gamma \rightarrow 1$ in the adiabatic case (Section 2.2), then $D_S \rightarrow D$, $L_S \rightarrow L_T$ and $N_r \rightarrow 0$. However, the Expressions (10)–(12) do not reduce to their locally isothermal analogues (17)–(19) in this limit. This observation highlights the *singular nature* of the locally isothermal approximation and clearly shows why this approximation leads to different results when compared to the adiabatic case with $\gamma \rightarrow 1$, something that has been pointed out in Miranda & Rafikov (2019b).

Only in the *globally* isothermal limit ($T = \text{const}$), when $L_T \rightarrow \infty$, Equations (10)–(12) become identical to the Equations (17)–(19) and reduce to the master equation of Goldreich & Tremaine (1979).

2.4. Disks with Cooling

The adiabatic and locally isothermal disks considered in the previous subsections represent limiting cases of a more general disk thermodynamics, in which the disk temperature is relaxed toward an equilibrium profile on a finite timescale. We now derive the generalized master equation for disks with cooling.

Analogous to adiabatic disks, we adopt an ideal equation of state. We add a cooling term on the right-hand side of Equation (6), which relaxes e toward a prescribed equilibrium profile $e_0(r) = c_{s,\text{adi}}^2(r)/[\gamma(\gamma - 1)]$ on a cooling³ timescale t_c :

$$\left(\frac{\partial e}{\partial t} \right)_{\text{cool}} = -\frac{e - e_0}{t_c}. \quad (22)$$

Note that we allow t_c to be an arbitrary function of r . Equations (6) and (22) lead to the energy equation for the perturbed fluid variables,

$$\left(\frac{1}{t_c} - i\tilde{\omega} \right) \delta P - \left(\frac{1}{\gamma t_c} - i\tilde{\omega} \right) c_{s,\text{adi}}^2 \delta \Sigma = -\frac{\Sigma c_{s,\text{adi}}^2}{L_S} \delta u_r. \quad (23)$$

Note that Equation (23) reduces to corresponding Equation (16) for locally isothermal disks in the limit $t_c \rightarrow 0$ (noting that $c_{s,\text{adi}}^2/\gamma = c_{s,\text{iso}}^2$), and to the corresponding Equation (7) for adiabatic disks in the limit $t_c \rightarrow \infty$.

Combining Equations (2)–(4) and (23), we find that the master equation for the general cooling case is (see

³ The energy source term described by Equation (22) represents both heating and cooling of the gas toward a fixed temperature. Therefore, it represents thermal relaxation rather than strictly cooling. However, we will nonetheless loosely refer to it as “cooling.”

Equation (9)):

$$C_1 = \frac{d}{dr} \ln \left(\frac{r\Sigma}{D_c} \right) - \frac{1}{(1 - i\gamma\tilde{\beta})L_T}, \quad (24)$$

$$C_0 = -\frac{2m\Omega}{r\tilde{\omega}} \left[\left(\frac{L_T^{-1} - 2i\gamma\tilde{\beta}L_S^{-1}}{1 - i\gamma\tilde{\beta}} \right) + \frac{d}{dr} \ln \left(\frac{\Sigma\Omega}{D_c} \right) \right] + \frac{\gamma\tilde{\beta}L_S^{-1}(\gamma\tilde{\beta}L_S^{-1} + iL_T^{-1})}{(1 - i\gamma\tilde{\beta})^2} - \left(\frac{L_T^{-1} - i\gamma\tilde{\beta}L_S^{-1}}{1 - i\gamma\tilde{\beta}} \right) \frac{d}{dr} \ln \left[\frac{r\Sigma}{D_c} \left(\frac{L_T^{-1} - i\gamma\tilde{\beta}L_S^{-1}}{1 - i\gamma\tilde{\beta}} \right) \right] - \left(\frac{\kappa^2 - D_c}{\tilde{\omega}^2} \right) \frac{m^2}{r^2} - \left(\frac{1 - i\tilde{\beta}}{1 - i\gamma\tilde{\beta}} \right) \frac{\gamma D_c}{c_{s,\text{adi}}^2}, \quad (25)$$

with forcing due to the planetary potential

$$\Psi_m = -\frac{d^2\Phi_m}{dr^2} + \left[\frac{i\gamma\tilde{\beta}}{(1 - i\gamma\tilde{\beta})L_S} - \frac{d}{dr} \ln \left(\frac{r\Sigma}{D_c} \right) \right] \frac{d\Phi_m}{dr} + \left\{ \frac{2m\Omega}{r\tilde{\omega}} \left[\frac{d}{dr} \ln \left(\frac{\Sigma\Omega}{D_c} \right) - \frac{i\gamma\tilde{\beta}}{(1 - i\gamma\tilde{\beta})L_S} \right] + \left(\frac{\kappa^2 - D_c}{\tilde{\omega}^2} \right) \frac{m^2}{r^2} \right\} \Phi_m. \quad (26)$$

Here, $\tilde{\beta} = \tilde{\omega}t_c$ and

$$D_c = D - \frac{i\gamma\tilde{\omega}t_c}{1 - i\gamma\tilde{\omega}t_c} N_r^2. \quad (27)$$

It is easily verified that Equations (24)–(26) reduce to the master equation for adiabatic disks (Equations (10)–(12)) for $t_c \rightarrow \infty$, and to the master equation for locally isothermal disks (Equations (17)–(19)) for $t_c \rightarrow 0$, as expected. Note that the coefficients C_1 and C_0 (as well as the coefficients of Φ_m in the source term Ψ_m) are real in these limits but are complex in general for a finite t_c .

3. Numerical Validation

The master Equations (9) and (24)–(26) can be used to analytically determine the details of the density wave excitation and their subsequent propagation in the linear regime, as has been done in a number of past studies (Goldreich & Tremaine 1979; Zhang & Lai 2006; Tsang 2014). However, in this work, we take a different approach, making use of numerical solutions of the master equation (following e.g., Korycansky & Pollack 1993; Rafikov & Petrovich 2012) to characterize the waves excited by low-mass planets in protoplanetary disks.

We consider the interaction of a disk around a star of mass M_* with a planet of mass $M_p \ll M_*$. The planet is assumed to have a circular orbit with radius r_p and orbital period $t_p = 2\pi/\Omega_p$, where $\Omega_p = (GM_*/r_p^3)^{1/2}$. We present numerical solutions of the different versions of the master equation for planet-driven waves derived before and describe the basic global structure of the perturbations. We also present the results of direct numerical simulations of low-mass planets in disks,

for which the disk response is well-approximated by the linear regime, in order to validate our linear analysis.

3.1. Disk Model

The analysis presented in Section 2 is valid for disks with arbitrary profiles for the disk surface density $\Sigma(r)$ and temperature $T(r)$. For the purposes of the following numerical calculations, it is necessary to adopt a concrete disk model, and so we choose a simple power-law disk. The unperturbed disk has an *isothermal* sound speed (see Section 6.3) given by

$$c_{s,\text{iso}}(r) = h_p r_p \Omega_p \left(\frac{r}{r_p} \right)^{-q/2}, \quad (28)$$

where h_p is the disk aspect ratio, $h(r) = H/r = h_p (r/r_p)^{(1-q)/2}$, at r_p . Here, $H = c_{s,\text{iso}}/\Omega$ is the pressure scale height. The parameter q is the power-law index of the disk temperature $T \propto c_{s,\text{iso}}^2$. The surface density profile is

$$\Sigma(r) = \Sigma_p \left(\frac{r}{r_p} \right)^{-p}, \quad (29)$$

where the value of Σ_p is arbitrary and p is a constant. Accounting for the radial pressure support, the orbital frequency and radial epicyclic frequency resulting from centrifugal balance are, therefore,

$$\Omega = \Omega_K [1 - h^2(r)(q + p)]^{1/2}, \quad (30)$$

$$\kappa = \Omega_K [1 - h^2(r)(q + p)(2 - q)]^{1/2}, \quad (31)$$

where $\Omega_K = (GM_*/r^3)^{1/2}$ is the Keplerian frequency. The (inverse) length scales of the variation of entropy and temperature (Equations (8) and (21)) are

$$\frac{1}{L_S} = \frac{(\gamma - 1)p - q}{\gamma r}, \quad \frac{1}{L_T} = -\frac{q}{r}. \quad (32)$$

We choose the disk aspect ratio at the location of the planet $h_p = 0.1$, and temperature power-law index $q = 1$. As a result, the aspect ratio $h(r)$ is independent of r (this a theoretical convenience and is not necessarily intended to describe realistic disks). We choose a surface density power-law index $p = 1$, and (except in our locally isothermal calculations) adiabatic index $\gamma = 7/5$.

3.2. Linear Calculations

We compute the linear response of the disk to an orbiting planet by solving the master equation for either disks with cooling (Equation (9) with C_1 , C_0 , and Ψ_m given by Equations (24)–(26)) or for locally isothermal disks (in which case C_1 , C_0 , and Ψ_m are given by Equations (17)–(19)), for modes with different azimuthal numbers. The solution method, which closely follows that of Korycansky & Pollack (1993), is described in detail in the appendix of Miranda & Rafikov (2019a).⁴ We solve for the structure of modes with $m \leq m_{\text{max}}$, where $m_{\text{max}} = 80$ is sufficient to achieve convergence of the perturbation structure.

⁴ In setting the outgoing wave boundary conditions using the asymptotic (WKB) behavior of solutions, the variation of the wave amplitude is modified as appropriate based on which version of the master equation is being solved. The amplitude variation follows from the behavior of the wave AMF, see Section 4.

We then construct the full two-dimensional structure of the perturbed fluid variables by synthesizing them in real space.

For planets on circular orbits, the Fourier harmonics of the gravitational potential are

$$\Phi_m = -\frac{GM_p}{r_p} b_{1/2}^{(m)}(r/r_p), \quad (33)$$

where

$$b_{1/2}^{(m)}(\alpha) = \frac{1}{\pi} \int_0^{2\pi} \frac{\cos(m\psi) d\psi}{[1 - 2\alpha \cos(\psi) + \alpha^2 + \epsilon^2]^{1/2}} \quad (34)$$

are softened Laplace coefficients. We choose for the softening parameter $\epsilon = 0.6h_p$, corresponding to a softening length of $0.6H_p$ (representing the effect of the finite vertical extent of the disk). Note that for a circular orbit, the wave pattern frequency ω_p is equal to the orbital frequency of the planet Ω_p for all m . We ignore the indirect potential term, $\delta_{m,1} GM_p r/r_p^2$, associated with the orbital motion of the central star, as it has a negligible impact on the overall perturbation structure (Miranda & Rafikov 2019a).

3.3. Hydrodynamical Simulations

We also run 2D inviscid hydrodynamical simulations of planet–disk interaction using FARGO3D (Benítez-Llambay & Masset 2016). The numerical grid extends from $r_{\text{in}} = 0.05r_p$ to $r_{\text{out}} = 5.0r_p$ with logarithmic spacing in the radial direction and uniform spacing in the azimuthal direction. The number of grid cells is $N_r \times N_\phi = 3004 \times 4096$. Wave damping (de Val-Borro et al. 2006) is applied near the inner and outer boundaries ($r < 0.06r_p$ and $r > 4.5r_p$) to prevent wave reflection. The mass of the planet is gradually increased from zero to M_p over 10 orbits, and its potential is softened with a length scale $0.6H_p$, as in the linear calculations.

We choose either a locally isothermal EoS, or an ideal EoS with $\gamma = 7/5$ and cooling (Equation (22)). The same initial temperature profile is used in both cases, in order to provide the most direct comparison. Cooling is implemented using a simple implicit (backward Euler) step performed after the main hydro step. This implementation is stable and performs as expected (based on agreement with linear theory) for cooling timescales at least as small as $10^{-4} \Omega^{-1}$.

We choose a planet mass $M_p = 10^{-5} M_* = 0.01 M_{\text{th}}$, where

$$M_{\text{th}} = h_p^3 M_* \quad (35)$$

is the thermal mass. Since $M_p \ll M_{\text{th}}$, the response of the disk to the planet is well-approximated by the linear regime—at least close to the planet. The waves may undergo significant nonlinear evolution and develop into shocks at large distances from the planet. However, as described in Miranda & Rafikov (2019a), the waves excited by a $0.01 M_{\text{th}}$ planet never develop into shocks in the inner disk, as a result of the partitioning of the wave AMF into multiple spiral arms (see Rafikov 2002a). Shocks do develop at large radii in the outer disk, where there is only one spiral arm carrying all of the AMF.

3.4. Structure of Perturbations

Figure 1 shows the disk surface density perturbation, $\delta\Sigma$ (in terms of the background surface density Σ) for a planet in a locally isothermal disk and in disks with different dimensionless cooling timescales $\beta = \Omega t_c$. The chosen values of β , 10^{-2} – 10^2 ,

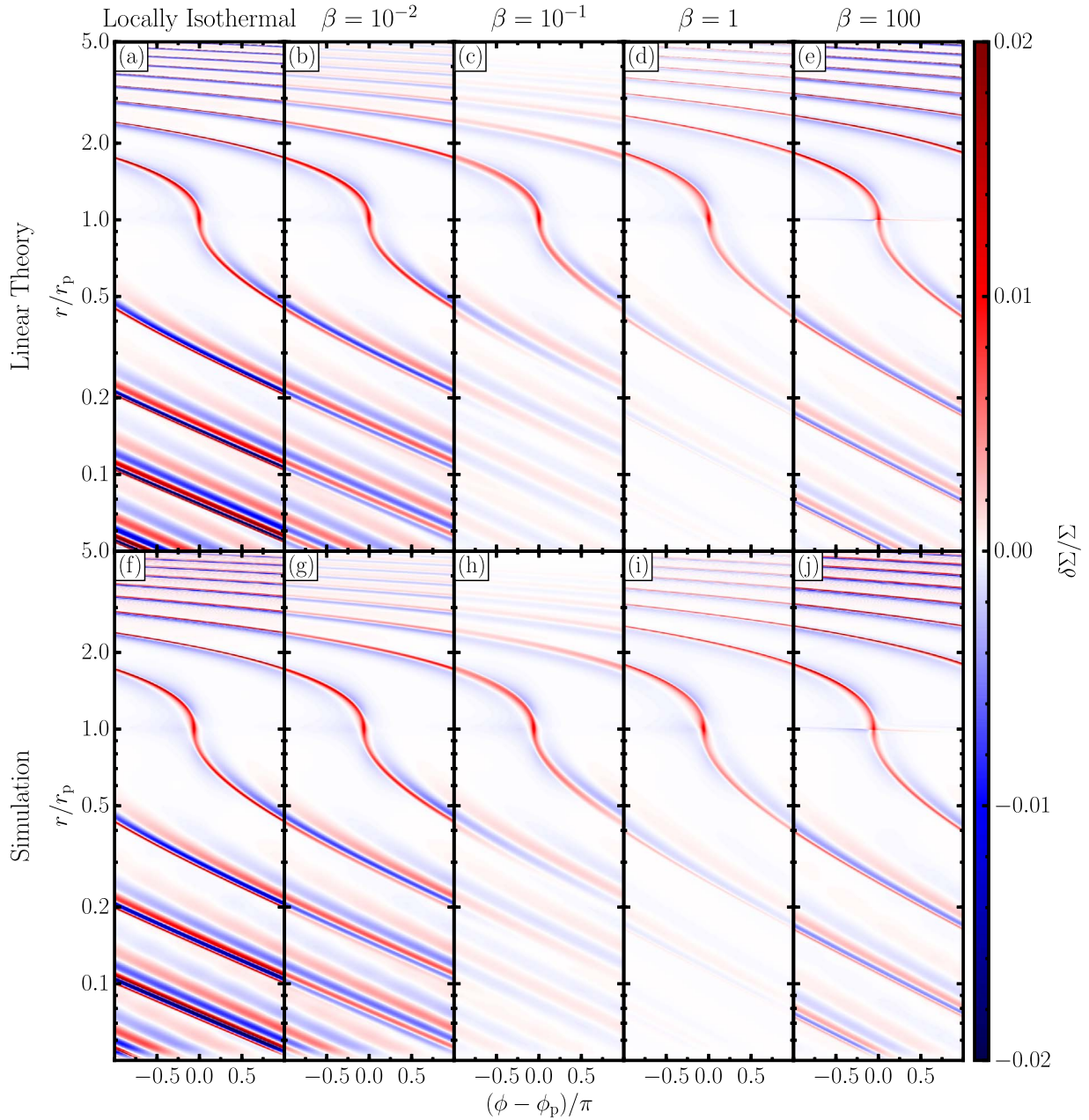


Figure 1. Fractional surface density perturbation $\delta\Sigma/\Sigma$, shown in polar coordinates for disks with different thermodynamics: locally isothermal ((a) and (f)), and with different constant values of the dimensionless cooling timescale $\beta = \Omega t_c$ (all other panels). The top row ((a)–(e)) shows the results of numerical solutions of the linear perturbation Equations (9) and (17)–(19) in panel (a) and Equations (9) and (24)–(26) in other panels, and the bottom row ((f)–(j)) shows the results of nonlinear simulations with a $0.01M_{\text{th}}$ planet (for which the perturbation is well in the linear regime) at 20 orbits. The disk has an aspect ratio $h_p = 0.1$, temperature power-law index $q = 1$, surface density power-law index $p = 1$, and (for the non-locally isothermal calculations) adiabatic index $\gamma = 7/5$.

along with the locally isothermal case, representing the limit $\beta \rightarrow 0$, represent the full range of behaviors for the different thermodynamics we consider. The results of both numerical solutions of the linear master equation (Figures 1(a)–(e)) and numerical simulations (Figures 1(f)–(j)) are shown. The simulation results show excellent agreement with the linear results, down to fine details, confirming the validity of our linear analysis, as well as of our numerical implementation of cooling in the simulations.

Generically, the perturbations take the form of a single spiral arm (i.e., an azimuthally narrow maximum of $\delta\Sigma$) in the outer disk, and multiple spirals (as many as four) gradually emerging in the inner disk. The spirals are the result of the complicated

interference of perturbations with different azimuthal mode numbers, which can be understood based on the dispersion relation for linear density waves (Bae & Zhu 2018; Miranda & Rafikov 2019a). This basic mechanism is evidently not strongly affected by the intricacies of the disk thermodynamics. However, the detailed structure of the spirals is affected by a particular choice of the nonzero cooling timescale in several key ways.

The first effect is that the radial range of the waves depends strongly on the cooling timescale. For fast cooling ($\beta \lesssim 10^{-2}$; Figures 1(a) and (b)) or slow cooling (e.g., $\beta = 100$; Figure 1(e)), the waves propagate far into both the inner and outer disk. That is, the wave amplitude at radii far from the

planet is comparable to (or in some cases larger than) the wave amplitude near the planet. However, for the cases with $\beta = 10^{-1}$ and $\beta = 1$ (Figures 1(c) and (d)), the wave amplitude decreases with distance from the planet. The waves are most strongly damped in the inner disk for the case $\beta = 1$ (Figure 1(d)), for which the amplitude decreases (relative to the amplitude close to the planet) by nearly an order of magnitude by the time the waves reach $0.5r_p$. In the outer disk, waves are most strongly damped when the cooling time is somewhat shorter, e.g., the case with $\beta = 10^{-1}$ (Figure 1(c)). In this case, the wave amplitude has decreased by an order of magnitude when the waves reach $2.5r_p$.

A second effect is that the overall tightness or pitch angle of the spirals (i.e., slope of density contours in Figure 1) varies with the value of β . The spirals are most tightly wound for very small values of β , and least tightly wound for large values of β . This is related to the dependence of the effective sound speed for density waves on the cooling timescale; see Section 6.2 for a more in-depth discussion.

Finally, the details of the structure of the multiple spirals in the inner disk—their relative amplitudes, separations, and azimuthal widths—are modified by cooling. Although we forgo a detailed analysis of the multiple spiral structure as carried out in Miranda & Rafikov (2019a), trends regarding the multiple spiral structure are evident in Figure 1. For $\beta = 10^{-1}$ –1, the primary spiral arm (i.e., the one that touches the planet) is wider than it is for either smaller or larger values of β , and the primary and secondary spirals are more widely separated. The case with $\beta = 10^{-1}$ represents an extreme modification of the spiral structure; far in the inner disk ($r \lesssim 0.1r_p$), the surface density is almost a pure $m = 2$ sinusoid. This pattern is very distinct from the narrow, azimuthally concentrated spiral arms found when β is very large or very small; it will be discussed further in Section 6.2.

4. Angular Momentum Flux

We can gain insight into the behavior of density waves by considering the behavior of the wave AMF,

$$F_J(r) = r^2 \Sigma(r) \oint \delta u_r(r, \phi) \delta u_\phi(r, \phi) d\phi, \quad (36)$$

where δu_r and $\delta u_\phi = u_\phi - r\Omega$ are the velocity perturbations. For a Fourier mode with azimuthal number m , the mode AMF is given in terms of the complex perturbations $\delta u_r(r)$, $\delta u_\phi(r) \propto \exp(im\phi)$ by

$$F_J^m = \pi r^2 \Sigma \operatorname{Re}(\delta u_r \delta u_\phi^*). \quad (37)$$

The total AMF F_J is found by summing over all of the modes,

$$F_J = \sum_{m=1}^{\infty} F_J^m. \quad (38)$$

We wish to describe the behavior of the AMF for free waves, i.e., not subject to an external potential. Evolution of the AMF of free waves is tied to wave-driven evolution of the disk (see Sections 5 and 6.4). Planet-excited waves can be considered free at locations far from the planet, where the torque density,

$$\frac{dT}{dr} = -r \oint \Sigma \frac{\partial \Phi_p}{\partial \phi} d\phi, \quad (39)$$

which describes wave excitation, is negligible (here, Φ_p is the gravitational potential of the planet). Practically speaking, this

is satisfied beyond about 2–3 scale heights H_p from the planet (Dong et al. 2011b; Rafikov & Petrovich 2012).

Our goal is to derive a conservation law for the Fourier AMF F_J^m of free waves for each of the thermodynamic assumptions described in Section 2. The general strategy is to express F_J^m (Equation (37)) in terms of δh (see Appendix B), so that the homogeneous version of the master Equation (9) for δh and either Equations (10)–(11), (17)–(18), or (24)–(25) (with $\Psi_m = 0$) can be used to ascertain the behavior of F_J .

4.1. Adiabatic Disks

For adiabatic disks, plugging δu_r and δu_ϕ given by Equations (80) and (81) into the Expression (37) for F_J , we find that the AMF for free waves is

$$F_J^m = \frac{\pi m r \Sigma}{D_s} \operatorname{Im}(\delta h \delta h^{*'}), \quad (40)$$

where the prime denotes the radial derivative. A global conservation law for F_J can be found by taking the complex conjugate of the general homogeneous version of the master Equation (9) with $\Psi_m = 0$, multiplying by δh , and taking the imaginary part of the resulting equation, leading to

$$\operatorname{Im}(\delta h \delta h^{*''}) + C_1 \operatorname{Im}(\delta h \delta h^{*'}) = 0. \quad (41)$$

On the other hand, differentiating (40) and making use of the Expression (10) for C_1 for adiabatic disks, one can easily see that Equation (41) is equivalent to

$$\frac{dF_J^m}{dr} = 0. \quad (42)$$

Therefore, F_J^m is constant or conserved (i.e., independent of r) in adiabatic disks (Goldreich & Tremaine 1979). Since the AMF of each wave mode is conserved, the *total* wave AMF must also be conserved:

$$\frac{dF_J}{dr} = 0. \quad (43)$$

4.2. Locally Isothermal Disks

For locally isothermal disks, using Expressions (82) and (83) for δu_r and δu_ϕ in Equation (37), we find that F_J is given by

$$F_J^m = \frac{\pi m r \Sigma}{D} \operatorname{Im}(\delta h \delta h^{*'}), \quad (44)$$

i.e., the same as for adiabatic disks (Equation (40)) but with $D_s \rightarrow D$. We follow the same procedure as for the case of adiabatic disks to find a global conservation law for F_J . In this case, C_1 is given by Equation (17). As a result, we see that Equation (41) is equivalent to

$$\frac{d}{dr} \left(\frac{F_J^m}{c_s^2} \right) = 0. \quad (45)$$

Therefore, F_J^m is not constant, but instead proportional to c_s^2 in locally isothermal disks (stated without a proof in Lee 2016). Since Equation (45) applies to all wave modes, the total wave

AMF obeys the same conservation law:

$$\frac{d}{dr} \left(\frac{F_J}{c_s^2} \right) = 0, \quad (46)$$

generally disagreeing with Equation (42) even when $\gamma \rightarrow 1$ in the latter. Only in the limit of a *globally* isothermal disk ($L_T \rightarrow \infty$) Equation (46) reduces to Equation (43).

4.3. Disks with Cooling

For disks with cooling, using Equations (84) and (85), we find that the Fourier AMF is given in terms of δh by

$$F_J^m = \frac{\pi m r \Sigma}{D^2 + \gamma^2 \tilde{\omega}^2 D_S^2} \left\{ (D + \gamma^2 \tilde{\omega}^2 D_S) \text{Im}(\delta h \delta h^*) - \gamma \tilde{\beta} N_r^2 \text{Re}(\delta h \delta h^*) + \gamma \tilde{\beta} \left[\frac{2m\Omega}{r\tilde{\omega}} N_r^2 - \frac{D}{L_S} + \frac{D_S}{L_T} \right] |\delta h|^2 \right\}. \quad (47)$$

In the limit $t_c \rightarrow 0$, we recover Equation (44) and for $t_c \rightarrow \infty$, we recover Equation (40).

The complexity of Equations (24)–(26) and (47) precludes us from finding a global conservation law for the Fourier AMF, as in the adiabatic and locally isothermal cases. Instead, we use a local WKB analysis to determine the approximate behavior of F_J^m (e.g., Takeuchi et al. 1996), the details of which are given in Appendix C. The result is

$$F_J^m(r) = F_J^m(r_0) \times \exp \left\{ \int_{r_0}^r \left[\frac{\Omega^2 L_T^{-1}}{\Omega^2 + \gamma^2 \tilde{\omega}^2 \beta^2} - 2 \text{Im}(k) \right] dr' \right\}, \quad (48)$$

where

$$\text{Im}(k) = \frac{(\gamma - 1)\Omega\tilde{\omega}\beta}{2(\Omega^2 + \gamma^2 \tilde{\omega}^2 \beta^2)} \text{Re}(k), \quad (49)$$

$$\text{Re}(k) = \frac{|D|^{1/2}}{c_{s,\text{eff}}}, \quad (50)$$

are the imaginary and real parts of the radial wavenumber k , $\beta = \Omega t_c$ is the dimensionless cooling time, and r_0 is an arbitrary reference radius. We have introduced the effective sound speed

$$c_{s,\text{eff}} = \left(\frac{\Omega^2 + \gamma^2 \tilde{\omega}^2 \beta^2}{\Omega^2 + \tilde{\omega}^2 \beta^2} \right)^{1/4} c_{s,\text{iso}}, \quad (51)$$

which varies from $c_{s,\text{iso}}$ for $\beta = 0$ to $\gamma^{1/2} c_{s,\text{iso}} = c_{s,\text{adi}}$ as $\beta \rightarrow \infty$. Equation (48) reduces to $F_J^m = \text{constant}$ in the adiabatic limit $\beta \rightarrow \infty$ or the globally isothermal limit $L_T \rightarrow \infty$ and $\gamma = 1$, and to $F_J^m \propto c_s^2$ in the locally isothermal limit $\beta \rightarrow 0$. The accuracy with which the WKB analysis reproduces the true behavior of the AMF components is examined in Appendix C.2.

Note that Equation (48) has explicit dependence on the azimuthal number m , unlike the equivalent expression for adiabatic (Equation (42)) or locally isothermal disks (Equation (45)). Therefore, the behavior of the total F_J , found by summing up the contributions for all azimuthal numbers, does not follow trivially from Equation (48). In general, the variation of F_J depends on the relative magnitudes of all of the F_J^m components at some reference radius.

4.3.1. Inner Disk

In general, even the WKB Equation (48) must be evaluated numerically to determine the behavior of F_J^m . However, its asymptotic behavior in the inner disk can be expressed in a simple analytic form if we assume that β is a constant, i.e., that the cooling timescale is a fixed fraction of the orbital period, as in our numerical calculations. A simple expression for the behavior of F_J^m in the inner disk is of interest because of the complex wave phenomena that occur there (e.g., multiple spiral arms). The behavior of F_J^m in the outer disk cannot be described by a simple expression for constant β , and so we leave detailed characterization of the AMF evolution in this region to the numerical results presented in Section 4.4.

At small radii in the inner disk, $\Omega \gg \omega_p$ and, hence, $\tilde{\omega} \approx -m\Omega$. Making use of this approximation, the WKB Equation (48) reduces to

$$F_J^m(r) = F_J^m(r_0) \left[\frac{c_s(r)}{c_s(r_0)} \right]^{2/(1+\gamma^2 m^2 \beta^2)} \exp \left[\xi g \left(\frac{r}{r_0} \right) \right], \quad (52)$$

where

$$\xi = \frac{(\gamma - 1)(m^2 - 1)^{1/2} m \beta}{h_0(1 + \gamma m^2 \beta^2)} \left(\frac{1 + m^2 \beta^2}{1 + \gamma m^2 \beta^2} \right)^{1/4}, \quad (53)$$

with $h_0 = h(r_0)$, and

$$g(x) = \begin{cases} 2(q - 1)^{-1} [x^{(q-1)/2} - 1] & (q \neq 1), \\ \ln x & (q = 1). \end{cases} \quad (54)$$

For the case $q = 1$, further simplification is possible:

$$F_J^m(r) = F_J^m(r_0) \left(\frac{r}{r_0} \right)^{\xi - 1/(1+\gamma^2 m^2 \beta^2)}. \quad (55)$$

There are two distinct contributions to the radial variation of the wave AMF for free waves in disks with cooling (Equation (48)). The first is related to the disk temperature profile, described by the term in Equation (48) involving L_T . As typically, the disk temperature varies on a global scale, i.e., $L_T \sim r$, this usually constitutes a slow variation. The second contribution formally describes a linear damping, as it is associated with an imaginary part of the radial wavenumber (Equation (49)). Damping occurs because thermal relaxation counteracts the adiabatic heating/cooling associated with compression/expansion of a fluid element, reducing the restoring action of the pressure for the wave.

We characterize this damping by examining the asymptotic behavior of the WKB AMF F_J^m in the inner disk, as $r \rightarrow 0$. Using Equation (52), we define the characteristic length scale of the radial variation of the AMF,

$$l_F^m = \left(\frac{d \ln F_J^m}{dr} \right)^{-1}. \quad (56)$$

Note that the sign of l_F^m is important, as it indicates whether the magnitude of F_J^m increases or decreases with radius. For inward traveling waves in the inner disk, a positive l_F^m means that F_J^m decreases as the wave propagates toward the inner disk, and a negative l_F^m means that it grows as the wave propagates inward.

We first focus on the case of a disk with a $q = 1$ temperature profile, which has a constant aspect ratio $h_0 = h = \text{const}$ in Equation (53). In this case, F_m has a power-law dependence on r in the inner disk (Equation (55)). Therefore, AMF varies with

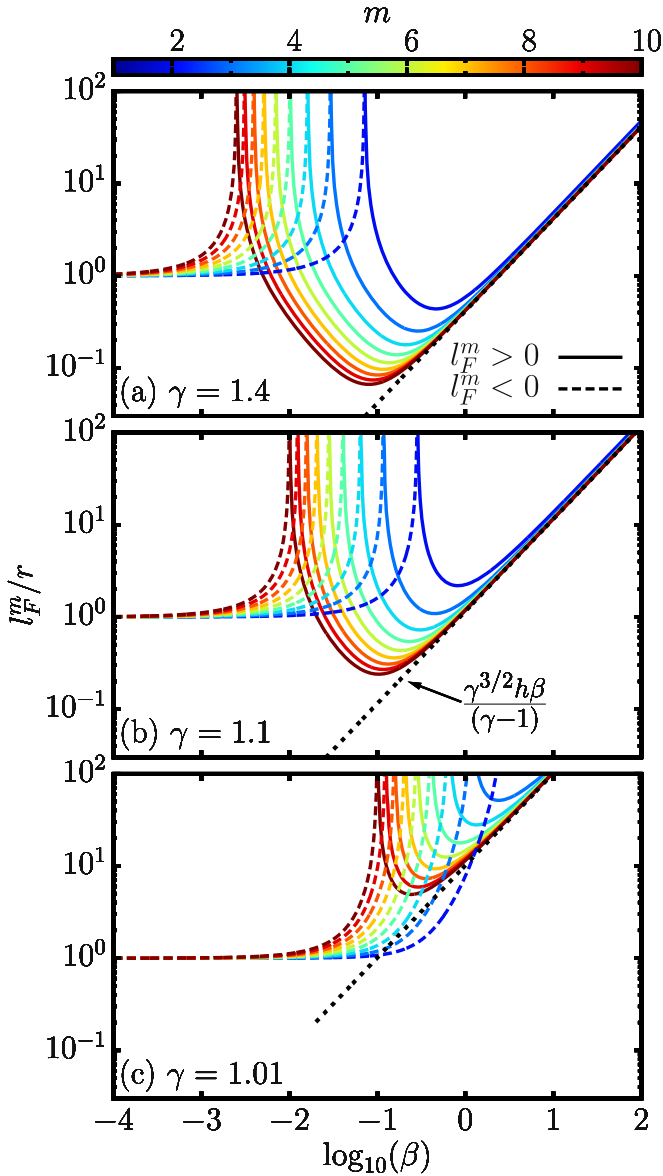


Figure 2. Asymptotic behavior of F_j^m , the AMF of Fourier modes, for free waves in the WKB limit in the inner disk, where the wave pattern frequency is slow compared to the orbital frequency. The length scale l_F^m (defined by Equation (56) and normalized by r) associated with the variation of F_j^m is shown as a function of the dimensionless cooling timescale β for different azimuthal mode numbers m (different colored lines), and for different values of the adiabatic index γ (different panels). The disk has a temperature power-law index $q = 1$ and aspect ratio $h = 0.1$. Solid lines denote that l_F^m is positive, so that F_j^m decreases toward the center of the disk, and dashed lines denote that l_F^m is negative, hence, F_j^m increases toward the center. The dotted line in each panel (labeled in (b)) indicates the asymptotic behavior for large β given by Equation (61).

a length scale proportional to r , i.e., l_F^m/r is independent of radius. Specifically, from Equation (55), we have

$$\frac{l_F^m}{r} = \left(\xi - \frac{1}{1 + \gamma^2 m^2 \beta^2} \right)^{-1}. \quad (57)$$

Because of this simplification, a disk model with $q = 1$ serves as a convenient case for analysis. Figure 2 shows l_F^m/r as a function of β for waves with different azimuthal numbers in the inner disk for a disk with $q = 1$ and aspect ratio $h = 0.1$. The

azimuthal numbers shown are $m = 2-10$, which carry most of the AMF for planet-excited waves in a disk with this aspect ratio (note that $m = 1$ waves are evanescent in the inner disk).

In Figure 2(a), we consider the case with $\gamma = 1.4$, appropriate for protoplanetary disks. We highlight several key features of the behavior of the F_j^m illustrated by this figure. First, for this γ , there is always a range of values of β for which l_F^m/r is positive and $\lesssim 1$; for the plotted values of m this range falls into an interval $10^{-2} \lesssim \beta \lesssim 1$, with higher m modes reaching lower l_F^m/r . This means that F_j^m decreases rapidly toward the inner disk, i.e., waves are *damped*, with a damping length significantly shorter than r , for all values of m .

Second, for large values of β ($\gtrsim 1$), l_F^m is positive and increases with β . For $\beta \gtrsim 10$, $l_F^m/r > 1$, corresponding to weak damping, and for $\beta \gtrsim 100$, $l_F^m/r \gtrsim 10$, so that the waves are effectively undamped and the adiabatic limit is recovered. Note also that l_F^m is nearly independent of m for values of β in this range.

Finally, for small values of β (depending on m but $\lesssim 10^{-1}$), l_F^m is negative, so that F_j^m grows as the wave propagates toward the inner disk. For $\beta \approx 10^{-4}$, $l_F^m \approx -1$ (i.e., $F_j^m \propto 1/r$) for all values of m that we consider, so that the locally isothermal limit is recovered. Note that even for β as small as 10^{-3} , l_F^m still shows significant variation with m , meaning that the locally isothermal limit is not valid.

The details of this behavior can be understood by examining Equation (57). We find that the minimum positive l_F^m , i.e., the shortest possible damping length, occurs when $\beta = \beta_{\text{crit}}$, where

$$\beta_{\text{crit}} \approx \frac{1}{\gamma^{1/2} m}. \quad (58)$$

Note that, aside from the order unity factor $\gamma^{1/2}$, β_{crit} can be interpreted as an approximate equality of the cooling timescale t_c and $1/(m\Omega)$, the time for a fluid element to cross through one wavelength of the m -fold perturbation (in the inner disk).

When $\beta = \beta_{\text{crit}}$, the damping coefficient ξ in Equation (52) reaches

$$\xi_{\text{max}} = \frac{(\gamma - 1)(m^2 - 1)^{1/2}}{2\gamma^{3/4} h}, \quad (59)$$

and the shortest possible damping length is

$$l_{\text{d,min}}^m \approx \frac{r}{\xi_{\text{max}}} = \frac{2\gamma^{3/4} H}{(\gamma - 1)(m^2 - 1)^{1/2}}. \quad (60)$$

Note that $l_{\text{d,min}}^m \propto m^{-1}$ for large m , just as observed in Figure 2. For planet-driven waves in a disk with aspect ratio h_p , the dominant mode has $m \approx m_* \approx (2h_p)^{-1}$ (Ogilvie & Lubow 2002). For this mode, $\beta_{\text{crit}} \approx 2h_p/\gamma^{1/2}$. For $\gamma = 7/5$ and $h = 0.1$, the minimum damping length for the m_* mode is $l_{\text{d,min}} \approx 1.3H_p$, i.e., very short.

Using Equation (57), we also find that the asymptotic behavior for large values of β is

$$\left(\frac{l_F^m}{r} \right)_{\beta \gg \beta_{\text{crit}}} \approx \frac{\gamma^{3/2} h \beta}{\gamma - 1}, \quad (61)$$

which is independent of the value of m . This behavior is in perfect agreement with the results shown in Figure 2.

Additionally, l_F^m becomes negative for $\beta \lesssim \beta_{\pm}$, where

$$\beta_{\pm} \approx \frac{h}{(\gamma - 1)m^2}. \quad (62)$$

For planet-driven waves with $m = m_*$ carrying most of the angular momentum,

$$\beta_{\pm}(m_*) \approx \frac{4}{\gamma - 1} h_p^3, \quad (63)$$

which is about $10h_p^3$ for $\gamma = 7/5$.

This value of the cooling rate has a very important meaning: it sets an *upper limit* on the value of β below which the AMF behavior begins to converge to the locally isothermal regime. One can see from Figure 2 that β has to be at least an order of magnitude smaller than the value $\beta_{\pm}(m_*)$, at which l_F^m changes sign, for l_F^m/r to finally converge to -1 , which signals the ultimate transition to the locally isothermal behavior. Thus, it is natural to expect that the locally isothermal behavior can be reproduced in disks with cooling only when $\beta \lesssim 0.1\beta_{\pm}(m_*) \sim h_p^3$. This point is further discussed in Section 6.5.

The case of a smaller value for the adiabatic index, $\gamma = 1.1$ (chosen only to explore the dependence on γ), is shown in Figure 2(b). We see that the behavior of F_J^m is qualitatively very similar to the case with $\gamma = 1.4$. In particular, there are regions of strong wave damping for β in the range $10^{-2} \lesssim \beta \lesssim 1$. In agreement with Equation (60), the damping is weaker, i.e., l_F^m/r is larger by a factor of a few as compared to the case with $\gamma = 1.4$, but nonetheless $l_F^m/r \ll 1$ for these values of β . In this case, the locally isothermal limit is reached for $\beta \lesssim 3 \times 10^{-3}$ —larger than the limiting value for $\gamma = 1.4$, in agreement with Equation (62).

Finally, for γ very close to unity, $\gamma = 1.01$ (see Figure 2(c)), the damping is very weak for all values of β . The behavior of F_J^m transitions almost monotonically from the locally isothermal limit to the adiabatic limit in this case. Note that for this γ , the locally isothermal limit is valid for $\beta \lesssim 0.03$. This is much larger than the limiting value for $\gamma = 1.4$, again due to the $(\gamma - 1)^{-1}$ dependence of β_{\pm} (Equation (62)), but it is still $\ll 1$.

The preceding analysis has focused on the case of a $q = 1$ temperature profile. For $q \neq 1$, for which l_F^m/r varies with r , the behavior is qualitatively similar to the $q = 1$ case shown in Figure 2 but quantitatively modified in two key ways (as described by Equations (52) and (56)). First, since the minimum damping length $l_{d,\min}$ is proportional to the local scale height $H(r)$ (see Equation (60), which is valid for general values of q), $l_{d,\min}/r$ is proportional to the local aspect ratio $h(r)$. This is independent of r for $q = 1$, but $\propto r^{(1-q)/2}$ in general. Second, for small values of β , in the nearly locally isothermal limit, in general, $l_F^m/r \rightarrow -1/q$ for all values of m . The limiting value of β at which this transition occurs is still given approximately by Equation (62), but modified by a factor of q . This is a relatively minor effect, since q lies within a fairly narrow range of values (i.e., varying by tens of percent) for most physically reasonable disk models.

4.4. Numerical Results

Profiles of the Fourier AMF F_J^m for planet-excited waves are shown in Figure 3, for disks with varied thermodynamics: a locally isothermal disk, and disks with cooling characterized by different values of β ranging from 10^{-4} to 10^2 . The AMF is

expressed in terms of the characteristic scale

$$F_{J,0} = \left(\frac{M_p}{M_*} \right)^2 h_p^{-3} \Sigma_p r_p^4 \Omega_p^2, \quad (64)$$

associated with the total angular momentum transfer at Lindblad resonances (Goldreich & Tremaine 1980; Ward 1997). In each panel, the results of numerical simulations (Section 3.3) for $M_p \ll M_{\text{th}}$, when the perturbation is well in the linear regime, are shown as solid lines, with different colors representing different azimuthal mode numbers. The dashed lines show the results of linear theory obtained by numerically solving the master Equation (9) with the isothermal (Equations (17)–(21)) or cooling (Equations (24)–(27)) inputs and fully accounting for the forcing by the planetary potential (see Section 3.2). Clearly, there is excellent agreement between the theoretical and numerical results, again validating our theoretical analysis. Note that the linear solutions self-consistently capture the behavior of F_J^m in the vicinity of the planet where wave excitation occurs. Beyond this region, the WKB approximation (see Section 4.3) provides an excellent description of the evolution of F_J^m , as we demonstrate in Appendix C.2. Profiles of the total AMF (resulting from the sum of all Fourier modes) are shown in Figure 4(a) (for the same cases as shown in Figure 3). Here, the solid and dashed lines again represent the results of numerical simulations and linear calculations, respectively, which show close agreement with one another.

We first examine the AMF in the inner disk. The AMF behavior for a locally isothermal disk (Figure 3(a)), for which $F_J^m \propto r^{-1}$ (for $q = 1$) for all m , is reproduced very closely for the case $\beta = 10^{-4}$ (Figure 3(b)). The case with $\beta = 10^{-3}$ (Figure 3(c)) is qualitatively very similar, although there are slight differences; the different F_J^m components and, hence, the total F_J (see Figure 4(a)), do not increase toward the inner disk quite as steeply.

For $\beta = 10^{-2}$ (Figure 3(d)), the results are substantially different. In this case, the damping rates of the different F_J^m are very different from one another. For modes with $m \lesssim 5$, F_J^m increases toward the inner disk, while for $m \gtrsim 5$, it decreases as a result of cooling-related damping. As a result, the total AMF (see Figure 4(a)) is approximately constant in the inner disk. The approximate behavior of the total AMF can be qualitatively deduced from the behavior of $F_J^{m_*}$, where $m_* = 5$ (for $h_p = 0.1$) is the dominant mode for planet-excited waves. For $\beta = 10^{-1}$ (Figure 3(e)), all of the F_J^m decrease toward the inner disk ($l_F^m > 0$), at a rate that increases with m . The resulting total AMF therefore decreases rapidly toward the inner disk.

For $\beta = 1$ (Figure 3(f)), the modes all have similar damping rates. For $\beta = 10$ (Figure 3(g)), all of the F_J^m behave essentially the same, exhibiting only a weak decay toward the inner disk. Finally, for $\beta = 100$ (Figure 3(h)), the AMF of all of the modes is approximately constant and, hence, so is the total AMF, effectively reproducing the behavior expected in a purely adiabatic disk.

In the outer disk, the behavior of the AMF is somewhat simpler. This is because the effects of cooling can only act to make F_J^m decrease with r in the outer disk (it would decrease even in the locally isothermal case). This is a result of either the AMF following or nearly following the global temperature gradient, e.g., for $\beta = 10^{-4}$ – 10^{-3} , or due to linear damping,

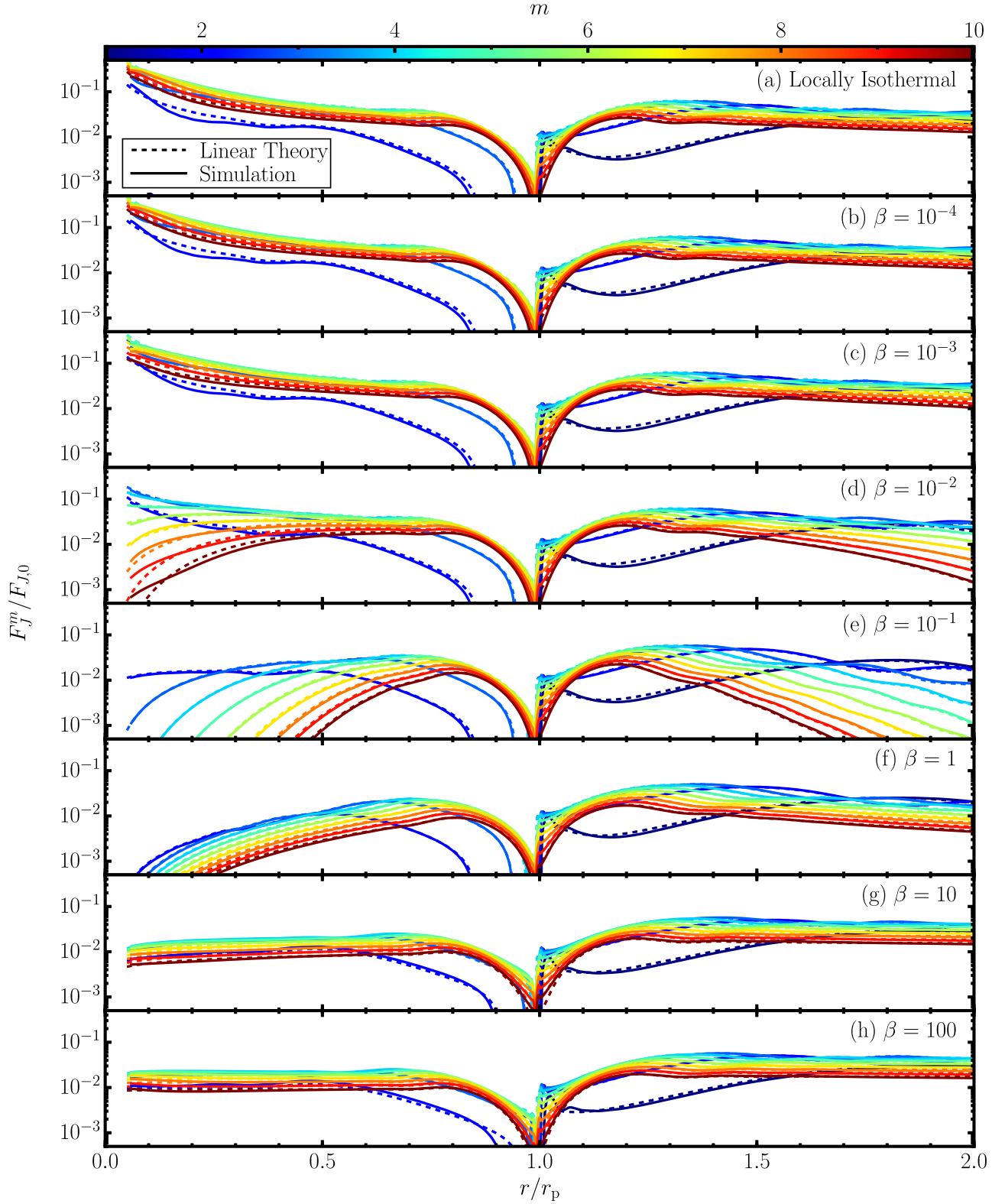


Figure 3. Profiles of the Fourier components of the wave AMF F_J^m , in terms of the characteristic scale $F_{J,0}$ (Equation (64)), for a disk with adiabatic index $\gamma = 7/5$, aspect ratio $h_p = 0.1$, temperature power-law index $q = 1$, and different thermodynamics: locally isothermal and cooling with different values of β . Solid lines of different colors, corresponding to different azimuthal numbers m , are the results of numerical simulations with a $0.01M_{\text{th}}$ planet at 20 orbits. The corresponding dashed lines are the F_J^m components for planet-excited waves computed from linear theory.

e.g., for $\beta = 10^{-2}$ –10. For intermediate values of β , the different modes have different damping rates in the outer disk, although the differences are less pronounced than in the inner disk.

Taken together, the results for the AMF illustrate the fact that the cooling time must be extremely short relative to the orbital timescale, with $\beta \lesssim 10^{-3}$, for the locally isothermal approximation to provide an accurate description of density

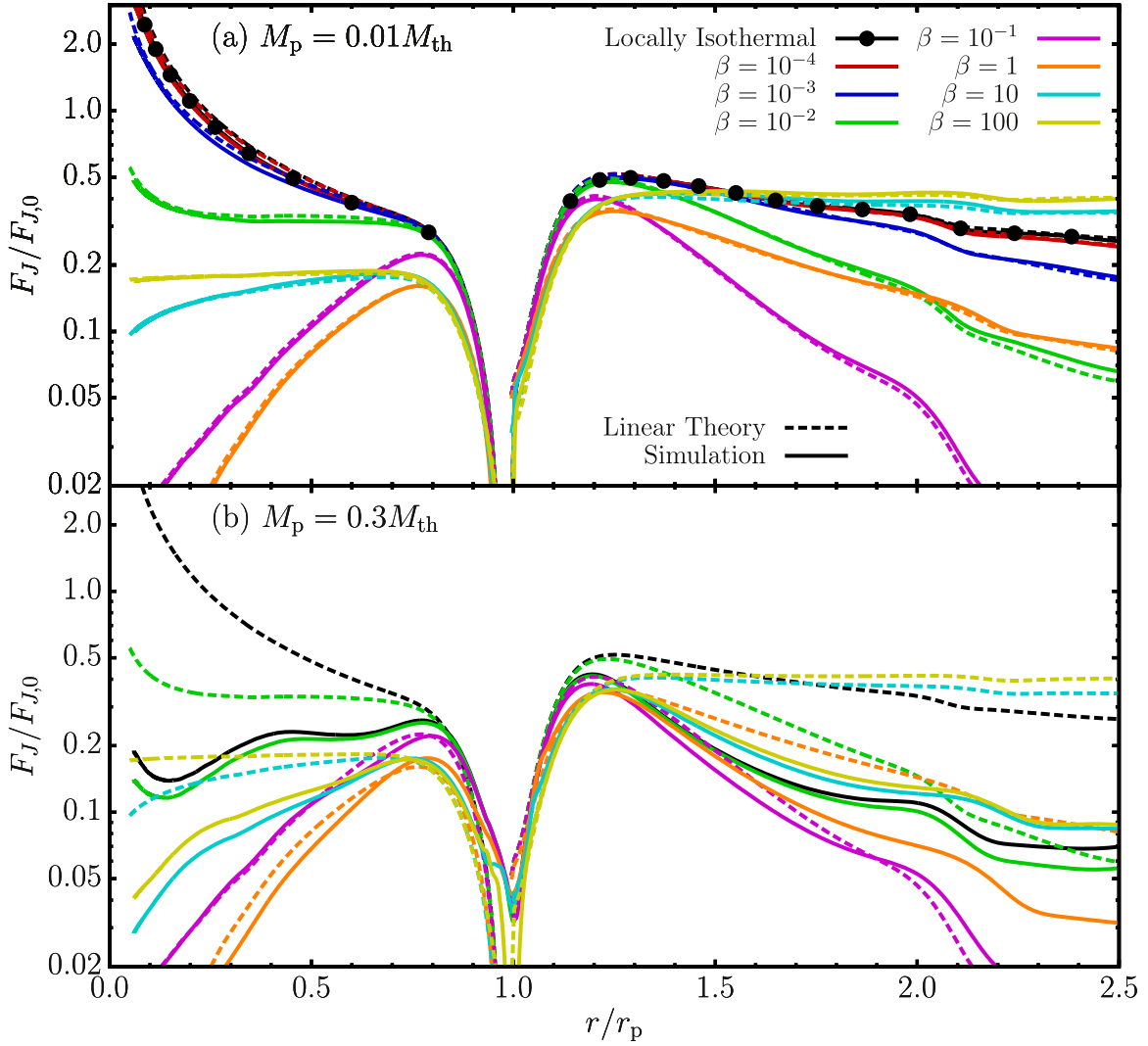


Figure 4. The total AMF F_J for planet-driven waves in disks with different thermodynamics (the same cases as in Figure 3). The dashed lines, which are the same in both panels, are computed using linear theory. The solid lines are the results of numerical simulations, with a $0.01 M_{th}$ planet in (a), and a $0.3 M_{th}$ planet in (b), taken at 20 orbits. The filled points in (a) highlight the curve for the locally isothermal simulation, in order to help distinguish it from the other similar curves for small values of β .

waves in the linear regime. The AMF behavior for purely adiabatic disks is reproduced for $\beta \gtrsim 10$. There is, therefore, a very wide range of cooling times, with β spanning about four orders of magnitude, for which neither the adiabatic nor locally isothermal approximations provide a good description of the density wave dynamics.

4.4.1. Torque

Our analysis has so far been primarily concerned with the radial variation of the wave AMF due to cooling that occurs outside the wave excitation zone. However, as seen in Figure 4, the “initial amplitude” of F_J (i.e., its value a few scale heights away from the planet) varies with the cooling timescale. This is indicative of a variation of the torque exerted on the disk with β . As this torque is eventually responsible for planet migration, we would like to quantify how it is affected by cooling. The net torque on the disk is

$$T_{\text{tot}} = \int_{r_{\text{in}}}^{r_{\text{out}}} \frac{dT}{dr} dr, \quad (65)$$

where dT/dr (Equation (39)) is the torque density. The orbital evolution of the planet is described by $\dot{L}_p = -T_{\text{tot}}$, where $L_p = M_p(GM_*r_p)^{1/2}$ is the angular momentum of the planet.

There are three different contributions to the net torque. These are the one-sided Lindblad torques T_{LR}^{\pm} (where + and – refer to the outer and inner disk), resulting from wave excitation at Lindblad resonances, and the corotation torque T_C , associated with angular momentum transfer at the corotation radius r_C , where $\Omega(r_C) = \Omega_p$ (e.g., Goldreich & Tremaine 1979). These torques are computed using profiles of the total (i.e., summed over m) AMF F_J and torque density dT/dr computed from linear theory. We forgo an analysis of the torques in our numerical simulations, as the corotation torque is subject to oscillations and saturation (e.g., Paardekooper & Papaloizou 2008) on timescales much longer than we have simulated.

In adiabatic disks, the one-sided Lindblad torques are identified as the asymptotic values of F_J at large distances from the planet. This is a consequence of the conservation of AMF for free waves—in adiabatic disks the wave AMF changes only as a result of excitation by the planetary potential,

described by dT/dr , except in the immediate vicinity of the corotation resonance, where it experiences a discontinuous jump. As we have shown, cooling results in additional radial variation of F_J , described by the difference between dF_J/dr and dT/dr . Therefore, the one-sided Lindblad torques can be computed via the following formula:

$$T_{\text{LR}}^{\pm} = \pm F_J(r_{\text{out/in}}) \mp \int_{r_c^{\pm}}^{r_{\text{out/in}}} \left(\frac{dF_J}{dr} - \frac{dT}{dr} \right) dr. \quad (66)$$

The second term in Equation (66) compensates for the decay (or growth) of AMF due to cooling that occurs between just outside the corotation radius and the disk edge. The lower limit of integration in this term is $r_c^{\pm} = r_c \pm \epsilon$, with $0 < \epsilon \lesssim 0.005 r_p$, chosen to cut out the jump in F_J at r_c due to the corotation torque (which occurs over a finite radial distance in our numerical results). For $\beta \rightarrow \infty$, the integrand of the second term in Equation (66) vanishes, and we recover the usual definition of the one-sided Lindblad torque for adiabatic disks.

The total torque due to both the outer and inner Lindblad resonances—the so-called differential Lindblad torque—is then

$$T_{\text{LR}} = T_{\text{LR}}^{+} + T_{\text{LR}}^{-}, \quad (67)$$

and finally the corotation torque is given by

$$T_C = T_{\text{tot}} - T_{\text{LR}}, \quad (68)$$

The results of this decomposition are shown in Figure 5. Results for several additional values of β not shown in Figures 3 and 4 are displayed in order to fully illustrate the behavior of the torques as a function of β . All of the components of the torque shown in Figure 5 exhibit approximately the same dependence on β . Each one is constant for $\beta \lesssim 10^{-1}$, undergoes a moderate decrease as β is increased to ≈ 100 , and is again constant for $\beta \gtrsim 100$. In the locally isothermal limit ($\beta \rightarrow 0$), the torques are 30%–35% larger than in the adiabatic limit ($\beta \rightarrow \infty$). This is related to the variation of the effective sound speed $c_{s,\text{eff}}$ (see Equation (51)) from $c_{s,\text{iso}}$ to $c_{s,\text{adi}}$ as β is increased. The smaller $c_{s,\text{eff}}$ for small values of β results in a stronger response of the disk to the planetary potential and, therefore, a larger torque. Note that the total torque on the disk (Figure 5(d)) is always positive, resulting in inward migration of the planet.

Analytic studies of wave excitation at Lindblad resonances (Goldreich & Tremaine 1980; Ward 1997) find the characteristic scale of the one-sided Lindblad torques T_{LR}^{\pm} to be $F_{J,0}$ (Equation (64)), while the characteristic scale of the differential Lindblad torque T_{LR} should be smaller by a factor h_p . Hence, $T_{\text{LR}}^{\pm} \propto c_s^{-3}$ and $T_{\text{LR}} \propto c_s^{-2}$. However, in practice, T_{LR}^{\pm} and T_{LR} are found to be similar in magnitude when $h_p \approx 0.1$ (Ward 1997; Papaloizou et al. 2007) as we have considered in this work. As such, all of the torques in Figure 5 are shown in terms of the differential Lindblad torque scale $h_p F_{J,0}$.

The aforementioned scalings suggest that $T_{\text{LR}}^{\pm}(\beta \rightarrow 0)/T_{\text{LR}}^{\pm}(\beta \rightarrow \infty) \approx \gamma^{3/2}$ and $T_{\text{LR}}(\beta \rightarrow 0)/T_{\text{LR}}(\beta \rightarrow \infty) \approx \gamma$. In our calculations, T_{LR} obeys the expected scaling, whereas T_{LR}^{\pm} exhibits a somewhat weaker variation with β than expected. This may be related to the relatively large value of h_p used, or the softening of the planetary gravitational potential in our calculations, which is not included in many analytic studies. We note that our value for T_{LR} in the adiabatic limit is in good agreement (within a few percent) with the adiabatic Lindblad torque formula given by Paardekooper et al. (2010), which is

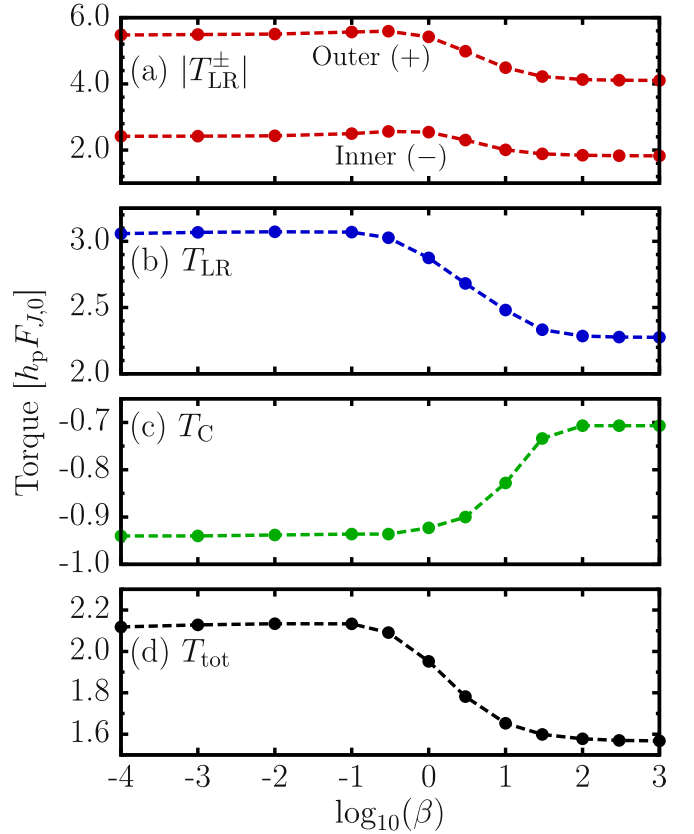


Figure 5. Different components of the torque exerted by the planet on the disk, computed using linear theory, as a function of $\beta = \Omega t_c$: (a) magnitude of the one-sided Lindblad torques T_{LR}^{\pm} (note that $T_{\text{LR}}^{-} < 0$), (b) differential Lindblad torque T_{LR} , (c) corotation torque T_C , and (d) total torque T_{tot} . See Section 4.4.1 for details. The disk has an aspect ratio $h_p = 0.1$, temperature and surface density power-law indices $q = 1$ and $p = 1$, respectively, and adiabatic index $\gamma = 7/5$.

based on numerical solutions of the linear perturbation equations (including softening).

We conclude that cooling has only a modest effect (tens of a percent) on the magnitude of the linear torque associated with planet migration for the disk model we have considered. However, more dramatic effects associated with cooling might be possible. In particular, the corotation torque, which is prone to becoming nonlinear even for low-mass planets, is sensitive to the entropy gradient of the disk, and its behavior has previously been shown to be sensitive to cooling. Reversal of the direction of migration is possible when these effects are taken into account (Paardekooper & Papaloizou 2008). A thorough exploration of this topic, requiring an exploration of different disk profiles and a full consideration of nonlinear effects, is beyond the scope of this work.

5. Massive Planets and Disk Evolution

Simulations shown in the previous section explored the linear regime of the planet–disk coupling. We now explore numerically the role of cooling on the planet–disk interaction in the presence of the nonlinear effects. It is well known that density waves launched by massive planets in adiabatic disks undergo rapid nonlinear evolution resulting in their shocking and dissipation (Goodman & Rafikov 2001; Rafikov 2002a; Dong et al. 2011a). Our goal will be to explore the interplay between the wave evolution due to cooling (which is a linear

phenomenon) and due to the nonlinear effects. Another goal is to examine the effect of the different levels of cooling on the long-term evolution of the disk—the variation of its surface density caused by the deposition of the angular momentum of the density wake.

We consider a planet mass that is a moderate fraction of the thermal mass, $M_p = 0.3M_{\text{th}}$. The setup for these simulations is the same as described in Section 3.3, except that we reduce the numerical resolution by a factor of two, to $N_r \times N_\phi = 1502 \times 2048$, in order to facilitate a longer evolution timescale of $500t_p$. This is long enough for the surface density profile of the disk to undergo significant evolution due to the planet-driven waves. We make use of profiles of the wave AMF as a diagnostic of the wave-driven evolution of the disk (e.g., Miranda & Rafikov 2019b). As before, we consider both locally isothermal disks and disks with cooling with several values of β , which are representative of the variety of different behaviors of F_J in the linear regime studied before (Section 4.4).

5.1. AMF Profiles

Figure 4(b) shows the profiles of the total AMF at 20 orbits (solid lines), along with the AMF profiles computed using linear theory (dashed lines). For the locally isothermal case, F_J begins to deviate from the linear prediction at about one or two scale heights from the planet, which is consistent with the theoretical shocking distance $l_{\text{sh}} \approx (M_p/M_{\text{th}})^{-2/5} \approx 1.6H_p$ (Goodman & Rafikov 2001). Beyond this distance, the actual F_J is always smaller than the linear F_J , as a result of nonlinear dissipation. In most other cases (i.e., for different values of β), F_J also begins to drop below the linear prediction at a distance $\approx l_{\text{sh}}$ from the planet.

Two particular cases shown in Figure 4(b), locally isothermal and $\beta = 10^{-2}$, have very similar AMF profiles, despite the fact that the corresponding linear AMF profiles are quite different. Given that the linear AMF for the locally isothermal case is always larger than it is for the case with $\beta = 10^{-2}$, this indicates that in the locally isothermal case, the waves experience more nonlinear dissipation, resulting in the actual F_J being comparable between these two cases almost everywhere in the disk. This similarity is probably coincidental, resulting from the particular strength of the nonlinear dissipation for our chosen planet mass and would not occur for a different mass. Note, however, that, regardless of these details, the fact that the AMF profile is nearly the same for these two cases indicates that the resulting wave-driven disk evolution should also be nearly the same.

The most important feature of Figure 4(b) is the fact that for the case with $\beta = 10^{-1}$, the AMF profile is not substantially different from the corresponding linear AMF profile. This is also true for case with $\beta = 1$, although, mostly in the inner disk—there is some deviation in the outer disk. This indicates that nonlinear dissipation plays a much smaller role in the evolution of the density waves in these cases, i.e., for values of β in this range. Instead, evolution of the wave AMF is mostly controlled by the linear damping associated with cooling. Evidently, this damping is so strong that it is the dominant source of dissipation even for a fairly massive planet such as the one considered here ($0.3M_{\text{th}}$).

The fact that nonlinear dissipation is subdominant to linear dissipation for $\beta \approx 10^{-1}$ –1 suggests that wave-driven evolution of the disk operates differently for disks with dimensionless cooling timescales in this range. We may therefore expect

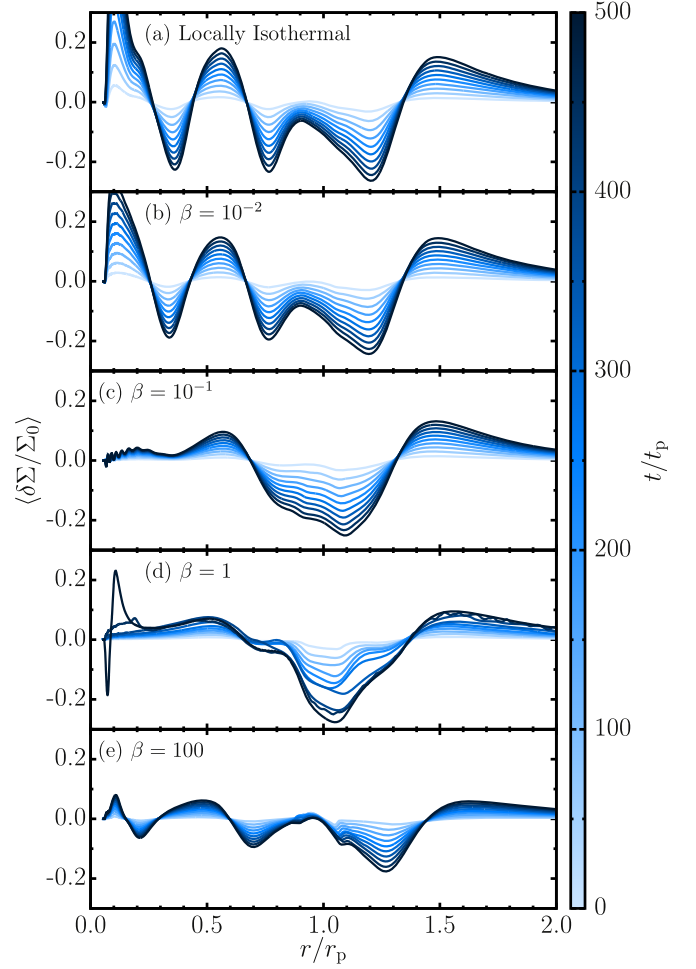


Figure 6. Evolution of the disk surface density for $M_p = 0.3M_{\text{th}}$ in disks with different thermodynamics. The azimuthally averaged surface density perturbation $\delta \Sigma$, relative to the initial surface density profile Σ_0 , is shown every 50 orbits for 500 orbits.

one of two different types of disk evolution, one associated with nonlinear wave dissipation (for very long or very short cooling timescales), and one dominated by linear wave dissipation (in the aforementioned range of β). The exact separation between these two regimes in terms of the cooling timescale should be a function of M_p/M_{th} .

5.2. Surface Density Evolution

Figure 6 shows the evolution of the disk surface density profile for the different β cases shown in Figure 4(b). There is a distinct dichotomy in the resultant disk structures. For sufficiently short and sufficiently long cooling times (Figures 6(a), (b), and (e)), for which wave dissipation happens primarily through shocks, the disk exhibits multiple gap and ring structures. For these cases, three gaps (surface density minima) and four rings (surface density maxima) are formed. Note that the surface density perturbation in the case of $\beta = 10^{-2}$ is almost the same as in the isothermal case, which is to be expected based on Figure 4(b) and the discussion in the previous section.

For the intermediate cooling times, for which linear damping is more important than nonlinear dissipation ($\beta = 10^{-1}$ and $\beta = 1$; Figures 6(c)–(d)), the resulting structure is dominated by a single wide gap around the orbit of the planet. In these cases, the gap extends from about $0.5r_p$ to about $1.5r_p$, so that

its fractional width $\Delta r/r$ is ≈ 1 . Some hints of additional structure are present, but the multiple ring/gap structure is highly suppressed.

For the cases that develop multiple ring/gap structures, the features are more pronounced, i.e., the gaps are deeper, for short cooling timescales as compared to long cooling timescales. This is a consequence of the fact that for short cooling timescales (i.e., in locally isothermal or approximately locally isothermal disks), waves in the inner disk gain AMF from the background disk flow at larger radii and carry it to small radii, where it is returned to the disk through the nonlinear dissipation. This results in stronger features, as compared to disks with long cooling timescales, i.e., effectively adiabatic disks, for which waves do not gain AMF from the disk as they propagate but can only dissipate their initial AMF. This difference in behaviors has been explored in Miranda & Rafikov (2019b).

6. Discussion

6.1. Implications for Protoplanetary Disks

In protoplanetary disks, thermal relaxation is mediated by a combination of (i) radiative cooling via thermal dust emission from the surface of the disk and (ii) radiative diffusion along the plane of the disk. In general, the full energy evolution equation describing one or both of these processes may be linearized, resulting in a cooling law of the form given by Equation (22). The specific form of the resulting cooling timescale t_c or β depends on the details of the cooling processes under consideration. Note that, in general, β must vary with the distance from the star, unlike the idealized constant β scenario considered in this work. However, our theoretical framework is not limited to the constant β case and can be applied to arbitrary $\beta(r)$ profiles.

In the case of radiative cooling from the surface of the disk, estimates of the cooling timescale $\sim e\Sigma/(\sigma T_{\text{eff}}^4)$ (where T_{eff} is the disk effective temperature) for typical disk parameters lead to values of β ranging from ~ 100 at several astronomical units to $\sim 10^{-2}$ at 100 astronomical units (e.g., Zhu et al. 2015). We have shown (see Section 4.4) that in the linear regime, the locally isothermal approximation is applicable only for $\beta \lesssim 10^{-3}$. For more massive planets, $\beta = 10^{-2}$ is only marginally compatible with the isothermal description (see Section 5.1). Based on this, we conclude that even in the outer regions of protoplanetary disks, radiative cooling is not sufficiently rapid for the locally isothermal approximation to provide an accurate description of wave dynamics.

However, in the optically thick inner disk (radii less than a few tens of astronomical units), radiative diffusion may instead be the dominant source of cooling. In this case, the cooling timescale in Equation (22) should be identified with the diffusion timescale $(\eta k^2)^{-1}$, where η is the radiative diffusion coefficient and k is the radial wavenumber of the perturbation (e.g., Lin & Youdin 2015). Note that in this case, the cooling timescale depends on the perturbation wavelength. In the context of the linear theory presented in this work, this translates to a dependence of the cooling timescale on the azimuthal mode number m . Rather than a universal cooling time, each Fourier harmonic of the internal energy is thermally relaxed on its own cooling timescale $t_{c,m}$.

It can be shown that the radiative diffusion cooling timescale is smaller than the (optically thick) radiative cooling timescale

by a factor $(kH)^2 \sim m^2$ for a density wave with azimuthal number m . For planet-driven waves with a characteristic azimuthal number $\sim h_p^{-1} \sim 10$, this suggests that the value of β associated with radial diffusion may range from ~ 1 at several astronomical units, down to some minimum value, perhaps $\sim 10^{-3}$, occurring at the optically thick/thin transition radius. Therefore, when cooling is mediated by thermal diffusion, the locally isothermal approximation may provide an accurate treatment of wave dynamics in the outermost regions of protoplanetary disks, but not at smaller radii, where a full consideration of the effects of cooling is indispensable. In a forthcoming work (R. Miranda & R. Rafikov 2020, in preparation) we will explore the evolution of density waves using a detailed prescription for the cooling timescale, which accounts for both radiative cooling and diffusion, as appropriate for realistic protoplanetary disks.

6.2. Implications for Multiple Spiral Arms

In Section 3.4, we pointed out several modifications of the multiple spiral arm structure of planet-driven density waves in disks with cooling. We now discuss the reasons behind these modifications in the context of AMF conservation.

As we noted, the basic fact that multiple spirals are formed in the inner disk is not modified by cooling. Multiple spiral arms are formed as a consequence of the mode interference governed by the dispersion relation for spiral density waves. This dispersion relation is not strongly modified by cooling (see Equation (50)), with one key exception: the effective sound speed for density waves varies from $c_{s,\text{iso}}$ for small values of β to $c_{s,\text{adi}}$ for large values of β . As a result, the pitch angle of the spirals gets effectively larger by a factor $\gamma^{1/2}$ (i.e., 18% larger for $\gamma = 7/5$) for adiabatic disks as compared to locally isothermal disks, see Figures 1(a) and (e).

In regards to this fact, note that in Miranda & Rafikov (2019a), we pointed out that the spiral structure is nearly independent of the value of γ for adiabatic disks. This statement applies when the comparison is made between disks that have the same adiabatic sound speed profile, i.e., for which $\gamma T(r)$ is the same. If, instead, we consider disks with different values of γ but with the same temperature profile $T(r)$ (i.e., the same isothermal sound speed profile), then the spiral structure in the linear regime does in fact depend on the value of γ . Specifically, different values of γ lead to different adiabatic sound speeds (proportional to $\gamma^{1/2}$). Correspondingly, the pitch angles of the spirals are larger, and the evolution of the single spiral into multiple spirals proceeds more slowly with distance from the planet, for larger values of γ (and the same temperature profile). See Section 6.3 for a further discussion of this issue.

In Miranda & Rafikov (2019a), we performed a detailed characterization of the spiral structure for adiabatic disks. However, with a minor modification, it can also be applied to locally isothermal disks. Specifically, the quantity $\delta\Sigma_{\text{lin}}$, which characterizes the overall scaling of the wave amplitude as a function of r as dictated by AMF conservation (see Equation (16) of Miranda & Rafikov 2019a), should be modified to account for the fact that F_J is not constant in this case but is proportional to c_s^2 . Otherwise, the details of the emergence of the different spirals at different radii, their relative amplitudes and separations, and so on, are the same for both adiabatic and locally isothermal disks. This can be traced to the fact that the AMF for different wave modes F_J^n all obey the same radial

scaling in both cases: they are all constant in adiabatic disks and all proportional to c_s^2 in locally isothermal disks.

The situation is rather different in cooling disks with $\beta \approx 10^{-2}$ –1. In such disks, the radial scaling of F_j^m and, hence, of the amplitudes of the harmonics of the surface density perturbation $\delta\Sigma_m$, varies with m (see Figures 2 and 3). As a result, the structure of the spirals is different than in either the short or long cooling timescale regimes. Therefore, the secondary spiral, tertiary spiral, and so on, form at different locations in the inner disk, and their relative amplitudes and widths evolve differently in these cases. For example, according to Figure 2(a), in a disk with $\beta = 0.1$ all of the modes except for $m = 2$ have small and positive l_F^m/r , implying rapid decay of these modes toward the inner disk. On the contrary, $m = 2$ mode has large l_F^m/r in a disk with this value of β . As a result, only the $m = 2$ mode is present in the inner disk—note the two broad, well azimuthally separated arms, clearly visible in Figure 1(c) (see also the discussion in the end of Section 3.4). We leave further quantitative analysis of the multiple spiral structure for disks with cooling to future work.

As a result of the modification of the spiral structure due to cooling in the linear regime, the subsequent nonlinear evolution of the spirals may also be modified. For example, if cooling reduces the amplitude of a spiral arm or broadens it, then it will develop into a shock at a larger distance from the planet (or potentially not at all). This results in a shift of the axisymmetric gap associated with the spiral developing into a shock. The modified shape of the spiral arm due to cooling changes the amount of AMF it carries and can ultimately deposit into the disk, therefore also affecting the depth of the resultant gap.

6.3. Isothermal Versus Adiabatic Sound Speed

It is important to distinguish between the different sound speeds relevant to wave dynamics in disks with different thermodynamics. These are the isothermal and adiabatic sound speeds:

$$c_{s,\text{iso}} = \left(\frac{k_B T}{\mu} \right)^{1/2}, \quad c_{s,\text{adi}} = \left(\frac{\gamma k_B T}{\mu} \right)^{1/2}, \quad (69)$$

where μ is the mean molecular weight. Note that $c_{s,\text{adi}} = \gamma^{1/2} c_{s,\text{iso}}$. In Section 2, we showed that $c_{s,\text{iso}}$ is the appropriate sound speed in locally isothermal disks, while $c_{s,\text{adi}}$ is the appropriate sound speed in adiabatic disks. We also showed that, with a finite cooling timescale, the appropriate sound speed is neither $c_{s,\text{adi}}$ nor $c_{s,\text{iso}}$, but it can be represented by an effective sound speed intermediate between the two (see Equation (51)). This effective sound speed is nonuniversal, with a nontrivial dependence on the cooling timescale, as well as on the location in the disk and the azimuthal number of the perturbation.

In numerical simulations, the disk temperature profile is typically prescribed indirectly through the sound speed, which itself is often parameterized by specifying the disk aspect ratio $h = H/r$, with the disk thickness $H = c_s/\Omega$, where c_s is either the isothermal sound speed or the adiabatic sound speed. In locally isothermal simulations, H must be defined in terms of $c_{s,\text{iso}}$, since this is both the actual propagation speed of sound waves, and the only sound speed that is properly defined in this case. But in simulations with an ideal equation of state (with or

without cooling), we can, in principle, choose to define H in terms of either of the two sound speeds. This ambiguity is resolved by noting that, if H is defined in terms of $c_{s,\text{iso}}$, then disks with the same $h(r)$ profile have the same temperature profile, regardless of other thermodynamic considerations (adiabatic index or cooling timescale). If we regard the temperature as a fundamental physical property of the disk—as opposed to the sound speed, which depends on thermodynamic assumptions—then this is the preferred way of defining H . Parameterizing the disk temperature in terms of $h = c_{s,\text{iso}}/(r\Omega)$, as we have done in this paper, facilitates the most direct comparison of simulations of disks with different thermodynamics—locally isothermal, adiabatic, or cooling.

6.4. Anomalous Mass Flux

As a result of angular momentum conservation in the disk + density wave system, evolution of the wave AMF must result in disk evolution. In particular, the mass flux $\dot{M} = -\oint r \Sigma u_r d\phi$ (here defined to be positive for inflow) is related to the evolution of the wave AMF according to (Rafikov 2002b)

$$\dot{M} = \left(\frac{dl}{dr} \right)^{-1} \left(\frac{dF_j}{dr} - \frac{dT}{dr} \right), \quad (70)$$

where $l = r^2\Omega$ is the specific angular momentum. For simplicity, we have neglected an additional term related to the time evolution of the angular momentum profile of the disk, although in some situations, this term can be important (e.g., Arzamasskiy & Rafikov 2018).

The physical interpretation of Equation (70) is as follows. The evolution of the wave AMF, described by dF_j/dr , occurs for two different reasons. The first is wave excitation—transfer of angular momentum to the wave by an external torque, described by dT/dr . Far from the planet, this term can be neglected. The second is the transfer of angular momentum from the wave to the background disk (e.g., due to damping of the wave), which necessarily leads to disk evolution. Therefore, the disk evolution is determined by the difference between the total variation of the AMF and the external torque density; hence, this difference appears in Equation (70).

Consider planet-driven waves that are subject to dissipation (either linear or nonlinear), so that F_j decreases as waves propagate away from the planet. In this case, Equation (70) indicates that $\dot{M} < 0$ in the outer disk (as $dF_j/dr < 0$ there) and $\dot{M} > 0$ in the inner disk (where $dF_j/dr > 0$). In other words, the effect of wave dissipation is effectively to *repel* mass from the orbit of the planet. In the absence of dissipation, the wave has no effect on the state of the disk (Goldreich & Nicholson 1989).

However, in locally isothermal disks, the wave AMF, described by Equation (45), increases as waves propagate toward the inner disk, and decreases in the outer disk (see Figure 4(a)), provided that the disk temperature decreases with r . As a result, in the locally isothermal case, using Equation (70), we find (setting $dT/dr = 0$ as applicable for free waves)

$$\dot{M} = \frac{2F_j}{r\Omega} \frac{d \ln c_s^2}{d \ln r} = -\frac{2qF_j}{r^2\Omega}, \quad (71)$$

where the second equality applies to disks with $T \propto r^{-q}$. In this case, provided that $q > 0$, $\dot{M} < 0$ everywhere in the disk, with $|\dot{M}| \propto r^{-(q+1/2)}$ because $F_j \propto c_s^2 \propto r^{-q}$. This means that in the

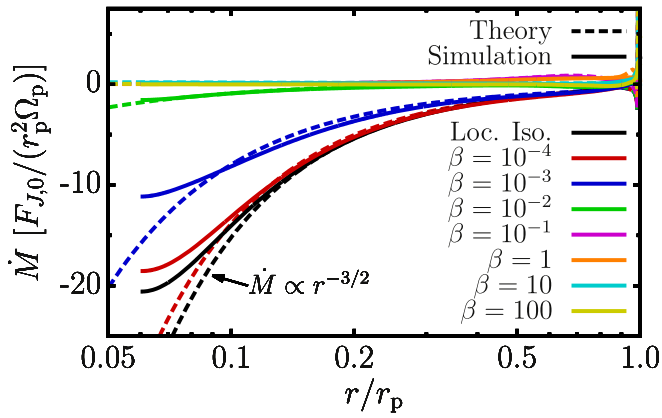


Figure 7. Mass flux \dot{M} due to planet-driven waves in the inner disk for disks with different dimensionless cooling timescales. The dashed lines show the theoretical mass flux (Equation (70)) based on linear theory, and the solid lines show the results of numerical simulations (with $M_p = 0.01M_{th}$). The asymptotic behavior for $r \ll r_p$ in the locally isothermal limit (and for a $q = 1$ temperature profile) is indicated in the lower left.

inner disk, there is a flow of mass *toward* the planet in the linear regime, before the waves shock, in contrast to the conventional expectation. In the outer disk, even before the wave shocks, there is a flow of mass *away* from the planetary orbit. This represents anomalous wave-driven disk evolution that is not associated with wave dissipation in the usual sense (cf., Goldreich & Nicholson 1989).

The case of a disk with cooling is distinct from the adiabatic and locally isothermal cases: even in the absence of nonlinear dissipation, evolution of the wave flux due to cooling (see Figure 4(a)) will give rise to nonzero \dot{M} . However, depending on the value of β , \dot{M} in the inner disk can be either positive (e.g., for intermediate values of $\beta \sim 0.1$) or negative (e.g., for small $\beta \lesssim 10^{-3}$, close to the locally isothermal limit).

The behavior of the anomalous mass flux described above is illustrated in Figure 7. The mass flux in the inner disk from numerical simulations in the linear regime (for $M_p = 0.01M_{th}$, when the wave does not shock) is shown along with the theoretical mass flux (Equation (70)) predicted by linear theory (using F_J computed in linear theory, see Figure 4(a)). The locally isothermal case and cases with cooling and different constant values of β are shown. The numerical and theoretical results show good agreement, except at small radii ($r \lesssim 0.1r_p$).⁵ For $\beta \lesssim 10^{-2}$, as well as in the locally isothermal case, there is a substantial outward flow of gas, i.e., toward the planet ($\dot{M} < 0$). For $\beta = 10^{-1}$ –1, one can see a small positive \dot{M} , as expected from the $F_J(r)$ behavior.

In Figure 7, \dot{M} is expressed in terms of $F_{J,0}/(r_p^2 \Omega_p)$ (see Equation (64)). If we equate this scale to the mass flux for a steady state viscous disk, $3\pi \nu \Sigma$, with ν parameterized using the α -prescription, then the characteristic scale of \dot{M} due to planet-driven waves is equivalent to having

$$\alpha \sim \left(\frac{M_p}{M_{th}} \right)^2 h_p. \quad (72)$$

For $M_p \sim M_{th}$, this characteristic α can substantially exceed the usual “viscous” α .

⁵ Discrepancies at these small radii may be the result of nonlinear evolution of the waves, or boundary effects associated with the presence of the inner wave damping zone.

However, in practice the aforementioned anomalous mass flux (away from the central star) can have a substantial impact on the disk evolution only when the wave amplitude is small enough for waves to travel far into the inner disk without shocking and nonlinear dissipation. This requires planets to be low-mass with M_p below roughly a few percent of M_{th} . In this case, the effective α would be $\lesssim 10^{-4}$. Nonlinear dissipation of the density waves driven by more massive ($M_p \sim M_{th}$) planets can give rise to substantially higher values of effective α (Goodman & Rafikov 2001; Rafikov 2002a).

6.5. Other Implications

It is commonly assumed that the locally isothermal approximation should be valid provided that $\beta = \Omega t_c \lesssim 1$, i.e., the cooling timescale is short compared to the orbital timescale. However, we showed that this expectation is too naive and that, in fact, there are two key timescales to which t_c should be compared in order to assess the effects of cooling on the dynamics of density waves, neither of which are the orbital timescale.

The first relevant timescale is the time for a fluid element to cross through one of the crests/troughs of the perturbation with azimuthal number m , which is $|\tilde{\omega}|^{-1}$. This can be understood by noticing that β enters Equations (24)–(26) only in the combination $\tilde{\beta} = \tilde{\omega} t_c = m\beta(\omega_p - \Omega)/\Omega$. If the wave pattern frequency is slow, e.g., in the inner disk, then $|\tilde{\beta}| \approx m\beta$. When $|\tilde{\beta}| \sim 1$, there is a change in the behavior of the equations. Density waves experience strong linear damping when $|\tilde{\beta}| \sim 1$. Damping is minimal for $|\tilde{\beta}| \gg 1$, which represents the adiabatic limit, in which the wave AMF is constant. Damping is also weak when $|\tilde{\beta}| \ll 1$; however, this constraint, which is already more stringent than $\beta \lesssim 1$, still does not guarantee convergence to the locally isothermal limit, in which the AMF actually rises with the disk temperature.

This constraint is improved upon by considering the second relevant (dimensionless) timescale $\beta_{\pm} \sim h/m^2$ (Equation (62)), for which the radial derivative of the AMF changes sign in the inner disk. Only when $\beta \lesssim \beta_{\pm}$, does the AMF grow as waves travel inwards. This represents a transition toward the locally isothermal regime, in which $F_J \propto r^{-q}$. Based on our numerical results, we require β to be much smaller than β_{\pm} (by at least an order of magnitude) for the AMF behavior to fully converge to the locally isothermal limit. For waves excited by planets, the dominant azimuthal number of perturbations is $m \sim h_p^{-1} \sim 10$ (Goldreich & Tremaine 1980), and so the locally isothermal limit corresponds to $\beta \lesssim h_p^3 \sim 10^{-3}$ in realistic disks (see Equation (63)). As waves become nonlinear, this requirement relaxes to $\beta \lesssim 10^{-2}$, see Section 5. For other types of density waves (i.e., not driven by planets), the applicability of the locally isothermal approximation depends on their characteristic azimuthal wavenumber.

As a result of the stringent constraints on the cooling timescale necessary for the locally isothermal approximation to provide a good description of wave dynamics, its use is likely to be unjustified in many studies involving wave dynamics in disks. This is especially relevant for situations involving the long-range propagation of planet-excited density waves. These include modifications to disk-driven migration due to communication between planets via density waves (Podlowska-Gaca et al. 2012), and due to wave reflection at disk edges (Tsang 2011; Miranda & Lai 2018). Under more realistic

thermodynamics, the range over which the density waves can propagate may be limited, reducing the efficacy of these mechanisms. In studies of circumbinary accretion (e.g., Moody et al. 2019; Muñoz et al. 2019), use of the locally isothermal approximation may lead to a misestimation of the size of the circumbinary cavity and of the circumstellar disks (or “mini-disks”) around the individual stars. This could affect the balance of gravitational torques exerted on the binary by circumbinary and circumstellar disks, artificially influencing the resulting orbital evolution of the binary.

A number of other problems critically dependent on globally propagating density waves have been studied using locally isothermal simulations. These include the origin of multiple narrow rings and gaps (Bae et al. 2017; Dong et al. 2017, 2018; Nazari et al. 2019; Weber et al. 2019), similar to the substructures seen in submillimeter observations of protoplanetary disks (ALMA Partnership et al. 2015; Andrews et al. 2016, 2018; Isella et al. 2016; Loomis et al. 2017; Long et al. 2018), Papaloizou–Pringle instability (Barker & Ogilvie 2016), excitation of disk eccentricity in binaries (Regály et al. 2011), and Lidov–Kozai oscillations in disks (Martin et al. 2014), among many others. It may be useful to reconsider some aspects of these problems in terms of AMF conservation as we do in our work. The impact of the locally isothermal EoS on the results of these studies, as well as any modifications of the results that may arise under the consideration of realistic thermodynamics should be identified.

The analysis presented in this work is 2D. We, therefore, expect its validity to depend on the extent to which the structure of planet-excited waves is confined to the midplane of the disk. Irradiated disks may possess an increasing vertical temperature gradient. In the presence of such a temperature gradient, a portion of the angular momentum in planet-excited waves is carried by buoyancy waves, which are channeled toward the upper layers of the disk, possibly leading to enhanced nonlinear dissipation (Lee & Gu 2015). It is unclear whether or not such an effect would be subdominant to the dissipation associated with cooling described in this work. However, our 2D analysis with cooling may set a lower bound to the amount of dissipation expected in 3D.

7. Summary

In this work, we explored the properties of density waves in disks with varied thermodynamics, focusing on the effects of disk cooling on the wave propagation and damping. We used both linear theory and numerical simulations and used the behavior of the AMF of the waves as a diagnostic of wave-driven disk evolution. Our main results can be summarized as follows.

1. While the AMF of free waves F_J is strictly conserved in adiabatic disks (in the absence of linear or nonlinear dissipation), it varies in locally isothermal disks as $F_J \propto c_s^2$ (or the disk temperature T).
2. In disks with a more general thermodynamics, in which the temperature is cooled/relaxed toward an equilibrium profile on a characteristic timescale t_c , the adiabatic and locally isothermal limits are recovered when the cooling timescale is very long and very short, respectively.
3. However, for the locally isothermal approximation to provide a good description of wave dynamics in the linear regime, the cooling timescale must be very short, with

$\beta = \Omega t_c \lesssim h_p^3 \approx 10^{-3}$. This constraint relaxes to $\beta \lesssim 10^{-2}$ for more massive planets because of the nonlinear wave damping.

4. The adiabatic limit (i.e., conservation of F_J in the linear regime) becomes applicable in disks with cooling for $\beta \gtrsim 10$.
5. The transition between the two limiting regimes is non-monotonic and highly nontrivial. For intermediate cooling times, $\beta = 10^{-2}$ –1, the wave AMF rapidly decays due to linear damping.
6. Differences in the decay rates of different Fourier modes of the wave significantly modify the appearance of the planet-driven spiral structure in the inner disk for intermediate cooling times.
7. Non-conservation of the wave AMF in locally isothermal and rapidly cooling disks gives rise to anomalous mass flux driving disk evolution even in the absence of viscosity or nonlinear dissipation.
8. In idealized disks with a radially constant dimensionless cooling time β , the structure of the gaps and rings carved in the disk by a moderately massive planet is strongly affected by the value of β . When β is small ($\lesssim 10^{-1}$) or large ($\gtrsim 1$), multiple narrow rings and gaps (one ring/gap pair exterior to the orbit of the planet and several more interior to the orbit) are formed. For intermediate values of β , a single wide gap centered on the orbit of the planet is formed instead.

Our results should provide guidance for future efforts to better understand the appearance and evolution of density waves and to interpret observations of protoplanetary disks.

We are grateful to an anonymous referee for comments that helped improve the clarity of our presentation. Financial support for this work was provided by NASA via grant 15-XRP15-2-0139.

Appendix A Perturbation Equations

A.1. Adiabatic

For adiabatic disks, Equations (2)–(4) become

$$-i\tilde{\omega} \frac{\Sigma}{c_{s,adi}^2} \delta h + \frac{\Sigma}{L_S} \delta u_r + \frac{1}{r} \frac{\partial}{\partial r} (r \Sigma \delta u_r) + \frac{im\Sigma}{r} \delta u_\phi = 0, \quad (73)$$

$$-i \left(\frac{\tilde{\omega}^2 - N_r^2}{\tilde{\omega}} \right) \delta u_r - 2\Omega \delta u_\phi = -\frac{\partial}{\partial r} (\delta h + \Phi_m) + \frac{\delta h}{L_S}, \quad (74)$$

$$-i\tilde{\omega} \delta u_\phi + \frac{\kappa^2}{2\Omega} \delta u_r = -\frac{im}{r} (\delta h + \Phi_m). \quad (75)$$

A.2. Locally Isothermal

For locally isothermal disks, Equations (2) and (3) become

$$-i\tilde{\omega} \frac{\Sigma}{c_{s,iso}^2} \delta h + \frac{1}{r} \frac{\partial}{\partial r} (r \Sigma \delta u_r) + \frac{im\Sigma}{r} \delta u_\phi = 0, \quad (76)$$

$$-i\tilde{\omega} \delta u_r - 2\Omega \delta u_\phi = -\frac{\partial}{\partial r} (\delta h + \Phi_m) + \frac{\delta h}{L_T}, \quad (77)$$

and Equation (4) is the same as for adiabatic disks (75).

A.3. Disks with Cooling

For disks with cooling, Equations (2) and (3) become

$$\left(1 + \frac{i}{\gamma\tilde{\omega}t_c}\right)^{-1} \left[\left(\frac{1}{t_c} - i\tilde{\omega}\right) \frac{\Sigma}{c_{s,\text{adi}}^2} \delta h + \frac{\Sigma}{L_S} \delta u_r \right] + \frac{1}{r} \frac{\partial}{\partial r} (r \Sigma \delta u_r) + \frac{im\Sigma}{r} \delta u_\phi = 0, \quad (78)$$

$$-i \left(\frac{\kappa^2 - D_c}{\tilde{\omega}} \right) \delta u_r - 2\Omega \delta u_\phi = -\frac{\partial}{\partial r} (\delta h + \Phi_m) + \left(\frac{L_T^{-1} - i\gamma\tilde{\omega}t_c L_S^{-1}}{1 - i\gamma\tilde{\omega}t_c} \right) \delta h, \quad (79)$$

and again Equation (4) is the same as for adiabatic disks (75).

Appendix B Velocity Perturbations

B.1. Adiabatic

For adiabatic disks, the velocity perturbations are given in terms of δh by

$$\delta u_r = \frac{i}{D_S} \left[\left(\tilde{\omega} \frac{\partial}{\partial r} - \frac{2m\Omega}{r} \right) (\delta h + \Phi_m) - \frac{\tilde{\omega}}{L_S} \delta h \right], \quad (80)$$

$$\delta u_\phi = \frac{1}{D_S} \left[\left(\frac{\kappa^2}{2\Omega} \frac{\partial}{\partial r} - \frac{m}{r} \left(\frac{\tilde{\omega}^2 - N_r^2}{\tilde{\omega}} \right) \right) (\delta h + \Phi_m) - \frac{\kappa^2}{2\Omega L_S} \delta h \right]. \quad (81)$$

B.2. Locally Isothermal

For locally isothermal disks, the velocity perturbations are given in terms of δh by

$$\delta u_r = \frac{i}{D} \left[\left(\tilde{\omega} \frac{\partial}{\partial r} - \frac{2m\Omega}{r} \right) (\delta h + \Phi_m) - \frac{\tilde{\omega}}{L_T} \delta h \right], \quad (82)$$

$$\delta u_\phi = \frac{1}{D} \left[\left(\frac{\kappa^2}{2\Omega} \frac{\partial}{\partial r} - \frac{m\tilde{\omega}}{r} \right) (\delta h + \Phi_m) - \frac{\kappa^2}{2\Omega L_T} \delta h \right]. \quad (83)$$

B.3. Disks with Cooling

For disks with cooling, the velocity perturbations are given in terms of δh by

$$\delta u_r = \frac{i}{D_c} \left[\left(\tilde{\omega} \frac{\partial}{\partial r} - \frac{2m\Omega}{r} \right) (\delta h + \Phi_m) - \left(\frac{L_T^{-1} - i\gamma\tilde{\omega}L_S^{-1}}{1 - i\gamma\tilde{\omega}} \right) \tilde{\omega} \delta h \right], \quad (84)$$

$$\delta u_\phi = \frac{1}{D_c} \left[\left(\frac{\kappa^2}{2\Omega} \frac{\partial}{\partial r} - \frac{m}{r} \left(\frac{\kappa^2 - D_c}{\tilde{\omega}} \right) \right) (\delta h + \Phi_m) - \left(\frac{L_T^{-1} - i\gamma\tilde{\omega}L_S^{-1}}{1 - i\gamma\tilde{\omega}} \right) \frac{\kappa^2}{2\Omega} \delta h \right]. \quad (85)$$

Appendix C WKB Analysis for Disks with Cooling

Adopting the WKB ansatz, we write

$$\delta h(r) = A(r) \exp \left[i \int^r k(r') dr' \right], \quad (86)$$

where $k(r)$ is the radial wavenumber, and $A(r)$ is a slowly varying amplitude. The master equation for free waves (Equations (9) and (24)–(26) with $\Phi_m \rightarrow 0$) then reads

$$\frac{A''}{A} + \frac{2ikA'}{A} + ik' - k^2 + C_1 \left(\frac{A'}{A} + ik \right) + C_0 = 0. \quad (87)$$

Assuming $|kr| \gg 1$, to lowest order we have

$$k^2 = - \left(\frac{1 - i\tilde{\beta}}{1 - i\gamma\tilde{\beta}} \right) \frac{\gamma D_c}{c_{s,\text{adi}}^2}. \quad (88)$$

In Equation (88), we have assumed a thin disk, $c_s/(r\Omega) \ll 1$, so that only the last term in Equation (25) for C_0 is retained. We see that, in general, k is complex. We wish to obtain expressions for the real and imaginary parts of k (rather than k^2), as we are interested in the attenuation coefficient $\text{Im}(k)$. In doing so, we use the approximation $D_c \approx D$, since N_r^2 is smaller than D by $\mathcal{O}(h^2)$ (see Equation (27)). We find

$$k = \gamma^{1/2} \left(\frac{1 + \tilde{\beta}^2}{1 + \gamma^2 \tilde{\beta}^2} \right)^{1/4} \frac{|D|^{1/2}}{c_{s,\text{adi}}} \times \exp \left\{ \frac{i}{2} \tan^{-1} \left[\frac{(\gamma - 1)\tilde{\beta}}{1 + \gamma\tilde{\beta}^2} \right] \right\}. \quad (89)$$

It can be shown that the argument of the inverse tangent function inside the exponential, $(\gamma - 1)\tilde{\beta}/(1 + \gamma\tilde{\beta}^2)$, is always small. For example, if $\gamma = 7/5$, its maximum possible (absolute) value is ≈ 0.17 . We, therefore, expand Equation (89) to leading order in this quantity and obtain Equations (49) and (50) for the imaginary and real parts of k .

To the next leading order ($|kr| \gg 1$) after Equations (88), Equation (87) gives

$$\frac{2A'}{A} + \frac{k'}{k} + C_1 = 0, \quad (90)$$

and so the amplitude of δh is

$$A \propto \left(\frac{D}{r\Sigma k} \right)^{1/2} \exp \left[\frac{1}{2} \int^r \left(\frac{\Omega^2 + i\gamma\Omega\tilde{\omega}\beta}{\Omega^2 + \gamma^2\tilde{\omega}^2\beta^2} \right) \frac{dr'}{L_T} \right], \quad (91)$$

where again we have taken $D_c \approx D$.

C.1. Angular Momentum Flux

The AMF (Equation (47)) is, using $|k\delta h| \gg |\delta h/r|$, and again using $D_S \approx D$ (or equivalently $D_c \approx D$),

$$F_J^m \approx \frac{\pi m r \Sigma}{D} \text{Im}(\delta h \delta h^*) \propto \frac{m r \Sigma}{D} \text{Re}(k) |A|^2 \exp \left[-2 \int^r \text{Im}(k) dr' \right]. \quad (92)$$

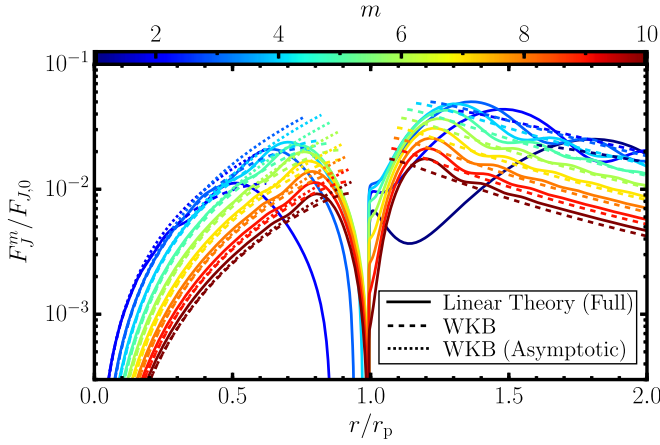


Figure 8. Comparison of different approximations for F_J^m , the Fourier components of the wave AMF for the case with a dimensionless cooling timescale $\beta = 1$. The results of a fully self-consistent linear calculation are shown as solid lines, with the different colors corresponding to different azimuthal mode numbers m . The WKB approximation for F_J^m , (Equation (48)), is represented by the dashed lines, and its asymptotic inner disk behavior (Equation (52)) by the dotted lines. The approximations for F_J^m are only shown exterior to the outer Lindblad resonances and interior to the inner Lindblad resonances. They are also scaled by arbitrary amplitudes (as they are defined only up to an overall constant) to facilitate comparison with the full linear results.

Therefore, using (49) and (91), we have

$$F_J^m(r) = F_J^m(r_0) \times \exp \left\{ \int_{r_0}^r \left[\frac{\Omega^2 L_T^{-1}}{\Omega^2 + \gamma^2 \tilde{\omega}^2 \beta^2} - 2 \text{Im}(k) \right] dr' \right\}, \quad (93)$$

where r_0 is an arbitrary reference radius. Note that we have taken a prefactor $\text{Re}(k)/|k|$ in this expression to be ≈ 1 , since including $\text{Im}(k)$ in this term gives only a small correction and because F_J depends on $\text{Im}(k)$ much more strongly through the exponential.

C.2. Validity of the WKB Approximation

The validity of the WKB approximation of the Fourier components of the wave AMF (Equation (48)) and its asymptotic inner disk approximation (Equation (52)) are examined in Figure 8 for the case with $\beta = 1$. The WKB approximation (dashed lines in Figure 8) gives a good description of the radial variation of F_J^m once the wave AMF has been fully accumulated, at distances from the planet $|r - r_p|$ larger than about $2|r_p - r_{LR}|$, where r_{LR} is the location of either the inner or outer Lindblad resonance (denoted by the endpoints of the dashed lines Figure 8). Therefore, it is typically valid beyond 2–3 scale heights from the planet, except for modes with small azimuthal numbers ($m \lesssim 2$). The inner disk WKB approximation (short dashed lines in Figure 8) has approximately the same region of validity.

ORCID iDs

Ryan Miranda <https://orcid.org/0000-0002-2804-3657>
Roman R. Rafikov <https://orcid.org/0000-0002-0012-1609>

References

- ALMA Partnership, Brogan, C. L., Pérez, L. M., et al. 2015, *ApJL*, 808, L3
 Andrews, S. M., Huang, J., Pérez, L. M., et al. 2018, *ApJL*, 869, L41
 Andrews, S. M., Wilner, D. J., Zhu, Z., et al. 2016, *ApJL*, 820, L40
 Arzamasskiy, L., & Rafikov, R. R. 2018, *ApJ*, 854, 84
 Bae, J., & Zhu, Z. 2018, *ApJ*, 859, 118
 Bae, J., Zhu, Z., & Hartmann, L. 2017, *ApJ*, 850, 201
 Barker, A. J., & Ogilvie, G. I. 2016, *MNRAS*, 458, 3739
 Baruteau, C., & Masset, F. 2008, *ApJ*, 672, 1054
 Benítez-Llambay, P., & Masset, F. S. 2016, *ApJS*, 223, 11
 de Val-Borro, M., Edgar, R. G., Artymowicz, P., et al. 2006, *MNRAS*, 370, 529
 Dong, R., Li, S., Chiang, E., & Li, H. 2017, *ApJ*, 843, 127
 Dong, R., Li, S., Chiang, E., & Li, H. 2018, *ApJ*, 866, 110
 Dong, R., Rafikov, R. R., & Stone, J. M. 2011a, *ApJ*, 741, 57
 Dong, R., Rafikov, R. R., Stone, J. M., & Petrovich, C. 2011b, *ApJ*, 741, 56
 Goldreich, P., & Nicholson, P. D. 1989, *ApJ*, 342, 1075
 Goldreich, P., & Tremaine, S. 1979, *ApJ*, 233, 857
 Goldreich, P., & Tremaine, S. 1980, *ApJ*, 241, 425
 Goodman, J., & Rafikov, R. R. 2001, *ApJ*, 552, 793
 Isella, A., Guidi, G., Testi, L., et al. 2016, *PhRvL*, 117, 251101
 Korycansky, D. G., & Pollack, J. B. 1993, *Icar*, 102, 150
 Lee, W.-K. 2016, *ApJ*, 832, 166
 Lee, W.-K., & Gu, P.-G. 2015, *ApJ*, 814, 72
 Lin, M.-K. 2015, *MNRAS*, 448, 3806
 Lin, M.-K., & Papaloizou, J. C. B. 2011, *MNRAS*, 415, 1445
 Lin, M.-K., & Youdin, A. N. 2015, *ApJ*, 811, 17
 Long, F., Pinilla, P., Herczeg, G. J., et al. 2018, *ApJ*, 869, 17
 Loomis, R. A., Öberg, K. I., Andrews, S. M., & MacGregor, M. A. 2017, *ApJ*, 840, 23
 Lunine, J. I., & Stevenson, D. J. 1982, *Icar*, 52, 14
 Martin, R. G., Nixon, C., Lubow, S. H., et al. 2014, *ApJL*, 792, L33
 Miranda, R., & Lai, D. 2018, *MNRAS*, 473, 5267
 Miranda, R., & Rafikov, R. R. 2019a, *ApJ*, 875, 37
 Miranda, R., & Rafikov, R. R. 2019b, *ApJL*, 878, L9
 Moody, M. S. L., Shi, J.-M., & Stone, J. M. 2019, *ApJ*, 875, 66
 Muñoz, D. J., Miranda, R., & Lai, D. 2019, *ApJ*, 871, 84
 Nazari, P., Booth, R. A., Clarke, C. J., et al. 2019, *MNRAS*, 485, 5914
 Ogilvie, G. I., & Lubow, S. H. 2002, *MNRAS*, 330, 950
 Paardekooper, S. J., Baruteau, C., Crida, A., & Kley, W. 2010, *MNRAS*, 401, 1950
 Paardekooper, S. J., & Papaloizou, J. C. B. 2008, *A&A*, 485, 877
 Papaloizou, J. C. B., Nelson, R. P., Kley, W., Masset, F. S., & Artymowicz, P. 2007, in *Protostars and Planets V*, ed. B. Reipurth, D. Jewitt, & K. Keil (Tucson, AZ: Univ. Arizona Press), 655
 Podlowska-Gaca, E., Papaloizou, J. C. B., & Szuszkiewicz, E. 2012, *MNRAS*, 421, 1736
 Rafikov, R. R. 2002a, *ApJ*, 569, 997
 Rafikov, R. R. 2002b, *ApJ*, 572, 566
 Rafikov, R. R. 2016, *ApJ*, 831, 122
 Rafikov, R. R., & Petrovich, C. 2012, *ApJ*, 747, 24
 Regály, Z., Sándor, Z., Dullemond, C. P., & Kiss, L. L. 2011, *A&A*, 528, A93
 Takeuchi, T., Miyama, S. M., & Lin, D. N. C. 1996, *ApJ*, 460, 832
 Tsang, D. 2011, *ApJ*, 741, 109
 Tsang, D. 2014, *ApJ*, 782, 112
 Ward, W. R. 1997, *Icar*, 126, 261
 Weber, P., Pérez, S., Benítez-Llambay, P., et al. 2019, *ApJ*, 884, 178
 Zhang, H., & Lai, D. 2006, *MNRAS*, 368, 917
 Zhu, Z., Dong, R., Stone, J. M., & Rafikov, R. R. 2015, *ApJ*, 813, 88