Multiple Spiral Arms in Protoplanetary Disks: Linear Theory

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

Abstract

Recent observations of protoplanetary disks, as well as simulations of planet–disk interaction, have suggested that a single planet may excite multiple spiral arms in the disk, in contrast to the previous expectations based on linear theory (predicting a one-armed density wave). We reassess the origin of multiple arms in the framework of linear theory by solving for the global two-dimensional response of a nonbarotropic disk to an orbiting planet. We show that the formation of a secondary arm in the inner disk, at about half of the orbital radius of the planet, is a robust prediction of linear theory. This arm becomes stronger than the primary spiral at several tenths of the orbital radius of the planet. Several additional, weaker spiral arms may also form in the inner disk. On the contrary, a secondary spiral arm is unlikely to form in the outer disk. Our linear calculations, fully accounting for the global behavior of both the phases and amplitudes of perturbations, generally support the recently proposed WKB phase argument for the secondary arm origin (as caused by the intricacy of constructive interference of the azimuthal harmonics of the perturbation at different radii). We provide analytical arguments showing that the process of a single spiral wake splitting up into multiple arms is a generic linear outcome of wave propagation in differentially rotating disks. It is not unique to planet-driven waves and also occurs in linear calculations of spiral wakes freely propagating with no external torques. These results are relevant for understanding formation of multiple rings and gaps in protoplanetary disks.

Key words: hydrodynamics – planet–disk interactions – protoplanetary disks – waves

1. Introduction

The gravitational interaction of young planets with their natal disks is known to produce spiral density waves. Recent high-resolution direct imaging has revealed spiral structures in several protoplanetary disks, including MWC 758 (Grady et al. 2013; Benisty et al. 2015), HD 100453 (Wagner et al. 2015, 2018), and SAO 206462 (Muto et al. 2012; Garufi et al. 2013; Stolker et al. 2016; Maire et al. 2017), which may be produced by the gravitational influence of planets or binary companions. A remarkable feature of these disks is that they display pairs of spiral arms separated by approximately $180°$, which has not been expected.

Indeed, in the conventional picture of planet–disk interactions, a planet is believed to give rise to only a single spiral arm (Ogilvie & Lubow 2002; Rafikov 2002a). Gravitational perturbations due to a planet excite many wave modes in the disk, each described by an azimuthal number m and launched at a corresponding Lindblad resonance—locations interior and exterior to the orbit of the planet where orbital commensurabilities occur (Goldreich & Tremaine 1979). Ogilvie & Lubow (2002) showed that these modes interfere constructively, leading to a characteristic one-armed spiral pattern. In this framework, two planets would be required to produce two arms (Benisty et al. 2015). However, for pairs of arms with similar azimuthal separations to be found in several unrelated protoplanetary disks would require fortuitous configurations of the orbital phases of the planets in these systems.

Recent three-dimensional simulations of planet–disk interactions have demonstrated that some of the observed multiple spiral features can, in fact, be produced by a single orbiting companion (Dong et al. 2015b, 2016a; Fung & Dong 2015; Zhu et al. 2015; Dong & Fung 2017). Notably, the spirals seen in HD 100453 were demonstrated to be consistent with the disturbances produced by the nearby M dwarf companion (Dong et al. 2016b; Wagner et al. 2018). A key finding of these numerical studies is that a single planet can produce multiple spirals, so it is not necessary to invoke the presence of multiple planets to explain the appearance of several spiral arms. In some cases, more than two spiral arms are produced. The number of arms, as well as the azimuthal separation of the two strongest spirals, was found numerically to depend on planet mass (Fung & Dong 2015; Zhu et al. 2015; Bae & Zhu 2018b). It was also shown (Bae et al. 2017) that multiple arms can be related to the formation of annular gaps in millimeter-size dust distribution in protoplanetary disks.

At the same time, Arzamasskiy & Rafikov (2018) recently demonstrated numerically that formation of secondary spirals does not necessarily require the presence of a planet (i.e., an orbiting point mass perturber) driving density waves. In their case, density waves were driven by an imposed boundary condition at the outer edge of the simulation domain and then freely propagated inward, without angular momentum injection by external torques. Such passive propagation of the waves sufficiently far into the inner disk was found to also naturally result in the formation of a secondary arm.

Despite these numerical experiments, the origin of secondary spiral arms has remained elusive. Some nonlinear mechanisms, such as mode coupling (Fung & Dong 2015; Lee 2016), have been proposed to explain their features. Recently, Bae & Zhu (2018a, 2018b) argued that the formation of multiple spirals can be explained by the radially dependent coherence of different azimuthal harmonics of the perturbations driven by a planet, essentially by an extension of the linear mode phase argument of Ogilvie & Lubow (2002). In their work, the phases of the multiple crests of each mode were shown to constructively

interfere in different parts of the disk (at different azimuthal locations) as the wave propagates away from the perturber. The different regions of interference are identified as the primary arm, secondary arm, tertiary arm, and so on. This argument was laid out in terms of the local (WKB) approximation for mode phases and essentially ignored the behavior of the mode amplitudes. Nevertheless, these findings were corroborated by two-dimensional numerical simulations, demonstrating that this idea is a promising step toward understanding the formation of multiple spirals. An important aspect of the work of Bae & Zhu (2018a) is that the emergence of multiple arms was understood, at least in part, within a linear framework.

In this paper, we directly apply the linear theory of density wave evolution to self-consistently compute the full two-dimensional structure of the response of a thin, locally isothermal disk to an orbital companion. By properly accounting for the global behavior of the mode amplitudes, as well as their phases (i.e., going beyond the WKB approximation), we show that multiple spiral arms are robustly formed in the inner regions of protoplanetary disks; under certain circumstances, they can also appear in the outer disk. We characterize the morphology of the spirals (e.g., their amplitudes, widths, and arm-to-arm separations) and its dependence on the disk properties—its aspect ratio, as well as profiles of the temperature and surface density.

The plan of this paper is as follows. In Section 2, we describe our setup and the details of our calculations. In Section 3, we present results on the formation of multiple spirals by a planet and characterize the properties of the spirals and their dependence on the disk parameters. In Section 4, we present calculations of the passive propagation of a spiral wake in a perturber-free disk, demonstrating that the emergence of a secondary spiral is a generic property of wave propagation in differentially rotating disks. In Section 5, we provide theoretical arguments based on linear mode phases in order to understand some key aspects of our calculations. We discuss our results in Section 6 and conclude with a summary of our main results in Section 7.

2. Problem Framework

We consider propagation of density waves in a two-dimensional fluid disk around a star of mass M_* in the linear regime. The foundations of the mathematical framework for studying this phenomenon were laid out in Goldreich & Tremaine (1979), and we heavily borrow from their results. We explore both the inhomogeneous and homogeneous versions of the problem.

In the inhomogeneous case (Section 3), the wave is explicitly driven by the gravitational potential of a planet of mass $M_p \ll M_*$ moving on a circular orbit with radius r_p and Keplerian frequency $\Omega_p = (GM_*/r_p^3)^{1/2}$. The torque due to the planetary gravity both excites the wave in the first place and modifies its subsequent propagation.

In the homogeneous case, the perturbation is imposed at the edge of the disk with no external torques affecting subsequent wave propagation (a setup analogous to Arzamasskiy & Rafikov 2018). This regime is studied using the same mathematical framework as the inhomogeneous case but with the planetary source terms set to zero (Section 4).

2.1. Basic Setup

We consider a very general disk model in which the entropy $S \propto \ln(P/\Sigma^\gamma)$ is allowed to vary with radius r. Here Σ is the disk surface density, $P = \Sigma c_s^2/\gamma$ is the (height-integrated) pressure, c_s is the adiabatic sound speed, and γ is the adiabatic index. We assume that in the unperturbed disk,

$$c_s(r) = h_p r_p \Omega_p \left(\frac{r}{r_p}\right)^{-q/2}, \qquad (1)$$

where h_p is the disk aspect ratio, $h(r) = H/r = h_p(r/r_p)^{(1-q)/2}$, evaluated at r_p, and $H = c_s/\Omega$ is the pressure scale height. The exponent q is the power-law index of the disk temperature, which is proportional to c_s^2. The disk surface density is

$$\Sigma(r) = \Sigma_p \left(\frac{r}{r_p}\right)^{-p}, \qquad (2)$$

where the value of Σ_p is arbitrary. As a result of specifying c_s and Σ independently, the entropy S can vary through the disk.

The orbital frequency of the disk fluid is modified from the pure Keplerian value $\Omega_K = (GM_*/r^3)^{1/2}$ by the pressure gradient:

$$\Omega^2 = \Omega_K^2 + \frac{1}{r\Sigma}\frac{dP}{dr}. \qquad (3)$$

The response of the disk is sensitive to the small deviations of Ω and the radial epicyclic frequency κ, given by

$$\kappa^2 = \frac{2\Omega}{r}\frac{d}{dr}(r^2\Omega), \qquad (4)$$

from the Keplerian frequency Ω_K.

2.2. Equations and Numerical Procedure

The surface density perturbation produced in response to the gravitational potential of the planet (or the externally imposed perturbation in the homogeneous case) is described in polar coordinates (r, ϕ) by $\delta\Sigma(r, \phi)$, which we decompose into Fourier modes according to

$$\delta\Sigma(r, \phi) = \sum_{m=1}^{\infty} \text{Re}\,[\delta\Sigma_m(r)e^{im(\phi - \phi_p)}], \qquad (5)$$

where $\phi_p = \Omega_p t$ is the azimuthal position of the planet (in the homogeneous case, $\phi_p \to 0$). Each $\delta\Sigma_m(r)$ is a complex quantity describing the radial variation of the amplitude and phase of the mode with azimuthal number m. The radial velocity and azimuthal velocity perturbations $\delta u_r(r, \phi)$ and $\delta u_\phi(r, \phi)$ are similarly expressed as sums of Fourier modes $\delta u_{r,m}(r)$ and $\delta u_{\phi,m}(r)$.

As a result of performing these steps, one arrives at the linear equation describing the quantity[3] $\delta h_m = \delta P_m/\Sigma$ (where δP_m is the pressure perturbation) of the mode with azimuthal number

---

[3] For barotropic disks, δh_m can be identified as the enthalpy perturbation. However, in the presence of an entropy gradient, this association no longer holds exactly. Instead, δh_m simply serves as a convenient variable for which a master equation can be found.

m (Baruteau & Masset 2008; Tsang 2014; here we adopt the notation of the latter),

$$\left\{\frac{d^2}{dr^2} + \left[\frac{d}{dr}\ln\left(\frac{r\Sigma}{D_S}\right)\right]\frac{d}{dr} - \frac{2m\Omega}{r\tilde{\omega}}\left[\frac{d}{dr}\ln\left(\frac{\Sigma\Omega}{D_S}\right) + \frac{1}{L_S}\right]\right.$$
$$\left. + \frac{m^2}{r^2}\left(\frac{N_r^2}{\tilde{\omega}^2} - 1\right)\right\}(\delta h_m + \Phi_m)$$
$$- \left[\frac{D_S}{c_s^2} + \frac{1}{L_S^2} + \frac{1}{L_S}\frac{d}{dr}\ln\left(\frac{r\Sigma}{L_S D_S}\right) + \frac{2m\Omega}{L_S r\tilde{\omega}}\right]\delta h_m$$
$$+ \frac{1}{L_S}\frac{d\Phi_m}{dr} = 0, \tag{6}$$

where $\tilde{\omega} = m(\Omega_p - \Omega)$ is the Doppler-shifted frequency of the tidal forcing due to the mth harmonic of the planetary potential,

$$\frac{1}{L_S} = \frac{1}{\gamma}\frac{dS}{dr} = \frac{(\gamma-1)p - q}{\gamma r} \tag{7}$$

is the inverse length scale associated with the radial variation of entropy,

$$N_r^2 = -\frac{1}{\Sigma^2}\frac{dP}{dr}\left(\frac{1}{c_s^2}\frac{dP}{dr} - \frac{d\Sigma}{dr}\right)$$
$$= \frac{(q+p)}{\gamma}\frac{c_s^2}{L_S r} \tag{8}$$

is the squared Brunt–Väisälä frequency, and $D_S = \kappa^2 - \tilde{\omega}^2 + N_r^2$. Note that in the barotropic (uniform entropy) limit, $1/L_S \to 0$ and $N_r \to 0$, Equation (6) reduces to the master equation of Goldreich & Tremaine (1979). The surface density perturbation $\delta\Sigma_m$ can be computed using solutions of Equation (6) according to

$$\delta\Sigma_m = \frac{\Sigma}{c_s^2}\delta h_m + \frac{i\Sigma}{L_S\tilde{\omega}}\delta u_{r,m}, \tag{9}$$

where the radial velocity perturbation $\delta u_{r,m}$ is given in terms of δh_m and its radial derivative $\delta h_m'$ by

$$\delta u_{r,m} = \frac{i}{D_S}\left[\tilde{\omega}(\delta h_m' + \Phi_m')\right.$$
$$\left. - \frac{2m\Omega}{r}(\delta h_m + \Phi_m) - \frac{\tilde{\omega}}{L_S}\delta h_m\right]. \tag{10}$$

The azimuthal velocity perturbation can be found in a similar fashion,

$$\delta u_{\phi,m} = \frac{1}{D_S}\left[\frac{\kappa^2}{2\Omega}(\delta h_m' + \Phi_m')\right.$$
$$\left. - \frac{m\tilde{\omega}}{r}\left(1 - \frac{N_r^2}{\tilde{\omega}^2}\right)(\delta h_m + \Phi_m) - \frac{\kappa^2}{2\Omega L_S}\delta h_m\right]. \tag{11}$$

The components of the gravitational potential of the planet are

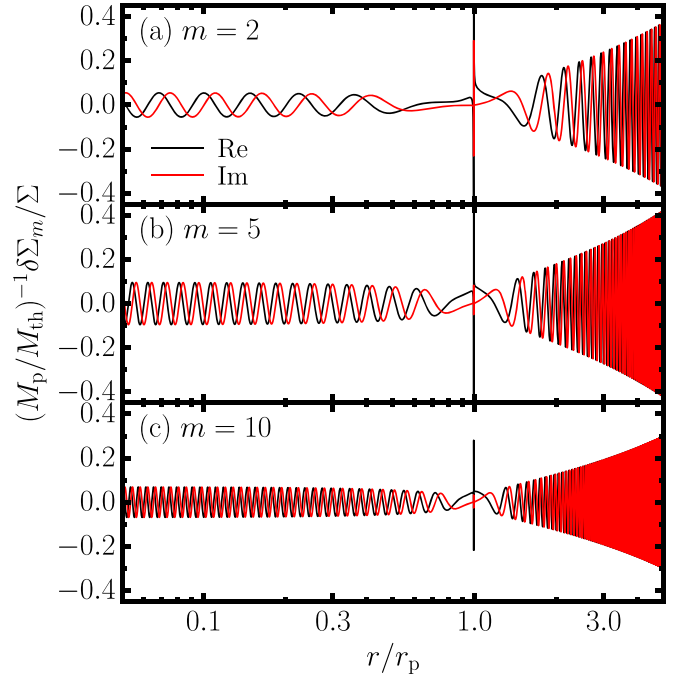$$\Phi_m = -\frac{GM_p}{r_p}b_{1/2}^{(m)}(r/r_p), \tag{12}$$

Figure 1. Radial profiles of the fractional surface density perturbation for several low-order modes for the case of the fiducial parameters $h_p = 0.1$, $q = 1$, and $p = 1$.

where $b_{1/2}^m$ are (softened) Laplace coefficients,

$$b_{1/2}^{(m)}(\alpha) = \frac{1}{\pi}\int_0^{2\pi}\frac{\cos(m\psi)d\psi}{[1 - 2\alpha\cos(\psi) + \alpha^2 + \epsilon^2]^{1/2}}. \tag{13}$$

For the softening parameter, we choose $\epsilon = 0.6h_p$, a value that is typically used in two-dimensional numerical simulations of planet–disk interaction to mimic the vertical averaging of the planetary gravity over the disk height. In our calculations, as in Bae & Zhu (2018a), we ignore the "indirect" potential term, $\delta_{m,1}GM_p r/r_p^2$, which arises due to the motion of the central star around the barycenter of the star + planet system. We motivate this choice in the Appendix.

We solve Equation (6) for a sufficient number of modes to fully capture the two-dimensional structure of the surface density perturbations. The mode solution method closely follows that of Korycansky & Pollack (1993, hereafter KP93), as well as Petrovich & Rafikov (2012) and Rafikov & Petrovich (2012), and is described in detail in the Appendix. The solution is obtained on a logarithmic grid with $r_{in} = 0.05r_p$ and $r_{out} = 5.0r_p$, which, for our fiducial parameters, has 6×10^4 grid points, or a resolution of about $1300/H$. All of the modes are solved on the same grid to facilitate their synthesis, and the grid resolution is set by the tight winding of the highest m modes near the grid boundaries.

Examples of the surface density perturbation profiles $\delta\Sigma_m(r)$ for several low-order azimuthal modes are shown in Figure 1. Note that modes with higher m are more tightly wound. Also note that here the $m = 5$ mode has the largest amplitude. This is because the response of the disk is dominated by modes with m close to $m_* \approx (2h_p)^{-1}$ (here $h_p = 0.1$ and $m_* = 5$). The dominant role of this characteristic m is related to the torque cutoff phenomenon (Goldreich & Tremaine 1980). The exact form of m_* (i.e., the factor of 1/2) is not fundamental, but this choice is supported by our numerical calculations.

3

Once the mode solutions are found, the two-dimensional surface density perturbation is then synthesized according to

$$\delta\Sigma(r, \phi) = \sum_{m=1}^{m_{max}} \text{Re}\left[\delta\Sigma_m(r) e^{im(\phi-\phi_p)}\right], \tag{14}$$

where m_{max} is the value of m necessary to achieve a converged perturbation structure; see the Appendix for details. The velocity perturbations $\delta u_r(r, \phi)$ and $\delta u_\phi(r, \phi)$ are computed in the same manner. Note that the mode perturbation, $\delta\Sigma_m(r)$, is a one-dimensional, complex quantity (described by a radially varying amplitude and phase), while the synthesized perturbation, $\delta\Sigma(r, \phi)$, is a two-dimensional, explicitly real quantity.

2.3. Parameters

The results of our calculations are fully determined by four dimensionless parameters: h_p, the disk aspect ratio at the orbital radius of the planet; q, the power-law index of the disk temperature; p, the power-law index of the disk surface density; and the adiabatic index, γ. For the fiducial parameters, we choose $h_p = 0.1$, $q = 1$ (corresponding to a constant disk aspect ratio h), $p = 1$, and $\gamma = 7/5$. We find that our results are almost completely insensitive to the value of γ (see Section 3.3), so unless otherwise stated, we keep its value fixed. We have explored the parameter space by performing calculations for which two of the three remaining parameters are fixed at their fiducial values and the third is varied over a plausible range of values for protoplanetary disks: $0.05 < h_p < 0.15$, $0 < q < 1$, and $0 < p < 3/2$. We find that the results are not sensitive to p, so we primarily focus our analysis and discussion on the effects of varying h_p and q.

3. Results for Planet-driven Spirals

We will start the presentation of our results with the case of a spiral pattern driven by the gravity of an embedded planet (forced or inhomogeneous case). To better highlight new findings, we start by outlining the existing picture of wave propagation in disks in Section 3.1. We then describe the general properties of the linear planet-driven density waves found in our linear calculations (Section 3.2) and provide a comparison with the results of direct numerical simulations (Section 3.4). These preliminaries form a basis for subsequent in-depth discussion of the properties of multiple spiral arms emerging in our calculations in Section 3.5.

3.1. Expectations Based on Simple Linear Theory

Existing linear calculations of the density wave propagation in disks provide some guidance on the expected outcome of our present calculation. Specializing to the inhomogeneous case, the linear response of the disk is expected to take the form of a one-armed spiral density wake (Ogilvie & Lubow 2002; Rafikov 2002a). For $|r - r_p| \gg H_p$, the position of the wake is given approximately by[4]

$$\phi_{lin} = \phi_p + \text{sgn}(r - r_p) \int_{r_p}^{r} \frac{\Omega(r') - \Omega_p}{c_s(r')} dr', \tag{15}$$

and its amplitude (peak height) is given by, to within a constant factor of order unity (Rafikov 2002a),

$$\frac{\delta\Sigma_{lin}}{\Sigma(r)} = \frac{M_p}{M_{th}}$$
$$\times \left[\frac{\Omega(r) - \Omega_p}{\Omega_p} \frac{\Sigma_p}{\Sigma(r)} \frac{r_p}{r}\right]^{1/2} \left[\frac{c_s(r)}{c_s(r_p)}\right]^{-3/2}. \tag{16}$$

Here

$$M_{th} = h_p^3 M_* \tag{17}$$

is the thermal mass. The first (constant) factor on the right-hand side of Equation (16) describes the initial amplitude of the density wake formed within a few scale heights of the planet (in the homogeneous case considered in Section 4, r_p is replaced with the radius at which the perturbation is imposed and M_p/M_{th} with an arbitrary constant), while the other factors indicate how the amplitude varies as the wake propagates away from the planet. The radial scaling is dictated by the conservation of angular momentum flux (AMF),

$$F_J(r) = r^2 \Sigma(r) \oint \delta u_r(r, \phi) \delta u_\phi(r, \phi) d\phi, \tag{18}$$

which, in the absence of explicit dissipation (linear or nonlinear), must be constant far from the planet, outside the wave excitation region (Goodman & Rafikov 2001; Rafikov 2002a). The characteristic scale of F_J, resulting from the sum of the one-sided Lindblad torques, is (Goldreich & Tremaine 1980; Ward 1997)

$$F_{J,0} = \left(\frac{M_p}{M_*}\right)^2 h_p^{-3} \Sigma_p r_p^4 \Omega_p^2. \tag{19}$$

These known linear results will be used as a reference for comparison in our current calculations.

3.2. Results of Linear Calculations: General Properties

In Figure 2, we show two-dimensional maps of the perturbed surface density (in polar coordinates r, ϕ) resulting from our typical linear calculations. The case with the fiducial parameters is shown (Figure 2(a)), as well as several other cases, discussed in Section 3.6, in which each of the three parameters (indicated at the top of the panel) is varied from its fiducial value.

The same maps of the perturbed surface density are also shown in Cartesian coordinates in Figure 3, in order to demonstrate the true geometry of the spirals as seen by an observer sensitive to the gas surface density perturbations. Note that in Figure 3, only the part of the disk interior to the orbit of the planet is shown, and the range of color scale for the perturbations is smaller than in Figure 2, in order to focus on the details of the perturbations in the inner disk.[5]

Also, to highlight the details of the spiral arm evolution, in Figures 4(a) and (b), we display the azimuthal profiles of $\delta\Sigma$ at a fixed radius r, normalized by the linear wake amplitude $\delta\Sigma_{lin}$

---

[4] Note that the analogous expression in Rafikov (2002a) has a different sign for the second term.

[5] We point out that radially narrow features seen at $r \approx r_p$ in Figures 2 and 3 are perturbations in the corotation region that result from the presence of a radial entropy gradient (i.e., due to terms depending on $1/L_S$ in Equation (6)). As they are very localized near the orbit of the planet and have no effect on the global spiral structure of the disk, we do not examine these features in detail in this work.
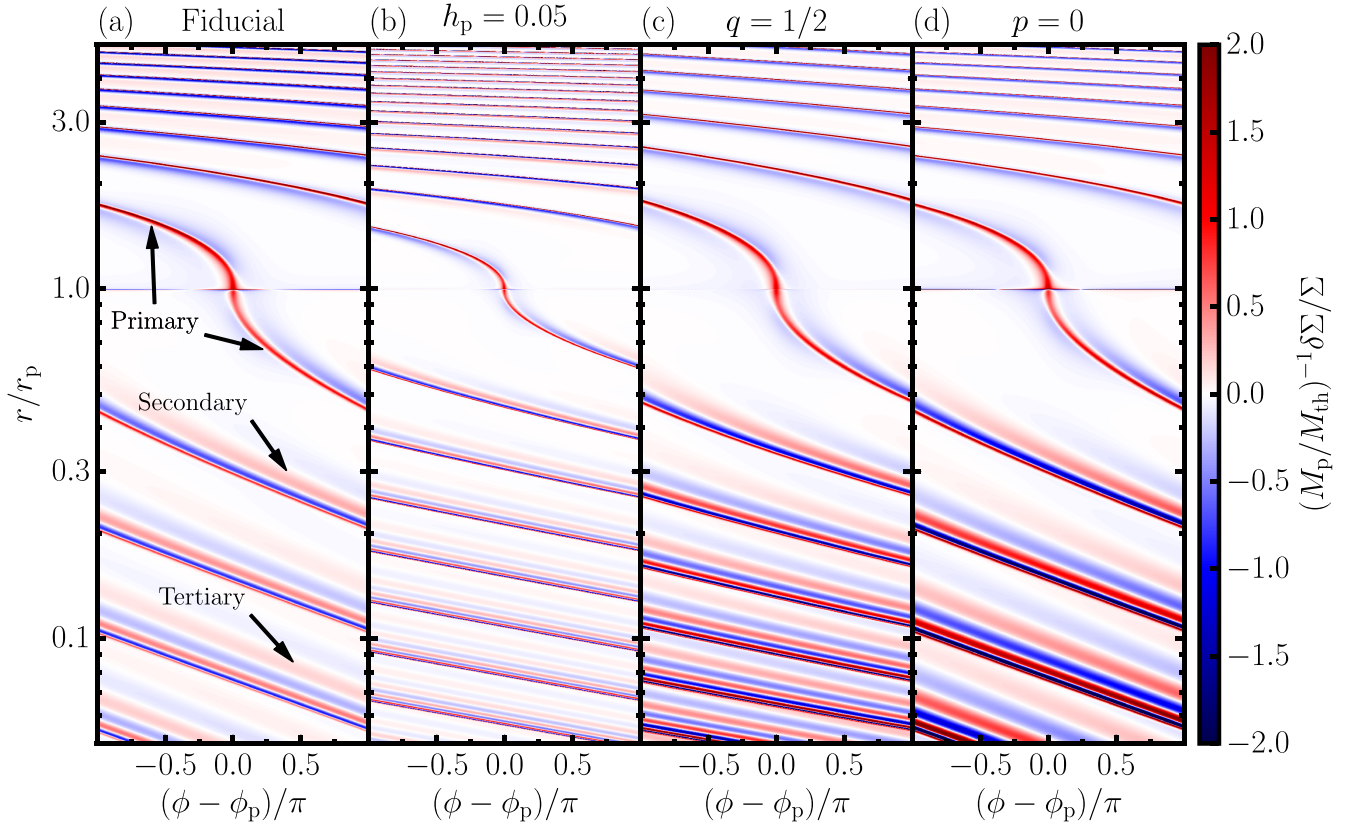
Figure 2. Two-dimensional fractional surface density perturbation (scaled by the ratio of the planet mass to the thermal mass), shown in polar coordinates, for the case of the fiducial parameters ($h_p = 0.1$, $q = 1$, $p = 1$; panel (a)) and several other cases, demonstrating the effect of varying each parameter: the disk aspect ratio, h_p (panel (b)); the temperature power-law index, q (panel (c)); and the surface density power-law index, p (panel (d)). The positions of the primary, secondary, and tertiary arms are indicated in panel (a).

given by Equation (16). These profiles can be thought of as horizontal cuts in Figure 2.

Generically, in the outer disk ($r > r_p$), a single strong spiral arm, i.e., a narrow structure with a $\delta\Sigma > 0$ peak, is present. Its azimuthal profile remains relatively unchanged as the arm winds up with increasing r; see Figure 4(a). It is accompanied by a comparatively wider surface density trough with $\delta\Sigma < 0$, which trails behind the arm (in the sense of the Keplerian rotation of the disk). The case with a smaller disk aspect ratio, $h_p = 0.05$, represents an exception, as it develops a weak additional spiral arm beyond $r \approx 3r_p$, although this is not clearly discernible in Figure 2 (see Section 3.5 for details).

In the inner disk ($r < r_p$), there is a single spiral arm (the "primary spiral") near the planet, accompanied by a trough that leads the arm. However, farther from the planet, the structure of the density wake deviates from the simple behavior seen in the outer disk. For the fiducial parameters, the presence of a second peak with $\delta\Sigma > 0$ (the "secondary spiral arm") becomes apparent at $r \approx 0.5r_p$, accompanied by a deepening of the initial trough; see Figure 4(b). Toward smaller radii, the strength of the secondary spiral increases, and it also becomes accompanied by a leading trough. At even smaller radii ($r \lesssim 0.1r_p$), a third arm (the "tertiary spiral") forms. These multiple spirals are robustly present for a variety of disk parameters. Note that the additional spirals (secondary, tertiary, and so on) are always located ahead of the primary spiral.

In Figures 4(c) and (d), we also show the azimuthal distribution of AMF,

$$f_J(r, \phi) = r^2 \Sigma(r) \delta u_r(r, \phi) \delta u_\phi(r, \phi), \qquad (20)$$

at different radii. This quantity is related to the AMF according to

$$F_J(r) = \oint f_J(r, \phi) d\phi. \qquad (21)$$

The evolution of the azimuthal profile of f_J with r allows us to trace the exchange of AMF between the multiple arms as the wake propagates away from the planet.

The AMF is peaked near the locations of both the peaks ($\delta\Sigma > 0$) and the troughs ($\delta\Sigma < 0$) of the surface density perturbation, since the AMF is approximately proportional to the square of $\delta\Sigma$ far from the planet (Rafikov 2002a):

$$f_J \approx \frac{rc_s^3}{|\Omega - \Omega_p|\Sigma}(\delta\Sigma)^2. \qquad (22)$$

In particular, in the outer disk, for $r/r_p \gtrsim$ a few, the trough trailing the primary arm of the surface density perturbation carries a significant fraction of the AMF.

We provide more in-depth discussion of the different features of the density wake in Section 3.5.
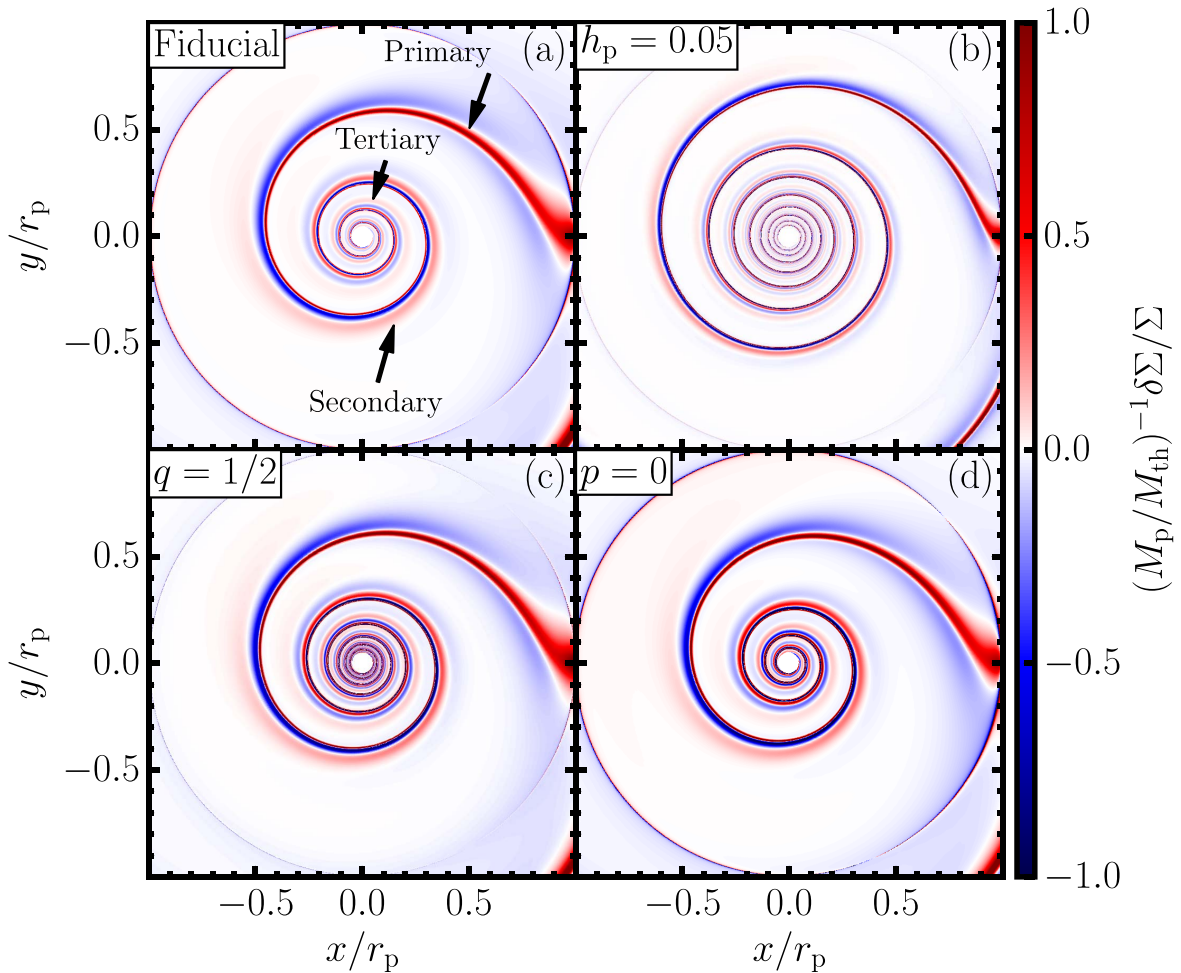
Figure 3. Two-dimensional fractional surface density perturbation for the fiducial parameters and several cases with varied parameters, as in Figure 2, but in Cartesian coordinates and focusing only on the spiral structure in the inner disk, interior to the orbit of the planet. The positions of the primary, secondary, and tertiary arms are indicated in panel (a).

3.3. Dependence on Adiabatic Index

In the calculations presented in this section, we have adopted $\gamma = 7/5$ as the fiducial adiabatic index. We have also carried out calculations for $h_p = 0.1$, $q = 1$, and $p = 1$ with different values of γ, the results of which are shown in Figure 5. Here the azimuthal profiles of $\delta\Sigma$ are shown at different radii for several values of γ in addition to $\gamma = 7/5$. The case $\gamma = 1.001$ corresponds to an almost (locally) isothermal disk. The case $\gamma = 4/3$ represents the effective two-dimensional adiabatic index corresponding to a three-dimensional adiabatic index $\Gamma = 7/5$. These different adiabatic indices are related according to $\gamma = (3\Gamma - 1)/(\Gamma + 1)$ (Goldreich et al. 1986; Ostriker et al. 1992) in the low-frequency limit, $\tilde{\omega} \ll \Omega$ (valid for $r \approx r_p$). For $\gamma = 2$, the disk has a uniform entropy profile (see Equation (7)), and we recover the barotropic limit.

The surface density profiles shown in Figure 5 demonstrate that our results are remarkably insensitive to the value of γ.[6] Mathematically, this is because γ-dependent contributions appear in Equation (6) only through terms varying on scales of $\sim r_p$. In Section 5, we show that the formation of multiple spirals is well described by considering the phases of different

modes in the local (WKB) limit, in which the globally varying terms in Equation (6) with explicit dependence on γ are unimportant. Therefore, it is sufficient to only consider the fiducial $\gamma = 7/5$ in our subsequent discussion.

3.4. Numerical Validation

To validate and understand the limitations of our semi-analytical linear calculations, we have also carried out a set of direct numerical simulations of planet–disk interaction in the low-mass regime using FARGO3D (Benítez-Llambay & Masset 2016). These simulations will be discussed in detail in a future work. Here we describe only the basic setup, as well as the results for the fiducial disk parameters, $h_p = 0.1$, $q = 1$, and $p = 1$.

We choose a planet mass $M_p = 10^{-5} M_*$, or $0.01 M_{th}$. Since $M_p \ll M_{th}$, the linear regime is appropriate for describing the disk response, except far ($\gtrsim 6H_p$) from the planet, where nonlinear effects (fully captured in simulations)—wake evolution into a shock and subsequent dissipation—become nonnegligible (Goodman & Rafikov 2001). A softening length of $0.6 H_p$ is applied to the potential of the planet. We use a logarithmically spaced radial grid extending from $r_{in} = 0.05 r_p$ to $r_{out} = 5.0 r_p$ with $N_r = 4505$ grid cells and a uniformly spaced azimuthal grid with $N_\phi = 6144$ grid cells. The resulting grid cells have a roughly square shape with a resolution of 98 cells per scale height. Wave damping zones

---

[6] The perturbation structure in the immediate vicinity of r_p (within about $0.1H$) has a strong dependence on γ, but this is not important for our present study of spiral arms far from the planet.
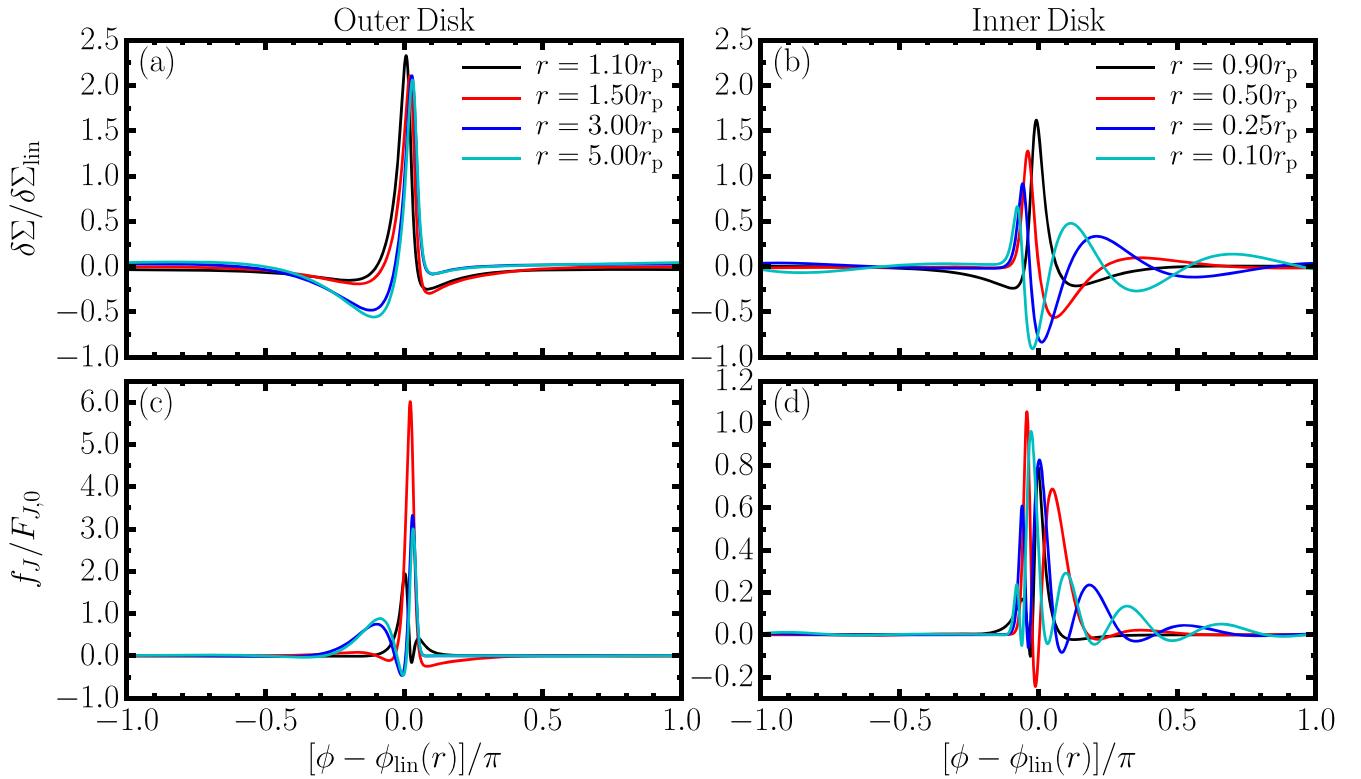
Figure 4. Top panels: profiles of the surface density perturbation $\delta\Sigma$, scaled by $\delta\Sigma_{\mathrm{lin}}$ (see Equation (16)) to facilitate comparison between the profiles at different radii. Bottom panels: profiles of the azimuthal distribution of AMF f_J, in terms of the characteristic AMF $F_{J,0}$ (Equation (19)). Profiles are shown for several different radii in the outer disk (left panels) and inner disk (right panels) for the fiducial disk parameters.

(e.g., de Val-Borro et al. 2006) are implemented for $r < 0.06r_{\mathrm{p}}$ and $r > 4.5r_{\mathrm{p}}$ in order to minimize wave reflection at the boundaries (note that direct comparison with the results of our linear calculations is not possible in the damping zones).

We adopt an ideal equation of state with $\gamma = 1.001$. This value of γ close to unity is chosen in order to avoid any significant heating of the disk due to wave dissipation during our simulations. Note that although this setup resembles that of a locally isothermal disk, the two cases are physically distinct. In locally isothermal disks, nonaxisymmetric perturbations are known to exchange angular momentum with the background flow (Lin & Papaloizou 2011; Lin 2015), which can lead to anomalous results.[7] No explicit viscosity is included in the simulation.

We compare the results of the numerical simulation after 10 orbits (when a steady state is reached) with our linear calculations (using $\gamma = 1.001$ in order to facilitate direct comparison) in Figure 6. The profiles of the surface density perturbation $\delta\Sigma$ are shown at several different disk radii in the inner disk, highlighting the multiple spiral arm structure. The profiles are rotated by $\phi_{\mathrm{lin}}(r)$ and scaled by $\delta\Sigma_{\mathrm{lin}}(r)$, as in Figure 4.

The numerical results exhibit very good agreement with the linear prediction. The positions of the primary, secondary, and tertiary arms are closely reproduced, differing by $\lesssim 0.05$ rad from the linear prediction even at the smallest radii. In particular, the initial secondary-to-primary arm separation is $\approx 60°$, and this separation decreases toward smaller radii, in agreement with our calculations (see Section 3.5). This is a unique feature of the secondary spiral in the linear regime, which should be contrasted with the $180°$ separation that arises

when $M_{\mathrm{p}} > M_{\mathrm{th}}$ (e.g., Zhu et al. 2015). The amplitudes of the arms and troughs in the numerical simulation also show good agreement with the linear prediction, exhibiting essentially negligible differences, except for at small radii ($r \lesssim 0.1r_{\mathrm{p}}$), where differences of $\approx 20\%$ arise due to nonlinear effects (especially in the amplitude of the trough between the primary and secondary spirals). Notably, the amplitude of the secondary arm overtaking that of the primary for $r \lesssim 0.1r_{\mathrm{p}}$ is well reproduced in the simulation. The agreement between the numerical simulation and linear theory supports full applicability of our results to low-mass planets.

We note that no shocks are present in Figure 6, in contradiction to the results of Rafikov (2002a), which predict shock formation at about $0.3r_{\mathrm{p}}$. However, the calculations of Rafikov (2002a) were carried out for the case of a single spiral arm propagating in a self-similar fashion. In our calculations, there is an exchange of AMF from the primary spiral arm to the secondary spiral arm, which modifies this picture. Evidently, this exchange lowers the amplitude of the primary arm and suppresses shock formation for the planet mass ($M_{\mathrm{p}} = 0.01M_{\mathrm{th}}$) that we have considered here.

3.5. Structure of the Multiple Spirals

We now characterize the morphological properties—amplitude, width, and arm-to-arm separation—of the spiral arms, the radial dependence of these properties, and how they vary with the disk parameters. We first briefly describe the procedure for identifying and characterizing the spiral arms, the results of which are illustrated in Figures 7 and 8. We emphasize that the procedure for decomposing $\delta\Sigma$ into discrete spiral arms and troughs is heuristic and not unique; in practice, this decomposition can be done using different criteria than the ones we

---

[7] We look into this issue in more detail in R. Miranda & R. Rafikov (2019, in preparation).
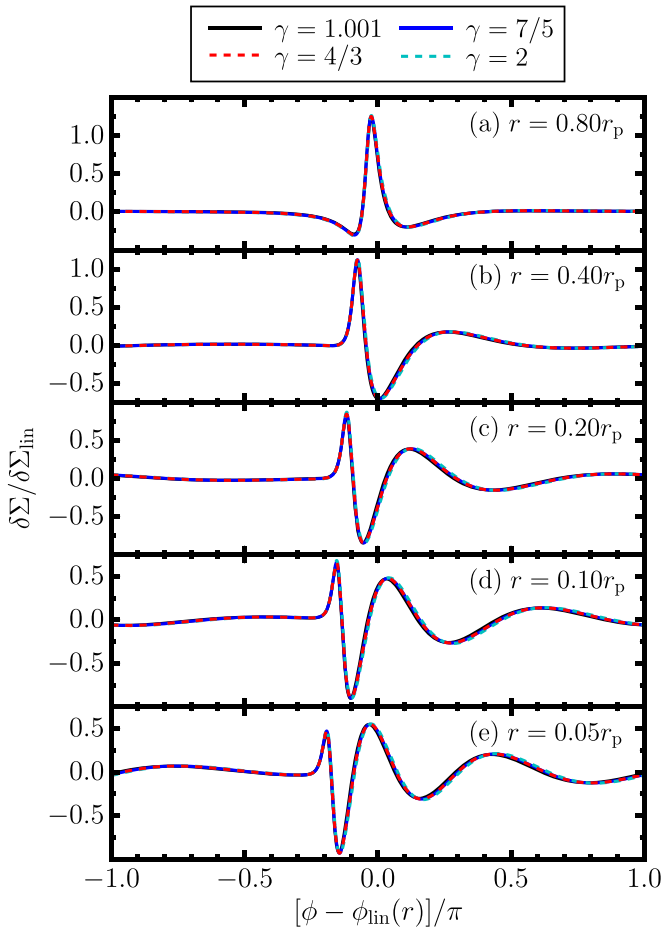
Figure 5. Surface density perturbation profiles in the inner disk for the fiducial disk parameters, $h_\mathrm{p} = 0.1$ and $q = p = 1$, and different values of the adiabatic index γ.
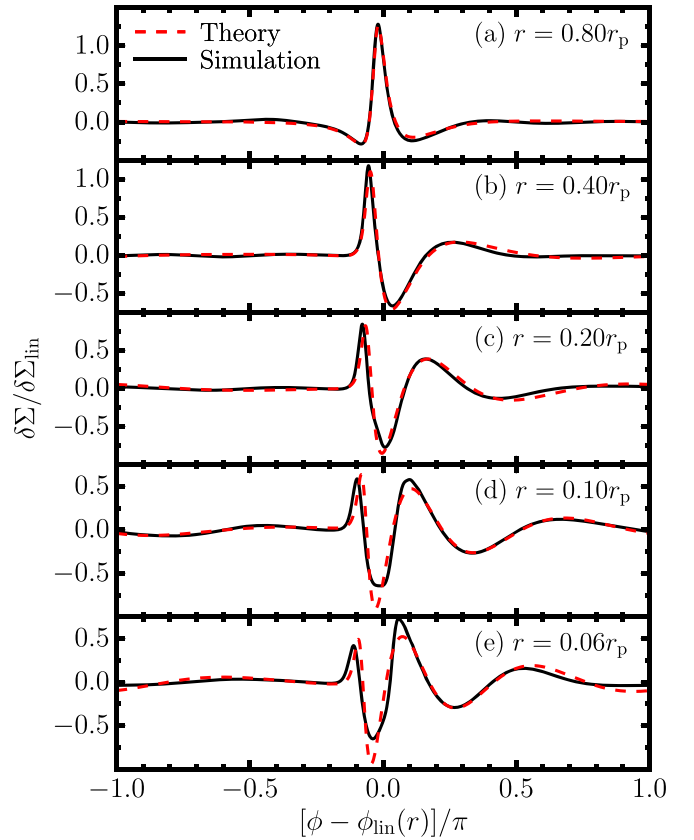

Figure 6. Profiles of the surface density perturbation at several radii in the inner disk, comparing the results of our numerical simulation (solid curves) and linear calculation (dashed curves) for the fiducial parameters (with $\gamma = 1.001$).

used. Therefore, the results should be taken as semiquantitative descriptions.

3.5.1. Identification and Characterization of Spiral Arms

At $r = r_\mathrm{p}$, $\delta\Sigma$ has a single (global) maximum, located at $\approx \phi_\mathrm{p}$, which we identify as the primary arm. By following the position of this maximum across different radii, we obtain the position of the primary arm as a function of r. If at some r, a second (local) maximum with an amplitude equal to at least 10% of the amplitude of the primary arm is present, it is identified as the secondary spiral arm. This relative amplitude threshold is necessary in order to avoid spurious "detections" of additional spirals. The position of the secondary is then followed across different radii in the same manner as the primary. The emergence of subsequent (e.g., tertiary, quaternary) arms is determined by a similar criterion, except we require the amplitude to be at least 10% of the strongest spiral arm at a given r, which may not necessarily be the primary (since its amplitude decays with distance from the planet). The positions of the spiral arms are denoted ϕ_P, ϕ_S, and so on.

We also identify the troughs, i.e., local minima of $\delta\Sigma$. We associate each trough with a spiral arm at each r, since features tend to appear in arm/trough pairs (see Figure 4). The troughs are therefore designated as the "primary trough," "secondary trough," and so on. Note that near the planet, the primary arm is accompanied by two troughs, one leading and one trailing.

Toward the outer disk, the trailing trough becomes the stronger of the two, so it is identified as the primary trough. In the inner disk, the leading trough becomes stronger, so it is identified as the primary trough instead. This is the origin of the discontinuity in the position of the primary trough in Figures 7(a)–(c) and 8(a) and (b). Multiple arms formed in the inner disk are each similarly accompanied by a leading trough. In the outer disk, when multiple arms are present (we find at most two, and only for our thinnest disk), they are instead accompanied by trailing troughs.

Using the profiles of $\delta\Sigma$ in the vicinity of the peaks and troughs, we quantify the amplitudes and widths of the arms/troughs using the maximum and FWHM of $\delta\Sigma$ for arms and $-\delta\Sigma$ for troughs. The arm amplitudes are denoted by $\delta\Sigma_\mathrm{P}$, $\delta\Sigma_\mathrm{S}$, etc.

Note that since troughs are simply identified as local minima, it is sometimes the case (for high-order troughs, e.g., tertiary and beyond) that $\delta\Sigma > 0$ at the trough location. In this case, the width of the trough is undefined, since the amplitude is measured relative to zero. But as the trough evolves, eventually, $\delta\Sigma$ at the minimum becomes negative, so the trough has a well-defined (but narrow) width. The trough width increases as its amplitude grows, until it resembles the initial widths of lower-order troughs, and then it becomes more narrow toward the inner disk, in accordance with the behavior of the other arms/troughs. This explains the anomalous behavior of some of the high-order trough widths seen in, e.g., Figures 7(f) and 8(c) and (d).

We also keep track the AMF F_J carried by each arm or trough by integrating the AMF distribution (Equation (20))
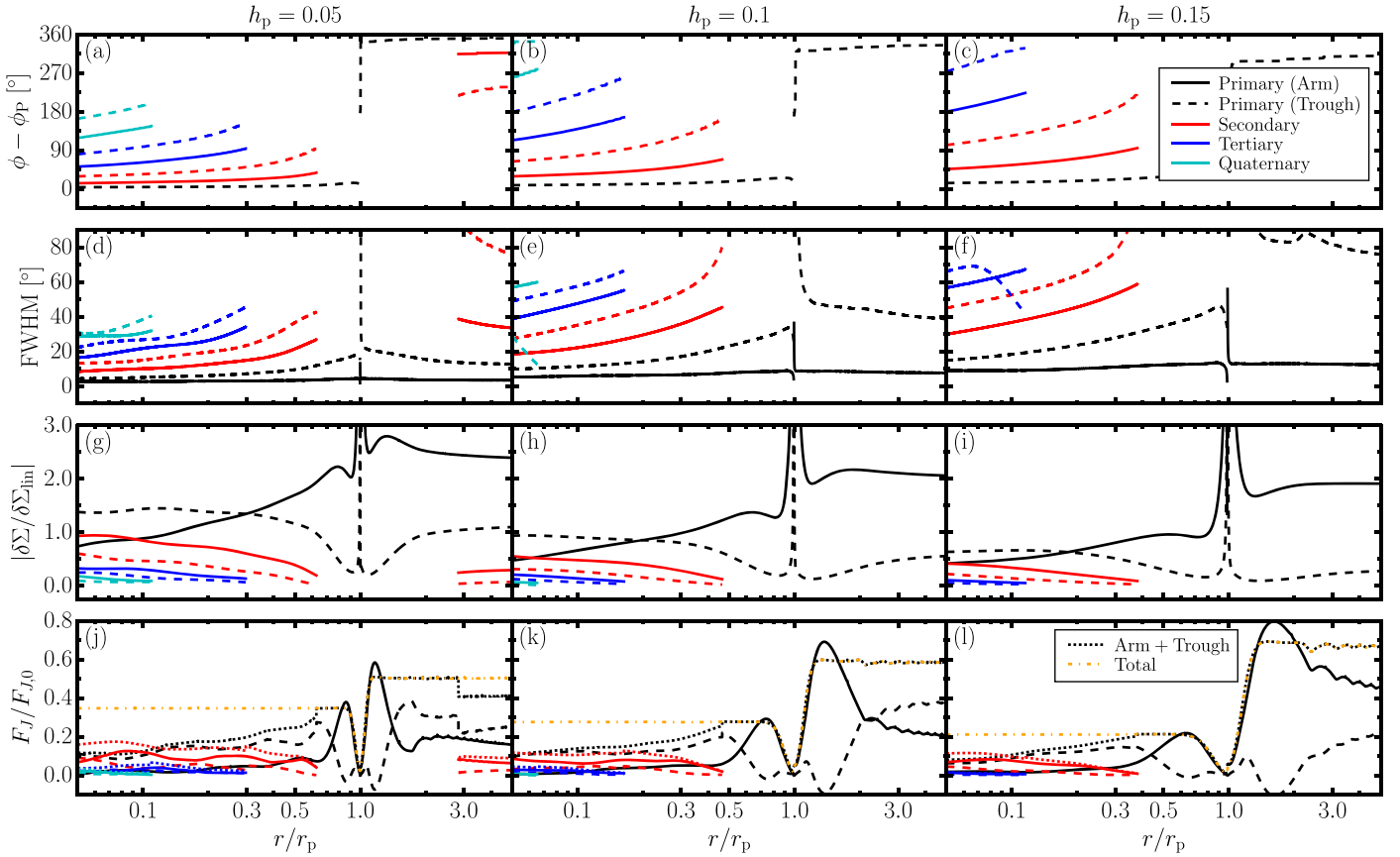
Figure 7. Properties of the multiple spiral arms (solid curves) and their associated troughs (dashed curves) as functions of radius for different values of the disk aspect ratio near the planet, h_p (left, right, and middle panels), and with fixed values of the temperature and surface density power-law indices $q = 1$ and $p = 1$. First row: positions of the arms and troughs relative to the primary arm. Second row: FWHMs of the arms and troughs. Third row: absolute values of the surface density perturbations associated with the arms and troughs, scaled by $\delta\Sigma_{lin}$ (Equation (16)). Fourth row: AMF F_J associated with each arm or trough, as well as the sum of the AMF of each arm and its associated trough (dotted curves) and the total AMF (sum over all of the arms and troughs; dotted–dashed curves).

over its azimuthal extent, which is delineated by the zeros of $\delta\Sigma$. These roughly correspond to the zeros of f_J, since the latter is approximately proportional to $(\delta\Sigma)^2$ (see Figure 4 and Equation (22)). Note that the AMF is divided up and assigned to the arms and troughs in a conservative manner, so that the sum of the AMF associated with all of the peaks and troughs at a given radius is equal to $F_J(r)$.

The extracted properties of the arms and troughs (position, width, amplitude, and AMF) are shown for the fiducial case and cases with different disk aspect ratios in Figure 7 and for cases with different temperature power-law indices in Figure 8. Note that the arm/trough positions are shown relative to the position of the primary arm, ϕ_P, which is well approximated by Equation (15) (although it may differ by ∼10° far from the planet; see Figure 4). Also note that the amplitudes are scaled by $\delta\Sigma_{lin}$, in order to remove the simple variation of amplitude with r resulting from AMF conservation. The AMF (lower panels) is scaled by the characteristic AMF due to the one-sided Lindblad torques given by Equation (19).

3.5.2. Outer Disk

The structure of the surface density perturbation (see Figures 7(g)–(i) and 8(e) and (f)) in the outer disk is fairly simple. Typically, the perturbation consists only of a primary arm, with a width of about 10°, and its associated trough, which is about four times wider. Their amplitudes vary slowly, with the gradual decay of the arm amplitude accompanied by the

gradual growth of the trough amplitude. The amplitude of the trough remains small relative to the arm (a few tenths at most), although it carries a significant fraction of the AMF (see Figures 7(j)–(l) and 8(g) and (h)) due to its large width. For $h_p = 0.1$ and smaller (with $q = 1$), the trough carries most of the AMF (i.e., more than the peak) for $r \gtrsim (1.5–2.0)r_p$.

For the thinnest disk we have considered ($h_p = 0.05$), a secondary arm does form at about $3r_p$ (see Figure 7(g)), which is explained in Section 5.2.2. Its amplitude grows very slowly with r and is only about 12% of that of the primary at $5r_p$. Note that we have solved for the structure of the perturbations only out to $5r_p$, so it is possible that a secondary arm emerges beyond this radius for larger h_p as well (although our analytic estimates indicate that this is unlikely; see Section 5.2.2). If so, it forms very far from the planet in comparison to the inner disk, where a secondary arm always forms at about half of the orbital radius of the planet. And as demonstrated by the $h_p = 0.05$ case, the outer secondary spiral is very weak, even when it exists. Therefore, we do not devote much attention to secondary spirals in the outer disk.

3.5.3. Inner Disk

In the inner disk, multiple (three to five) spiral arms are robustly formed for a variety of parameters. For the fiducial parameters (Figures 7(b), (e), (h), and (k)), the secondary arm first appears at $0.46r_p$, with an initial separation of about 70° from the primary arm and an initial width of 45°. At smaller
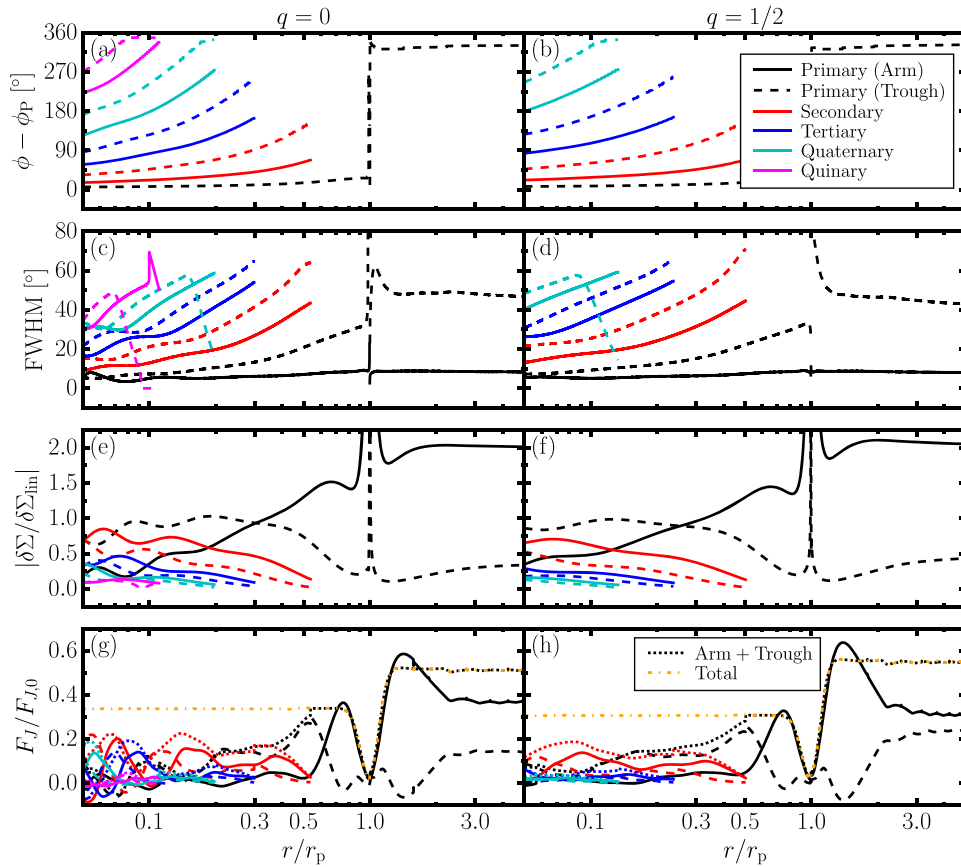
Figure 8. Same as Figure 7, but for different values of the temperature power-law index q, with $h_p = 0.1$.

radii, its separation relative to the primary arm decreases to about $30°$, and its width decreases to about $20°$. The amplitude of the secondary arm increases as that of the primary decreases. At $r = 0.17r_p$, the amplitude of the secondary is about half that of the primary, and at $r = 0.06r_p$, the secondary amplitude exceeds the primary amplitude.

A weak tertiary arm forms at about $0.3r_p$, and a weak quaternary arm forms at about $0.1r_p$. The tertiary and quaternary arms each form at $\approx 135°$ from the previous arm (roughly twice the secondary-to-primary separation), with an initial width of about $60°$; see Figures 7(b) and (e). After the quaternary arm forms, roughly the entire 2π extent of the disk is populated with spirals, so it is unlikely that any more well-defined arms would form.[8] Nonetheless, we emphasize that the tertiary and higher-order arms have very small amplitudes relative to the primary and secondary even at $r = 0.05r_p$ (see Figure 7(h)) and thus are relatively unimportant features.

As in the outer disk, the primary trough carries a significant portion of the total AMF (see Figures 7(j)–(l) and 8(g) and (h)) and typically carries more AMF than the primary arm at $r \lesssim (0.5–0.8)r_p$, which is a larger radius than the one at which the secondary arm forms or the secondary arm amplitude becomes stronger than the primary (in terms of amplitude). The AMF of the secondary arm exceeds the primary-arm AMF for $r \lesssim (0.4–0.5)r_p$. Far inside the inner disk ($r \approx 0.1r_p$), the AMF is primarily carried by the secondary arm, secondary trough,

and primary trough, while contributions from the other peaks and troughs are negligible.

3.6. Dependence on Disk Parameters

Figures 2(a) and (c) show that lowering h_p results in more tightly wound spiral arms. For such colder and thinner disks, the secondary arm (in the inner disk) forms closer to the primary (in azimuthal angle), with a smaller initial width and smaller initial arm-to-arm separation, and it also becomes stronger than the primary at a larger radius; see Figures 7(a), (d), (g), and (j).

For temperature profiles with q different from unity, Figures 2(a) and (b) show that the pitch angles of the spirals vary with radius. For $q < 1$, the initial properties of the secondary arm are relatively unchanged, but its amplitude becomes larger than the primary amplitude at larger radii (similar to the case of lower h_p); see Figure 8. Additional (tertiary and beyond) arms also form closer to the planet for thinner disks, and more arms are formed—as many as five for $h_p = 0.05$ (however, the full azimuthal extent of the inner disk is not fully populated with spirals in this case, as it is for larger aspect ratios).

The development of the secondary arm in the inner disk is summarized in Figure 9 for different disk parameters. It shows the radial location at which the amplitude of the secondary arm, $\delta\Sigma_S$, relative to the amplitude of the primary, $\delta\Sigma_P$, is equal to 0.2, 0.5, and 1. In each panel, two of the three disk parameters (h_p, q, and p) are held fixed, while the third one is varied. These properties depend most strongly on h_p, with the arm forming

---

[8] Although, since the arms become more narrow toward the inner disk, there may possibly be room for additional arms at even smaller radii.
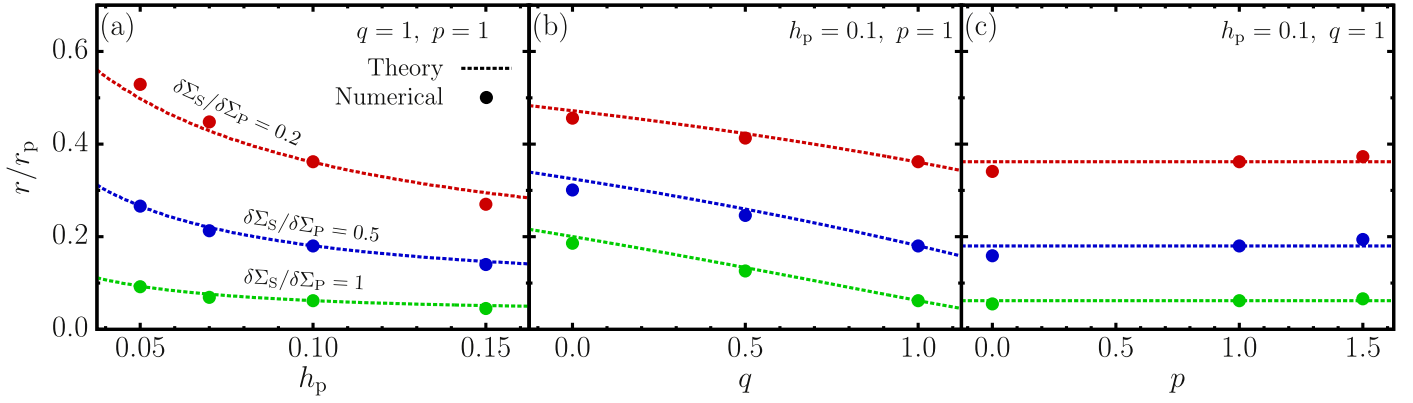
Figure 9. Radii at which the amplitude of the surface density perturbation associated with the secondary spiral arm, $\delta\Sigma_S$, relative to the amplitude of the primary arm, $\delta\Sigma_P$, is equal to 0.2 (red points), 0.5 (blue points), and 1 (green points). In each panel, two of three disk parameters (aspect ratio h_p, temperature power-law index q, and surface density power-law index p) are fixed at their fiducial values, while the value of the third parameter is varied. The dashed curves indicate the theoretical predictions given by Equations (38) and (39), calibrated to the fiducial disk parameters (i.e., using $h_p = 0.1$).

closer to the planet for smaller h_p. For values of q smaller than the fiducial $q = 1$, i.e., for more flared disks, the secondary arm forms somewhat closer to the planet. However, its amplitude grows with radius faster than the case with $q = 1$, reaching 50% and 100% of the primary amplitude at significantly larger radii.

The development of the secondary spiral is relatively insensitive to the slope of the surface density profile for a plausible range of values ($0 < p < 3/2$). This is evident in the structure of the 2D surface density maps in Figures 2 and 3, which show the cases $p = 1$ and 0 (note that the overall scaling of the amplitude of the perturbations with r differs between the two cases, but this is trivially described by AMF conservation; see Equation (16)). This is shown quantitatively in Figure 9(c). The radius at which the secondary spiral forms differs by only about 5% between the cases $p = 0$ and 3/2.

4. Results for Passively Propagating Spiral Waves

As will be shown in Section 5, the formation of the secondary spiral can be described fairly well by the local (WKB) approximation, which is valid far from resonances, where multiple spiral arms emerge and the gravitational influence of the planet is negligible. This indicates that the external potential of the planet plays a minimal role in this process. Therefore, motivated by the results of Arzamasskiy & Rafikov (2018), we may expect that any spiral wake formed by the phase coherence of a series of modes launched in a companion-free disk will also break up into multiple spirals.

To this end, we investigate the passive linear propagation of a spiral wake (not subject to the external potential of a planet). More specifically, we perform a calculation representing a linear analog of the nonlinear calculation of Arzamasskiy & Rafikov (2018), closely following their setup. In this setup, a surface density perturbation $\delta\Sigma(\phi)$ rotating with a pattern frequency ω_p is imposed at a radius r_{out} and allowed to propagate toward the disk center (only inward propagation is allowed provided that $\omega_p < \Omega_{out}$).

We solve for the structure of the modes of the homogeneous wave (Equation (6)) with $\Phi_m = 0$ by shooting solutions inward from r_{out} in order to satisfy the outgoing wave boundary condition at $r_{in} = 0.05 r_{out}$. For simplicity, here we choose a mode pattern frequency $\omega_p = 0$ (so that the wave pattern is fixed in the inertial frame), which corresponds to waves launched at a radius $\gg r_{out}$. Once the mode profiles are

obtained, any azimuthal profile $\delta\Sigma$ can be constructed from an appropriate linear combination of the different $\delta\Sigma_m$.

The perturbation imposed at r_{out} is chosen to be a Gaussian,[9]

$$\delta\Sigma(r_{out}) \propto \exp\left[-\frac{1}{2}\left(\frac{\phi}{\sigma}\right)^2\right]. \tag{23}$$

Thus, there are four parameters in our setup. These are the three disk parameters h_{out} and q that define the sound speed profile, now parameterized by $c_s(r) = h_{out} r_{out} \Omega_{out}(r/r_{out})^{-q/2}$, and the surface density power-law index p, as well as the initial azimuthal width σ of the surface density perturbation. For the fiducial parameters, we choose $h_{out} = 0.1$, $q = 1$, $p = 1$, and $\sigma = 0.06$. The fiducial σ is chosen to produce a surface density profile similar to the one produced near the planet for the case of the fiducial parameters in the inhomogeneous problem.

Figure 10 shows two-dimensional maps of the fractional surface density perturbation for the fiducial parameters and several cases with varied parameters. In all of these cases, the initial, single-peaked perturbation profile breaks up into multiple arms, in agreement with the findings of Arzamasskiy & Rafikov (2018). The azimuthal profiles of $\delta\Sigma$ at different radii are shown in Figure 11 for the fiducial parameters. They are qualitatively very similar to the profiles that develop in the inner disk in the case of the spiral wake driven by a planet; see Figure 4. The azimuthal distribution of the AMF similarly becomes divided up into several different peaks. We emphasize that the appearance of multiple spirals in this setup, in which there are no resonances in the disk, indicates that the formation of multiple spirals is not connected to resonances.

Figure 12 summarizes the dependence of secondary-arm amplitude on the disk parameters and initial width of the spiral wake. The dependence on the surface density profile index p is extremely weak. In fact, it is much weaker than in the case with a planet: the radius at which the secondary spiral emerges

---

[9] The actual profile imposed at r_{out} differs slightly from the form specified by Equation (23). Specifically, the actual profile is "missing" the $m = 1$ Fourier component. This is because the $m = 1$ mode is evanescent interior to its outer Lindblad resonance (OLR) (which is assumed to be located at $r \gg r_{out}$) and exterior to its inner Lindblad resonance (ILR) (which is formally located at $r = 0$). Therefore, the amplitude of the $m = 1$ mode is assumed to be zero at r_{out} as a result of its evanescent propagation from large r. Note that if we had instead chosen the amplitude of the $m = 1$ mode to be finite at r_{out} (so that $\delta\Sigma$ is given exactly by Equation (23)), it would decay away toward small r anyway, and it would not strongly affect our results.
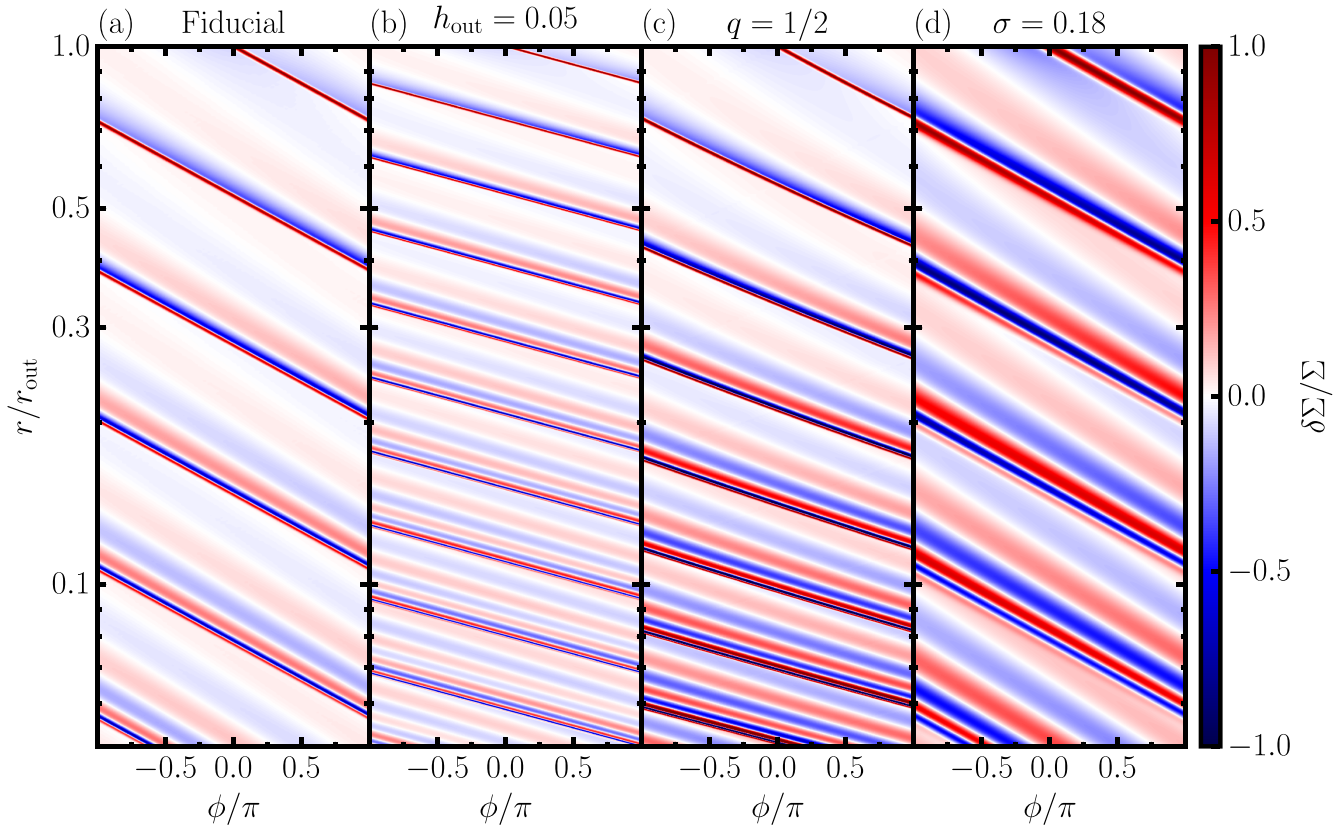
Figure 10. Two-dimensional fractional surface density perturbations for the propagation of a passive spiral wake with an initially specified profile at r_{out} (see Section 4). The left panel shows the case of the fiducial disk parameters $h_{out} = 0.1$, $q = 1$, $p = 1$, and $\sigma = 0.06$. In the other panels, the disk aspect ratio, temperature power-law index, and initial width are varied. In all cases, the single spiral arm imposed at r_{out} splits up into two or more arms as it propagates inward.

changes by less than 1% when the value of p is varied between zero and 1. This lack of sensitivity to the Σ profile is explained in Section 5.

On the other hand, the properties of the secondary spiral depend strongly on the initial width of the perturbation; see Figure 12(c). Wider initial perturbations lead to the development of the secondary spiral at larger radii, which is discussed in Section 5. In this experiment, we have isolated the effect of the initial perturbation width, as it is controlled by the parameter σ, which is unrelated to the disk parameters. In the case of a density wave launched by a planet, the width of the spiral wake near the planet is set by the disk aspect ratio, since the perturbation is dominated by modes with $m \sim 1/h_p$, so thinner disks produce narrower initial wakes. In the homogeneous case, by separating these two effects, we see that variation of the disk aspect ratio has an even stronger effect on the emergence of the secondary spiral than it does in the case with a planet: changing h_{out} by a factor of 3 changes the radius at which the secondary emerges by a factor of 2. Evidently, these two effects somewhat cancel out in the case with a planet.

5. Analytical Understanding of Our Results

In this section, we present theoretical arguments, following those made by Ogilvie & Lubow (2002) and Bae & Zhu (2018a), that interpret spiral arm formation as resulting from constructive interference among different azimuthal modes. We
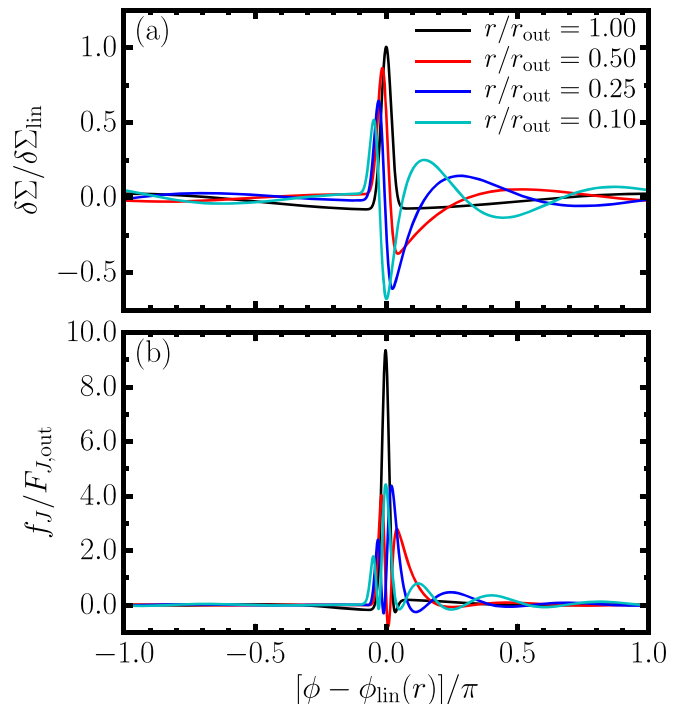

Figure 11. Profiles of the surface density perturbation (top panel) and azimuthal distribution of AMF (bottom panel) at different radii, as in Figure 4, but for the homogeneous problem of a passive spiral wake.
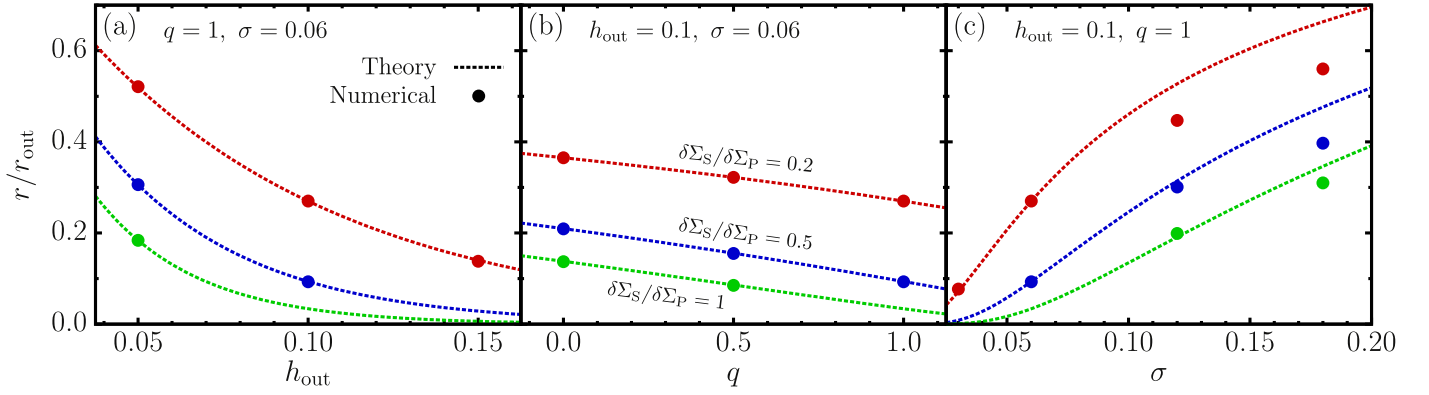
Figure 12. Radii at which the amplitude of the secondary arm reaches 0.2 (red points), 0.5 (blue points), and 1 (green points) relative to that of the primary arm, as in Figure 9, but for the homogeneous problem. In the different panels, the aspect ratio, temperature power-law index, and initial perturbation width are varied, while the other parameter values are held fixed. Note that "missing" points indicate that the secondary amplitude did not reach the specified threshold anywhere for $r > r_{in} = 0.05$ for the specified parameter value. The dashed curves show the theoretical predictions given by Equations (38) and (44), calibrated to the fiducial parameters.

then use these arguments to explain and interpret the results of our numerical calculations.

5.1. Mode Phases and Interference

The surface density perturbation of the mode with azimuthal number m is

$$\delta\Sigma_m(r, \phi) = |\delta\Sigma_m(r)| \exp[i\psi_m(r)], \quad (24)$$

where

$$\psi_m(r) = \text{Arg}[\delta\Sigma_m(r)] + m(\phi - \phi_p). \quad (25)$$

The phase of $\delta\Sigma_m$ can be expressed as (Ogilvie & Lubow 2002)

$$\text{Arg}[\delta\Sigma_m(r)] = \text{sgn}(r - r_p)\frac{\pi}{4} + \int_{r_\pm}^{r} k_m(r')dr', \quad (26)$$

where

$$r_\pm = \left(1 \pm \frac{1}{m}\right)^{2/3} r_p \quad (27)$$

is the location[10] of the outer (+)/inner (−) Lindblad resonance, and the $\pi/4$ term is the phase shift associated with the resonance. The radial number is given in the local (WKB) limit (and in the Keplerian limit $\kappa = \Omega$) by

$$k_m(r) = \frac{1}{c_s}[m^2(\Omega - \Omega_p)^2 - \Omega^2]^{1/2}. \quad (28)$$

Note that for $m \gg 1$, $k_m(r) \approx \tilde{k}_m(r)$, where

$$\tilde{k}_m(r) = \frac{m}{c_s}|\Omega - \Omega_p|, \quad (29)$$

which is simply proportional to m. Defining

$$\Delta k_m(r) = k_m(r) - \tilde{k}_m(r), \quad (30)$$

we can write

$$\psi_m = \text{sgn}(r - r_p)\frac{\pi}{4} + m(\phi - \phi_P)$$
$$+ \int_{r_\pm}^{r} \Delta k_m(r')dr' - \int_{r_p}^{r_\pm} \tilde{k}_m(r')dr'. \quad (31)$$

Here we have defined[11]

$$\phi_P = \phi_p - \int_{r_p}^{r} \tilde{k}_m(r')dr'$$
$$= \phi_p - \int_{r_p}^{r} \frac{|\Omega(r') - \Omega_p|}{c_s(r')}dr'. \quad (32)$$

Note the distinction between ϕ_p (with a lowercase subscript), the position of the planet, and ϕ_P (with an uppercase subscript), the position of the primary spiral arm. Also, in Equations (26), (31), and (32), note the different limits of integration.

The perturbation $\delta\Sigma_m$ has a maximum found by setting $\psi_m = 0$. For modes with $m \gg 1$, the last two terms in Equation (31) become negligible, so that the maxima of these modes have positions $\phi \approx \phi_P$. For general m, the maximum of $\delta\Sigma_m$ interferes constructively with these modes if

$$\int_{r_\pm}^{r} \Delta k_m(r')dr' - \int_{r_p}^{r_\pm} \tilde{k}_m(r') < \Delta\phi_0, \quad (33)$$

where $\Delta\phi_0$ is the maximum phase difference that results in constructive interference. In general, this criterion is indeed satisfied (for some range of r and values of m), resulting in the formation of the primary spiral arm. Therefore, ϕ_P gives the approximate position of the primary spiral (Ogilvie & Lubow 2002; Rafikov 2002a). To understand the origin of higher-order arms (secondary, tertiary, etc.), let us note that maxima of $\delta\Sigma_m$ are also attained at $\psi_m = 2\pi n$, where $n = 1, 2, ..., m - 2, m - 1$ is an integer. Maxima in this range of n are distinct in a sense that, at a given r, each of them corresponds to a well-defined azimuthal location $\phi = \phi_{m,n}$ (in a frame corotating with the planet), where

$$\phi_{m,n}(r) = \phi_p + \frac{1}{m}[-\text{Arg}(\delta\Sigma_m(r)) + 2\pi n]. \quad (34)$$

Values of n outside the interval $[1, m-1]$ yield $\phi_{m,n}$ coinciding with one of the azimuthal locations inside this interval.

---

[10] Note that this expression neglects $\mathcal{O}(h_p^2)$ terms due to radial pressure support and nonzero N_r.

[11] Note that ϕ_P is equivalent to ϕ_{lin} as defined in Equation (15). Here it has been redefined to emphasize that it gives a theoretical prediction for the azimuthal position of the primary spiral arm.
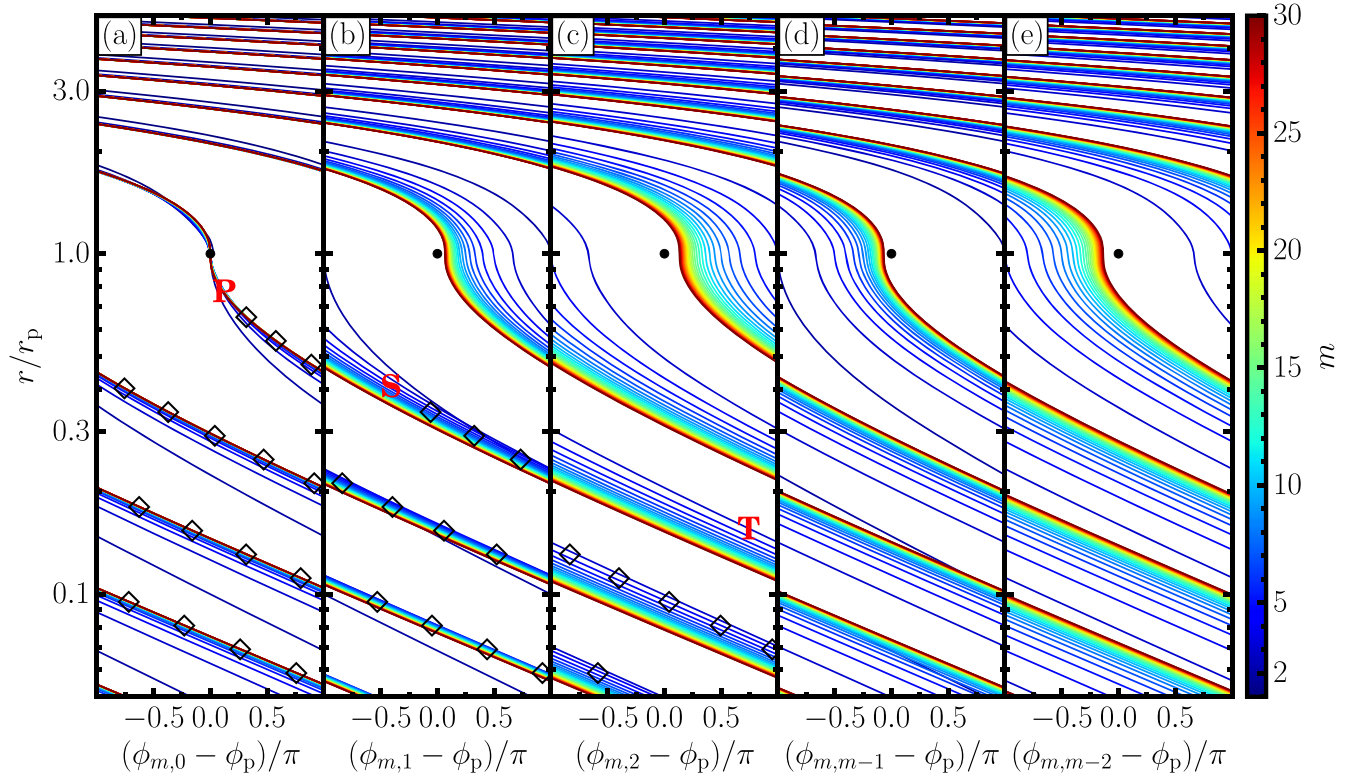
Figure 13. Phases of different surface density peaks associated with modes with different azimuthal numbers. Each $\phi_{m,n}$ corresponds to the nth peak associated with the mode of azimuthal number m. Constructive interference between peaks with different values of m but the same value of n may be responsible for the formation of the multiple spiral arms. The different colored curves correspond to different azimuthal numbers, and the different panels correspond to different values of n. In panels (a)–(c), the phases of the $n = 0$, 1, and 2 peaks, which are associated with the primary, secondary, and tertiary arms, respectively, are shown, along with diamonds indicating the actual location of these arms. In panels (d) and (e), the phases of peaks with $n = m - 1$ and $m - 2$ are shown. These peaks are not relevant to secondary spiral arm formation in the inner disk, as they are located behind, rather than in front of, the primary arm.

As pointed out by Bae & Zhu (2018a), these locations may define additional curves along which constructive interference occurs. In their formulation, a spiral arm may be formed as a result of constructive interference among the $\phi_{m,n}$ peaks with different values of m (but a fixed value of n). The primary spiral is a result of constructive interference of peaks with $n = 0$. In the inner disk, the secondary spiral is a result of interference of peaks with $n = 1$, the tertiary spiral is associated with $n = 2$, and so on. In the outer disk, $n = m - 1$ peaks are associated with the formation of a secondary spiral, $n = m - 2$ with a tertiary spiral, and so on.

The argument of Bae & Zhu (2018a) is based on the WKB approximation for the phases of the modes, and no account is given to the behavior of the mode amplitudes. In our calculations, the mode phases (as well as amplitudes) are computed exactly by solving for the global mode structure (i.e., not using WKB approximation). To illustrate how the mode interference idea works in our fully self-consistent calculation, in Figure 13, we show the phases of different crests of the modes defined by Equation (34), which were numerically computed using our full linear solutions. In agreement with Ogilvie & Lubow (2002), this figure indicates that the mode phases with $n = 0$ are tightly clustered in phase, and these phases closely follow that of the primary arm. But, in addition to that, one also sees that the phases with $n = 1$, which are initially very spread out (over a range of $\sim\pi$), become more clustered toward the inner disk. The phases of these modes follow the phase of the secondary spiral arm, indicating that the

constructive interference of $n = 1$ mode crests is indeed responsible for the emergence of the secondary arm.

The $n = 2$ phases also start out very spread out in azimuth at $r = r_{\rm p}$. They slowly converge as r decreases, although they still span a range of $\sim\pi/2$ even at $r = 0.05r_{\rm p}$. These phases are approximately coincident with that of the tertiary arm, although this correspondence is not as tight as it is for the cases of the primary and secondary arms. This highlights the fact that the phase information of the modes, while suggestive of the structure of the spiral arms, is not sufficient to fully capture their structure. Rather, a full consideration of the mode phases and amplitudes is required, which is done in this work.

None of the phases behind the $n = 0$ phases (i.e., $n = m - 1$, $n = m - 2$) are as tightly clustered as those with $n = 0$. Correspondingly, no spiral arm forms behind the primary in the inner disk (for the fiducial parameters).

5.2. Analytic Estimates

We now try to predict the radius at which the secondary arm emerges, as well as its location and width, using a phase interference argument. We do this by estimating the spread in the relevant $\phi_{m,n}$'s, and identifying the regions in which they are appropriately clustered as the locations of the different spiral arms. In order to do this, we make use of the WKB approximation, which we find reproduces the mode phases ($\phi_{m,n}$'s) found in our numerical calculations with reasonable accuracy. For modes with m close to $m_* \approx (2h_{\rm p})^{-1}$, the error in the WKB phases is $\sim20\%$ in the vicinity of the Lindblad

resonances and an order of magnitude smaller far from the resonances, where secondary spiral arm formation occurs.

From Equations (26) and (34), we have

$$\phi_{m,n} = \phi_{\mathrm{p}} - \mathrm{sgn}(r - r_{\mathrm{p}})\frac{\pi}{4m} + \frac{2\pi n}{m} - \int_{r_\pm}^{r} \frac{k_m(r')}{m}dr' \quad (35)$$

(see Bae & Zhu 2018a). The planet-induced density wake is dominated by modes with $m \approx m_*$. Therefore, defining $\phi_n \equiv \phi_{m_*,n}$, we follow the ansatz of Bae & Zhu (2018a) and identify in the inner disk the position of the secondary spiral as $\phi_{\mathrm{S}} = \phi_1$, the position of the tertiary spiral as $\phi_{\mathrm{T}} = \phi_2$, and so on. The positions of the secondary and higher-order spiral arms relative to the primary arm are then given approximately by

$$\phi_n - \phi_{\mathrm{P}} = -\mathrm{sign}(r - r_{\mathrm{p}})\frac{\pi}{4m_*} + \frac{2\pi n}{m_*}$$
$$- \int_{r_\pm}^{r} \frac{\Delta k_{m_*}(r')}{m_*}dr' + \int_{r_{\mathrm{p}}}^{r_\pm} \frac{\tilde{k}_{m_*}(r')}{m_*}dr'; \quad (36)$$

see Equations (32) and (35).

However, in order for a spiral arm to exist at $\phi = \phi_n$, the appropriate $\phi_{m,n}$'s must be sufficiently clustered. The spread in $\phi_{m,n}$ for values of m in the vicinity of m_* is

$$\delta\phi_{m,n} = \left| \left(\frac{\partial \phi_{m,n}}{\partial m} \right)_{m_*} \right| \Delta m, \quad (37)$$

where Δm represents the range of azimuthal numbers that contribute to the arm. We assume that this range is comparable to the critical m itself, taking $\Delta m = \zeta m_*$, with $\zeta \sim 1$.[12] In order for the $\phi_{m,n}$ to constructively interfere and form a spiral arm, we require that $\delta\phi_{m,n} < \Delta\phi_0/m_*$ (where, as in Equation (33), $\Delta\phi_0$ is a critical separation required for constructive interference, so that the variation of ψ_m is less than $\Delta\phi_0$). Evaluating the derivative in Equation (37) and simplifying, we find the condition for spiral arm formation

$$\delta\phi_n \equiv \zeta^{-1}m_*\delta\phi_{m,n} < \Delta\tilde{\phi}_0, \quad (38)$$

where $\Delta\tilde{\phi}_0 = \Delta\phi_0\zeta^{-1}$ is a new "phase-spread" constant and[13]

$$\delta\phi_n = \left| 2\pi n - \mathrm{sign}(r - r_{\mathrm{p}})\frac{\pi}{4} + \int_{r_\pm}^{r} \frac{\Omega^2(r')dr'}{c_s^2(r')k_{m_*}(r')} \right|. \quad (39)$$

Note that $\delta\phi_n$ represents the spread in mode phases, while $\delta\phi_n/m_*$ gives the approximate spiral arm width (provided that a spiral arm exists). The integral in Equation (39) is positive (negative) in the outer (inner) disk. Therefore, a necessary condition for $\delta\phi_n$ to become small enough to form a spiral arm is that $n \leqslant 0$ ($n \geqslant 0$) in the outer (inner) disk. Hence, secondary

----

[12] This is a reasonable assumption for the narrow primary and secondary spirals. However, the broader widths of the tertiary and higher-order spiral arms may indicate that they are dominated by modes with a smaller range of m, indicating that $\zeta < 1$ may be appropriate for these spirals.

[13] In Equation (39), n should take on the values 0, ± 1, ± 2, etc. While the definition of $\phi_{m,n}$ (Equation (34)) also permits values such as $n = m - 1$, $m - 2$, …, note that, e.g., $\phi_{m,m-1} = \phi_{m,-1}$.

Table 1
Values of $\delta\phi_n$ at Locations Corresponding to Several Relative Amplitudes of the Secondary/Tertiary Spiral Arm (Where, for Example, $r_{\mathrm{S}50}$ Indicates a 50% Relative Amplitude of the Secondary to the Primary)

h_{p}	q	p	$\delta\phi_1(r_{\mathrm{S}20})$	$\delta\phi_1(r_{\mathrm{S}50})$	$\delta\phi_2(r_{\mathrm{T}20})$
0.05	1	1	2.46	0.62	5.75
0.07	1	1	2.39	0.53	5.92
0.10	1	1	2.27	0.62	5.70
0.15	1	1	2.05	0.52	5.08
0.10	1/2	1	2.19	0.44	5.59
0.10	0	1	2.12	0.28	5.50
0.10	1	0	2.12	0.33	5.43
0.10	1	3/2	2.36	0.78	5.85

and higher-order spirals can only be attributed to constructive interference of $\phi_{m,n}$ with $n > 0$ in the inner disk and $\phi_{m,n}$ with $n < 0$ (or, equivalently, $n = m - 1$, $m - 2$, etc.) in the outer disk, should spiral arms be present there (Bae & Zhu 2018a). Note that Equation (39) can be used to estimate the phase spread of the primary spiral arm by taking $n = 0$. From this, we see that the formation of the secondary spiral is coincident with the dissolution of the primary spiral, since $\delta\phi_0$ necessarily becomes large as $\delta\phi_1$ becomes small. This explains a trend seen in Figures 7(g)–(i) and 8(e) and (f): the decrease in $\delta\Sigma_{\mathrm{P}}$ toward the inner disk is accompanied by an increase in $\delta\Sigma_{\mathrm{S}}$.

Up to this point, we have left the value of $\Delta\tilde{\phi}_0$ arbitrary. Roughly speaking, we expect that taking $\Delta\tilde{\phi}_0 \approx \pi$ in Equation (38) should qualitatively predict the presence of a spiral arm (Ogilvie & Lubow 2002). In practice, we can use our full numerical results to calibrate this criterion. We compute values of $\Delta\tilde{\phi}_0$ corresponding to different relative strengths of the secondary (and tertiary) spiral arm. We do this by calculating, using Equation (39), for example, $\delta\phi_1(r_{\mathrm{S}20})$, where $r_{\mathrm{S}20}$ denotes the location at which $\delta\Sigma_{\mathrm{S}}/\delta\Sigma_{\mathrm{P}} = 20\%$, inferred from our numerical results. These values are given in Table 1.

5.2.1. Passive Spirals

The same formalism can also be applied to our numerical experiments for passively propagating spirals presented in Section 4. In this setup, we assume that all modes are exactly in phase at r_{out}. We also assume that the (inner) Lindblad resonances for all modes are located exterior to r_{out}. Therefore, in Equation (26), we drop the $\pi/4$ term (associated with the resonance) and take r_{out} as the lower limit of the k_m integral. Therefore, we have

$$\phi_{m,n} = \frac{2\pi n}{m} - \int_{r_{\mathrm{out}}}^{r} \frac{k_m(r')}{m}dr'. \quad (40)$$

In the expression for the radial wavenumber (Equation (28)), the orbital frequency of the planet Ω_{p} should be replaced by the specified pattern frequency ω_{p}. Additionally, the initial width σ of the spiral launched at r_{out} is explicitly specified rather than being set by the disk aspect ratio, as it was for a wake launched by a planet. Therefore, we take $m_* \approx (2\sigma)^{-1}$ as an approximation of the dominant azimuthal mode number in the homogeneous case. We then have, for the arm-to-arm separations and phase spreads,

$$\phi_n - \phi_{\mathrm{P}} = \frac{2\pi n}{m_*} - \int_{r_{\mathrm{out}}}^{r} \frac{\Delta k_{m_*}(r')}{m_*}dr' \quad (41)$$

and

$$\delta\phi_n = \left| 2\pi n + \int_{r_{\mathrm{out}}}^{r} \frac{\Omega^2(r')dr'}{c_s^2(r')k_{m_*}(r')} \right|. \tag{42}$$

For the case $\omega_{\mathrm{p}} = 0$, as adopted in Section 4, the expressions for $\phi_n - \phi_{\mathrm{P}}$ and $\delta\phi_n$ take on simple forms,

$$\phi_n - \phi_{\mathrm{P}} = \frac{2\pi n}{m_*} + \left[1 - \frac{(m_*^2 - 1)^{1/2}}{m_*} \right] \frac{g(r/r_{\mathrm{out}})}{h_{\mathrm{out}}} \tag{43}$$

and

$$\delta\phi_n = \left| 2\pi n + \frac{g(r/r_{\mathrm{out}})}{h_{\mathrm{out}}(m_*^2 - 1)^{1/2}} \right|, \tag{44}$$

where

$$g(x) = \begin{cases} \dfrac{2}{q-1}[x^{(q-1)/2} - 1] & (q \neq 1), \\ \ln(x) & (q = 1). \end{cases} \tag{45}$$

5.2.2. Inner/Outer Disk Asymmetry

Our numerical results for planet-driven spirals indicate that multiple spirals robustly form in the inner disk, while only a single spiral forms in the outer disk (except for small h_{p}, for which a weak secondary spiral is present). Evidently, this asymmetry is related to the behavior of the integral in Equation (39) for small versus large r.

In the outer disk, the integral remains bounded. For $q = 1$, its value for $r \to \infty$ is exactly

$$\int_{r_+}^{\infty} \frac{\Omega^2(r')dr'}{c_s^2(r')k_m(r')} = -\frac{2}{3h_{\mathrm{p}}} \frac{\ln[m - (m^2 - 1)^{1/2}]}{(m^2 - 1)^{1/2}}. \tag{46}$$

For $q < 1$, the value of the integral is even less than given above, since the integrand is proportional to $r^{(q-6)/2}$ for large r. Therefore, for $q = 1$ and setting $m \approx (2h_{\mathrm{p}})^{-1}$, we have

$$\delta\phi_n(\infty) \approx \left| 2\pi n - \frac{\pi}{4} - \frac{4}{3}\ln(h_{\mathrm{p}}) \right|. \tag{47}$$

Consider $\delta\phi_n$ with $n = -1$. If this quantity becomes sufficiently small in the outer disk, a secondary spiral may form. For the fiducial parameters ($h_{\mathrm{p}} = 0.1$), $\delta\phi_{-1}(\infty) \approx 3.95$. Since this is larger than the critical phase spread necessary for secondary spiral arm formation given in Table 1, no such spiral forms in the outer disk.

However, Equation (47) also indicates that $\delta\phi_{-1}$ may become small enough to produce a secondary spiral if h_{p} is small enough. Indeed, in our numerical calculations, a weak secondary spiral is found for $h_{\mathrm{p}} = 0.05$. But the weak (logarithmic) dependence on h_{p} in Equation (47) indicates that $\delta\phi_{-1}$ can become small enough to produce a strong secondary spiral only if h_{p} is very small.

On the other hand, in the inner disk, for $r \ll r_{\mathrm{p}}$, one can show that

$$\int_{r_-}^{r} \frac{\Omega^2(r')dr'}{c_s^2(r')k_m(r')} \approx \frac{g(r/r_{\mathrm{p}}) + \eta}{(m^2 - 1)^{1/2}h_{\mathrm{p}}}, \tag{48}$$

where η is an order unity constant. For $q \leqslant 1$, $g(r/r_{\mathrm{p}}) \gg 1$ for small r, so $\delta\phi_n$ (Equation (39)) must cross zero (indicating constructive interference) before diverging as $r \to 0$. In other

words, it is always possible to find a sufficiently small r such that $\delta\phi_n < \Delta\phi_0$ ($n \geqslant 1$) for any $\Delta\phi_0$, so secondary arm formation is unavoidable in the inner disk.

5.3. Application of Theory to Our Numerical Results

We now apply the analytical results derived above to understanding certain features of the linear calculations presented in Sections 3 and 4. In Figures 14(a) and (b), we display the theoretical spiral arm phase spreads given by Equation (39) for two different values of h_{p}. The critical radii $r_{\mathrm{S}20}$, $r_{\mathrm{S}50}$, and $r_{\mathrm{T}20}$ found in our linear calculations are also indicated. We see that the values of the relevant $\delta\phi_n$ at these critical locations are relatively insensitive to h_{p} (as well as to the other disk parameters; see Table 1). Therefore, Equation (38) with properly calibrated $\Delta\tilde{\phi}_0$ can be used to reliably predict the location at which the secondary spiral forms. The phase spreads estimated from the numerical spiral arm widths (FWHM for the nth arm, FWHM$_n$), $\delta\phi_n = m_* \times$ FWHM$_n$, which can be compared to the theoretical $\delta\phi_n$, are also shown. The theoretical and numerical spiral arm phase spreads show qualitative agreement. They both vary similarly with r, although the exact values differ. This may be due in part to the ambiguity in quantifying the width of a spiral arm found in the numerical calculation (e.g., using the FWHM instead of some other metric). Figures 14(c) and (d) show the theoretical arm separations (Equation (36)), as well as the arm separations from our numerical results. The theoretical calculation accurately predicts the azimuthal separation between the primary and secondary spirals but only roughly predicts the separation of the tertiary and primary spirals. This discrepancy is possibly due to the fact that the tertiary spiral is not dominated by modes with $m \approx m_*$, as assumed in Equation (36), but rather by modes with $m \lesssim m_*$.

The theoretical predictions of the critical radii for secondary spiral formation (for the case of planet-driven spirals) obtained using Equations (38) and (39) are shown in Figure 9. These predictions are calibrated using the values of $\Delta\tilde{\phi}_0$ found for the fiducial parameters (i.e., for $h_{\mathrm{p}} = 0.1$). That is, the critical values of $\delta\phi_1$ corresponding to different secondary/primary amplitude ratios given in Table 1 for the fiducial parameters are taken to be universal values applicable for all parameters. These predictions based on simple phase arguments give excellent agreement with the numerical results over a range of disk parameters.

The equivalent predictions for the case of passive spirals are shown in Figure 12. These predictions also agree well with the numerical results. There are, however, some discrepancies for wide spirals (Figure 12(c)). This may be due to the fact that wide spirals are dominated by a small number of azimuthal modes, diminishing the accuracy of the approximation (Equation (37)) for the spread in $\phi_{m,n}$.

Given the success of our analytic arguments in explaining some key features of our numerical calculations, we can use them to interpret some features of secondary and higher-order spiral arms found in this work. The weak dependence of the characteristics of multiple arms on the surface density slope p found in both the inhomogeneous (Section 3) and homogeneous (Section 4) cases is easily understood in the context of mode interference. We showed that this process is well described by the WKB approximation, in which the mode phases are independent of the Σ profile. Rather, they depend
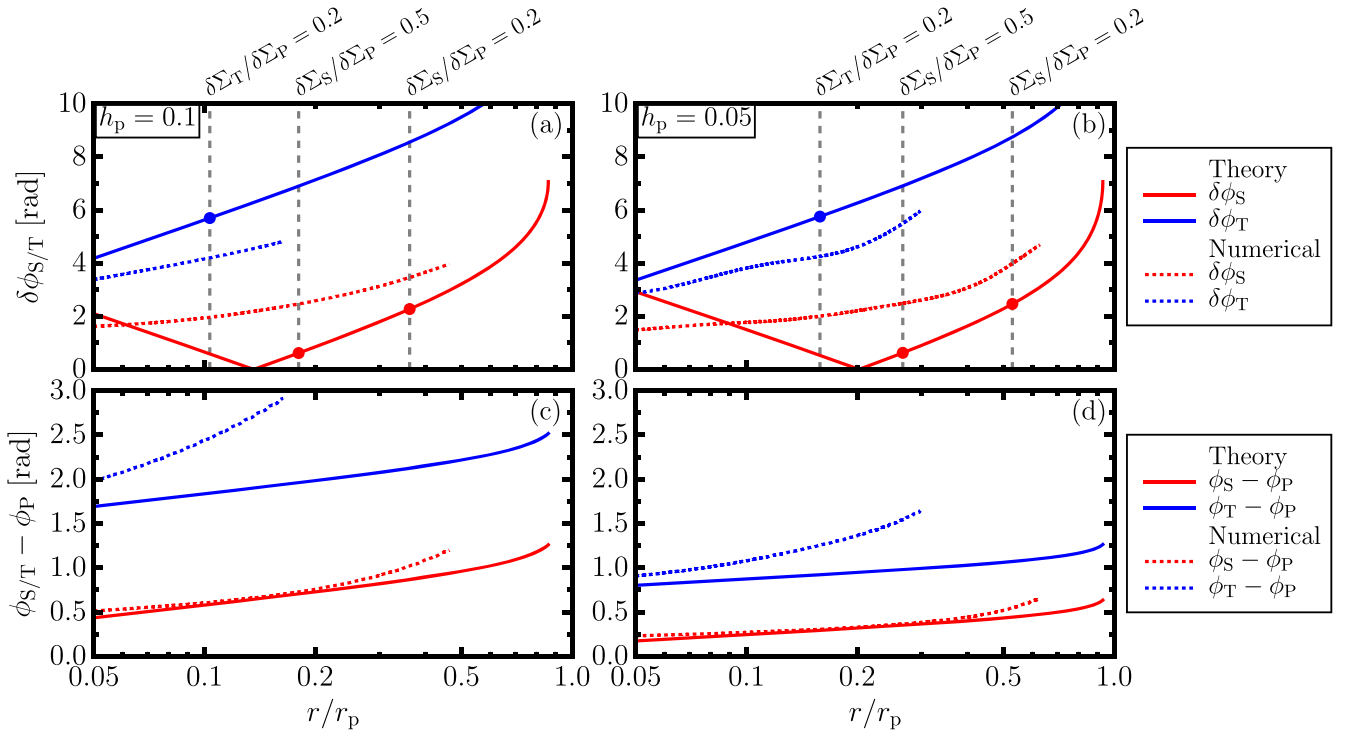
Figure 14. Secondary/tertiary phase spreads (Equation (39); top panels) and positions (Equation (36); bottom panels) predicted by our theoretical phase argument (solid curves) and estimated using numerical calculations (dotted curves). Two different cases are shown, $h_p = 0.1$ (left) and $h_p = 0.05$ (right), both with $q = p = 1$. In the top panels, the dashed vertical lines indicate the radii corresponding to several critical relative amplitudes of the secondary/tertiary spirals. The values of the theoretical phase spreads at these critical radii, highlighted by the filled points, are insensitive to the disk parameters. Note that the "reflection" feature of the theoretical $\delta\phi_S$ in panels (a) and (b) corresponds to a sign change of the quantity inside the absolute value symbol in Equation (39). This feature is not seen in the numerical $\delta\phi_S$, possibly due to the breakdown of the expansion (Equation (37)) when the spread in the mode phases is small, as a result of the discreteness of m.

only on $h(r)$, which in our parameterization is fully described by the parameters h_p and q.

The weak remaining dependence on p in the planet-driven case (see Figure 9(c)) is likely due to the shift of the Lindblad resonances at which the modes are launched. In the homogeneous case, our choice of $\omega_p = 0$ effectively places Lindblad resonances far outside the computation domain suppressing any sensitivity to p, as stated in Section 4.

We can also explain why, in the homogeneous case, (azimuthally) narrower patterns of $\delta\Sigma$ produce secondary arms at smaller radii; see Figure 12(c). Narrower perturbations have their power concentrated in modes of higher m (i.e., they have larger values of $m_* \approx (2\sigma)^{-1}$) compared to azimuthally wider perturbations. Equation (44) then predicts that $\delta\phi_n$ becomes small (resulting in phase coherence for $n = 1, 2$, and so on) at lower r/r_{out} as σ is decreased (and m_* is increased correspondingly). Note that $g(x) < 0$ in the inner disk; see Equation (45). Therefore, narrower perturbations produce high-order spiral arms at smaller radii than the wider ones do.

6. Discussion

The calculations presented in this work (except for Section 3.4) are explicitly linear. At the same time, it is well known that nonlinear effects play an important role in the propagation and damping of spiral waves, as well as the evolution of the disk (e.g., Goodman & Rafikov 2001; Rafikov 2002a, 2016; Arzamasskiy & Rafikov 2018). A planet-driven spiral wake begins to shock at a distance of $L_{sh} \sim (M_p/M_{th})^{-2/5}H_p$ from the planet (in the local approximation), evolving into a wide "N"-shaped wave (Goodman & Rafikov 2001). In the process, it deposits its angular momentum into the disk material, so that the AMF of the

wave is no longer conserved (as it is in linear theory). This process is modified by the presence of a secondary spiral (Arzamasskiy & Rafikov 2018). Injection of angular momentum originally carried by the spiral wave into the disk material drives the evolution of the disk (Arzamasskiy & Rafikov 2018) and causes gap opening (Rafikov 2002b).

Our linear calculations are strictly valid only as long as the appearance of the secondary spiral is not preceded by the shocking of the primary arm. This condition sets an upper limit on the allowed planet mass. Indeed, if the secondary spiral emerges after the wake travels a distance ζH_p in the inner disk, then the condition $L_{sh} \gtrsim \zeta H_p$ implies that $M_p \lesssim M_{th}\zeta^{-5/2}$. In our $h_p = 0.1$ calculation, the secondary arm forms at $\approx 5H_p$ interior to the planet, meaning that $\zeta \approx 5$ and $M_p \lesssim 0.02M_{th}$ are needed for our linear calculation to capture the formation of multiple spirals in quantitative detail. However, at the qualitative level, our calculation should remain valid at substantially higher values of M_p (e.g., because nonlinear evolution has only a marginal effect on the analytical phase coherence calculation presented in Section 5).

Nonlinear evolution also affects the morphology of the spirals in the high-M_p regime. Numerical simulations have shown that the azimuthal separation of the secondary and primary arms is $\approx 60°$ for low-mass ($M_p \ll M_{th}$) planets, in agreement with our prediction from linear theory, but increases up to $\approx 180°$ for massive ($M_p \gg M_{th}$) planets (Dong et al. 2015b; Fung & Dong 2015). This transition is caused by the steady azimuthal broadening of the spiral wake due to its nonlinear evolution in the "N-wave" regime (Goodman & Rafikov 2001; Rafikov 2002a; Zhu et al. 2015). Therefore, the secondary spirals that we find to form by linear processes

should be regarded as just precursors to the fully fledged secondary spiral arms/shocks (see Section 6 of Arzamasskiy & Rafikov [2018](2018)). In our linear calculations, the tertiary/quaternary arms are always very weak, so it is unclear how they are affected by nonlinear effects. However, note that Dong & Fung ([2017](2017)) reported the presence of these higher-order arms in nonlinear simulations, although they found them to be destroyed by moderate viscosity.

We note that the global treatment, i.e., accounting for the cylindrical geometry (as opposed to the local, shearing sheet approximation), is critical for capturing the formation of multiple spiral arms. We find that the distance from the planet at which the secondary arm forms (which we have defined as the point at which its amplitude is 10% of the primary), measured in terms of H_p, is a decreasing function of h_p: the secondary arm forms at a distance of $\approx 5H_p$ from the planet for $h_p = 0.1$ and $\approx 7.5H$ from the planet for $h_p = 0.05$. In the local (shearing sheet/box) approximation, corresponding to the limit $h_p \to 0$, in which H_p is the characteristic length scale, this implies that secondary spirals form at $|r - r_{pl}|/H_p \to \infty$. In other words, higher-order spirals would not be captured in the shearing sheet approximation. Furthermore, secondary spirals form almost exclusively in the inner disk, not in the outer disk, an asymmetry that cannot arise in the shearing sheet framework.

6.1. Comparison with Other Work

The numerical calculations presented in this paper largely follow those of Ogilvie & Lubow ([2002](2002)), although we give a more detailed analysis of the results. However, Ogilvie & Lubow ([2002](2002)) did not report the presence of multiple spirals in their calculations. We find two main reasons for this. First, they only solved for the perturbation structure down to a radius of $0.3r_p$ (unlike our calculations, which extend to $0.05r_p$, allowing the secondary arm to be fully captured). From our Figure [7](7), we see that for our fiducial parameters, the secondary arm is still quite weak at that radius, with an amplitude about four times smaller than that of the primary arm. In their Figure 5, the first hint of a secondary arm becomes visible near the inner disk edge; however, it went unnoticed in their discussion. Second, the calculations by Ogilvie & Lubow ([2002](2002)) were restricted to the case of a constant disk aspect ratio, i.e., $q = 1$ in our notation. For flared disks with $q < 1$, the secondary arm emerges and overtakes the primary in amplitude closer to the planet (see Figures [8](8) and [9](9)), making its presence more apparent. Finally, we note that Ogilvie & Lubow ([2002](2002)) did point out that the phases ψ_m of modes with different m eventually diverge (logarithmically for the case $q = 1$) toward the inner disk so that their constructive interference fails, resulting in the partial dissolution of the primary arm. However, they missed the fact that the same process also results in convergence of ψ_m to an integer multiple of 2π (implying the same value of $\delta\Sigma_m$) at different azimuthal locations in the disk, giving rise to higher-order spirals (see Section [5](5)).

Rafikov ([2002a](2002a)) arrived at the one-armed spiral solution using a method different from that of Ogilvie & Lubow ([2002](2002)). In his case, the inability to capture the formation of higher-order arms is likely caused by a certain assumption used in the derivation of the linear wake shape, namely, the conservation of the Riemann invariant along the characteristics that cross (rather than follow) the wake. Small changes of this invariant at

the wake crossings, neglected in Goodman & Rafikov ([2001](2001)) and Rafikov ([2002a](2002a)), could be responsible for the eventual emergence of the secondary spiral. This conjecture is supported by the fact that the secondary spiral emerges closer to the planet in disks with lower h_p: the number of wake crossings by characteristics (per fixed radial interval) grows as h_p goes down, facilitating the breakdown of the one-spiral solution.

Some other ideas for the origin of secondary spirals have been advanced, in particular, nonlinear effects related to ultraharmonic resonances with the planet (Fung & Dong [2015](2015)). We do not find these explanations persuasive. First, we reproduce multiple spirals in the framework of a purely linear calculation. Second, our calculations of passive propagation of a wake with $\omega_p = 0$ in Section [4](4) do not feature any resonances, yet they do result in secondary spirals.

6.2. Applications

A possible connection between the multiple spirals driven by a planet and the multiple gaps and rings seen in some protoplanetary disks was suggested by Dong et al. ([2017](2017), [2018](2018)) and Bae et al. ([2017](2017)). In the picture put forth by these authors, a low-mass planet in a low-viscosity disk produces multiple spiral arms, each of which shocks, dissipates, and opens a gap at some distance from the planet. The location of the secondary gap, attributed to the dissipation of the secondary spiral, was given by Dong et al. ([2018](2018)) as a function of planet mass and disk thickness. Our linear prediction for the location at which the secondary spiral forms is exterior to their predicted gap location for small planet masses ($M_p \lesssim 0.2M_{th}$). This is consistent with the scenario in which the secondary spiral first forms in a linear fashion at some distance from the planet and then propagates inward, evolving nonlinearly, before shocking and opening a gap. For larger planet masses, Dong et al. ([2018](2018)) predicted a secondary gap at a location too close to the planet for a secondary spiral to have formed in linear theory. In this case, nonlinear effects clearly play a role not just in the dissipation of the spiral but also in its formation.

The multiple spiral features observed in some protoplanetary disks may be produced by planets (Dong et al. [2015b](2015b)). The 180° separation of these spirals requires massive planets to produce, so nonlinear effects cannot be neglected in these cases. We nonetheless expect the linear mechanism described in this work to play an important role in providing the conditions necessary for the formation of these structures (Arzamasskiy & Rafikov [2018](2018)).

However, planets are not the only possibility. As we showed in Section [4](4), any spiral arm, regardless of its origin, inevitably evolves into multiple spirals as it propagates through the differentially rotating disk (note that the assumption of a Keplerian profile for $\Omega(r)$ is not essential for the arguments advanced in Section [5](5)). Therefore, any mechanism capable of producing at least one spiral arm necessarily produces multiple spiral arms. Possible mechanisms include gravitational instability (e.g., Dong et al. [2015a](2015a)), accretion from an infalling envelope (Lesur et al. [2015](2015); Hennebelle et al. [2017](2017)), shadows/nonaxisymmetric illumination (Montesinos et al. [2016](2016)), and vortices (Paardekooper et al. [2010](2010)). We only require that density waves with a range of azimuthal mode numbers are excited and at least somewhat in phase with one another, so that one or a few well-defined spiral arms (rather than many flocculent spirals) are produced.

7. Summary

We explored the origin of multiple spiral arms, which are often observed in protoplanetary disks and found in numerical simulations of disks with massive perturbers. The two-dimensional structure of surface density perturbations induced by a planet (as well as that of a passive spiral) was computed using a linear theory of density wave excitation and propagation (Goldreich & Tremaine 1979) in the low planet mass (low-amplitude) regime.

We find that, in addition to the strong single spiral arm excited by the planet in agreement with past studies (Ogilvie & Lubow 2002; Rafikov 2002a), a secondary spiral arm (and often a tertiary arm, quaternary arm, and so on) robustly forms in the inner disk in the linear regime. The secondary arm first appears at about $r = (0.4–0.6)r_p$ and, though initially weak, becomes stronger and narrower toward the center of the disk, eventually exceeding the strength of the primary arm at $\approx 0.1r_p$. As the primary arm propagates into the inner disk, the AMF it carries gets steadily transferred to these higher-order spiral arms. In the outer disk, we find that a secondary spiral arm typically does not form, except for the coldest disk we considered, with $h_p = 0.05$.

We provide analytical arguments extending the reasoning of Ogilvie & Lubow (2002) that show that secondary spiral arms form as a result of the constructive interference among different azimuthal modes in the inner disk. Our treatment, which implicitly takes into account the global variation of both the phases and amplitudes of different linear modes in a self-consistent manner, thus corroborates the semiquantitative phase coherence picture previously put forth by Bae & Zhu (2018a) in the WKB limit. The gravitational potential of the planet does not play a role in this process. Rather, the planet only seeds the initial perturbation, which then propagates passively and spawns higher-order spiral arms. This is confirmed by the persistent emergence of multiple spirals in our linear calculations of the passive inward propagation of an imposed spiral wake, free from the influence of an external potential (following the setup of Arzamasskiy & Rafikov 2018).

Our results clearly demonstrate that the formation of secondary spirals is an intrinsically linear process that serves as a precursor for subsequent nonlinear evolution resulting in a formation of multiple shocks in the disk. These calculations should help us better understand the planet-driven evolution of protoplanetary disks (Goodman & Rafikov 2001; Rafikov 2002a), including the formation of multiple gaps in such disks (Bae et al. 2017; Dong et al. 2017, 2018). We use them to understand the details of the global distribution of the torque exerted by an embedded planet on a disk in R. Miranda & R. Rafikov (2019, in preparation).

Appendix
Numerical Procedure

A.1. Mode Solutions

Here we give a detailed description of the numerical method used for producing the solutions of Equation (6). We compute the values of the Laplace coefficients, as well as their derivatives required for Equation (6) (by first differentiating the integrand of Equation (13) with respect to α) using numerical quadrature. We remove the corotation pole ($\tilde{\omega} = 0$) in Equation (6) by replacing $\tilde{\omega}$ with $\tilde{\omega} + i\delta$, where δ is a small positive real constant (KP93). We choose $\delta = 10^{-6}$.

The solution method closely follows the technique described in detail by KP93, with several small differences. We first produce two linearly independent solutions, $\delta h_H^{(1)}$ and $\delta h_H^{(2)}$, of the homogeneous version of Equation (6) by integrating outward from corotation starting from arbitrary initial values, as well as an inhomogeneous solution, δh_{IH}. The desired solution, which satisfies any pair of specified boundary conditions at the boundaries r_{in} and r_{out}, can be expressed as

$$\delta h_m = a_1 \delta h_H^{(1)} + a_2 \delta h_H^{(2)} + \delta h_{IH}, \qquad (49)$$

where a_1 and a_2 are constants determined by the specific form of the boundary conditions. We choose outgoing wave boundary conditions with

$$\delta h_m'(r_{in}) = C_{in}\delta h_m(r_{in}), \ \delta h_m'(r_{out}) = C_{out}\delta h_m(r_{out}), \qquad (50)$$

where

$$C = ik + \frac{1}{2}\frac{d}{dr}\ln\left(\frac{D_S}{r\Sigma k}\right), \qquad (51)$$

representing an outgoing wave in the WKB limit (e.g., Tsang & Lai 2008). Here $k = (-D_S)^{1/2}/c_s$ is the radial wavenumber. Note that the second term in Equation (51), describing the slow change in amplitude of the wave, was not included by KP93 at this step (although they took it into account in an approximate fashion at a later step). Now that we have specified the boundary conditions, we have a system of equations that can be solved for a_1 and a_2:

$$\left[\frac{d}{dr}\delta h_H^{(1)}(r_{in}) - C_{in}\delta h_H^{(1)}(r_{in})\right]a_1$$
$$+ \left[\frac{d}{dr}\delta h_H^{(2)}(r_{in}) - C_{in}\delta h_H^{(2)}(r_{in})\right]a_2$$
$$= C_{in}\delta h_{IH}(r_{in}) - \frac{d}{dr}\delta h_{IH}(r_{in}), \qquad (52)$$

$$\left[\frac{d}{dr}\delta h_H^{(1)}(r_{out}) - C_{out}\delta h_H^{(1)}(r_{out})\right]a_1$$
$$+ \left[\frac{d}{dr}\delta h_H^{(2)}(r_{out}) - C_{out}\delta h_H^{(2)}(r_{out})\right]a_2$$
$$= C_{out}\delta h_{IH}(r_{out}) - \frac{d}{dr}\delta h_{IH}(r_{out}). \qquad (53)$$

As in KP93, once we have found the values of a_1 and a_2 and constructed a solution satisfying the boundary conditions, we generate a new inhomogeneous solution δh_{IH} using the "correct" values of δh and $\delta h'$ at corotation (i.e., the ones found from the previously constructed solution). The system of Equations (52) and (53) is then resolved for new values of a_1 and a_2. This process ensures that the final solution is robust with respect to the specific choice of (arbitrary) homogeneous solutions used in its construction (Equation (49)), because it tends to reduce the absolute values of a_1 and a_2 so that the contributions to the correct solution from the homogeneous solutions are small.
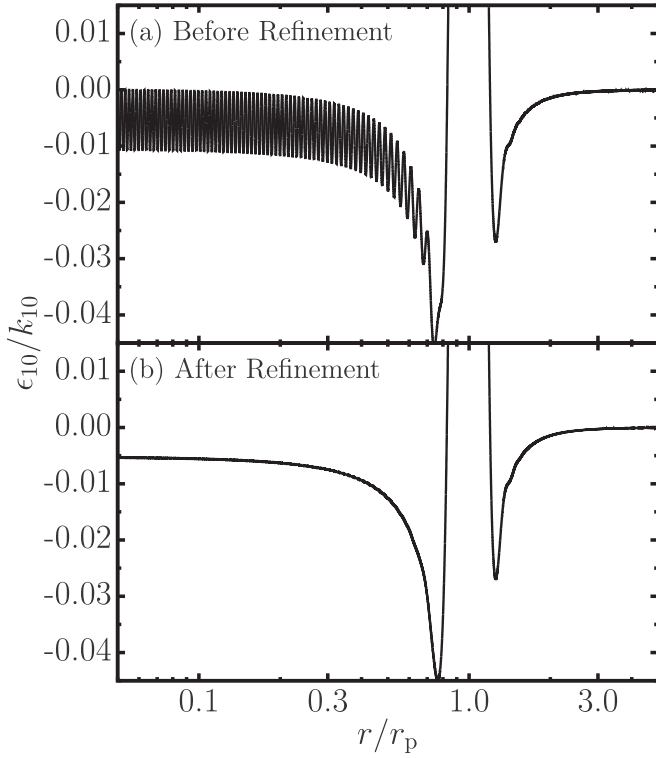
Figure 15. Fractional error in the phase gradient before and after refinement for the $m = 10$ mode, with the fiducial parameters $h_p = 0.1$, $q = 1$, and $p = 1$.

As a final refinement step, also described in KP93, we minimize the amplitude of the oscillations of the "phase gradient error" near the boundaries. The phase gradient error, defined as

$$\epsilon = \text{Arg}\,(\delta h)' - k, \qquad (54)$$

serves as a diagnostic of how close the solution is to an outgoing WKB wave. The phase gradient error of the initial solution exhibits oscillations, representing contamination by an incoming wave, near the boundaries. We seek to minimize the amplitude of these oscillations by slightly adjusting the constants C_{in} and C_{out} (so that they are no longer given exactly by Equation (51)) characterizing the boundary conditions. In principle, there are a variety of ways to achieve this (note that the exact procedure used by KP93 is not specified). We choose to minimize the peak-to-peak amplitudes of the phase gradient error ϵ within one local scale height of each boundary:

$$X_{in} = [\max(\epsilon) - \min(\epsilon)]_{r_{in} < r < r_{in} + H(r_{in})}, \qquad (55)$$

$$X_{out} = [\max(\epsilon) - \min(\epsilon)]_{r_{out} - H(r_{out}) < r < r_{out}}. \qquad (56)$$

Note that we also tried minimizing the derivative of the phase gradient error ϵ' at r_{in} and r_{out} but found that this was not as effective at getting rid of the oscillations. We numerically compute the Jacobian describing the derivatives of $\boldsymbol{X} = (X_{in}, X_{out})$ with respect to $\boldsymbol{C} = (C_{in}, C_{out})$ and use its inverse to perform one step of the secant method to find the root of $\boldsymbol{X}(\boldsymbol{C})$. This greatly reduces the amplitude of the oscillations of the phase gradient error; see Figure 15 for an example of this process. In practice, this refinement only requires changing the values of C_{in} and C_{out} by a very small amount ($\lesssim 1\%$). Also note that this refinement produces only very imperceptible changes in the form of $\delta h(r)$ and $\delta h'(r)$ but is potentially important for ensuring that different modes have the correct phase when the interference of many modes is considered.

As noted in Section 2, we do not include the indirect potential term $\delta_{m,1}GM_p r / r_p^2$ in our calculations. This term is proportional to r and so becomes large for $r \gg r_p$, in contrast to the direct terms (Equation (12)), which become small for $r \gg r_p$ (note that they both become small for $r \ll r_p$). Therefore, for $m \neq 1$, Equation (6) becomes effectively homogeneous for $r \gg r_p$. Our solution method exploits this fact by using knowledge of the asymptotic behavior of the homogeneous equation to set the outer disk boundary condition (and similarly for the inner boundary condition). However, for $m = 1$, when the indirect term is included, its anomalous behavior at large r calls into question the validity of setting the outer boundary condition in this way. We nonetheless carried out several tests in which the indirect potential was included and the outer boundary condition was set under the homogeneous assumption. We verified that including the indirect potential in this way only slightly modifies the profile of the density wake in the outer disk and does not otherwise affect our main results.

A.2. Mode Synthesis

The two-dimensional surface density perturbation $\delta\Sigma(r, \phi)$ is synthesized from the mode solutions δh_m using Equations (9) and (14). In order to produce an accurate solution, a sufficient number of modes (up to some m_{max}) must be used. The solution must be converged with respect to m_{max}; i.e., the perturbation structure should not change as more modes are added. The value m_{max} required for convergence of the two-dimensional surface density is several times larger than the cutoff parameter $m_{cut} \approx h_p^{-1}$. Figure 16 illustrates this point by revealing spurious features in the distribution of $\delta\Sigma(r, \phi)$ for

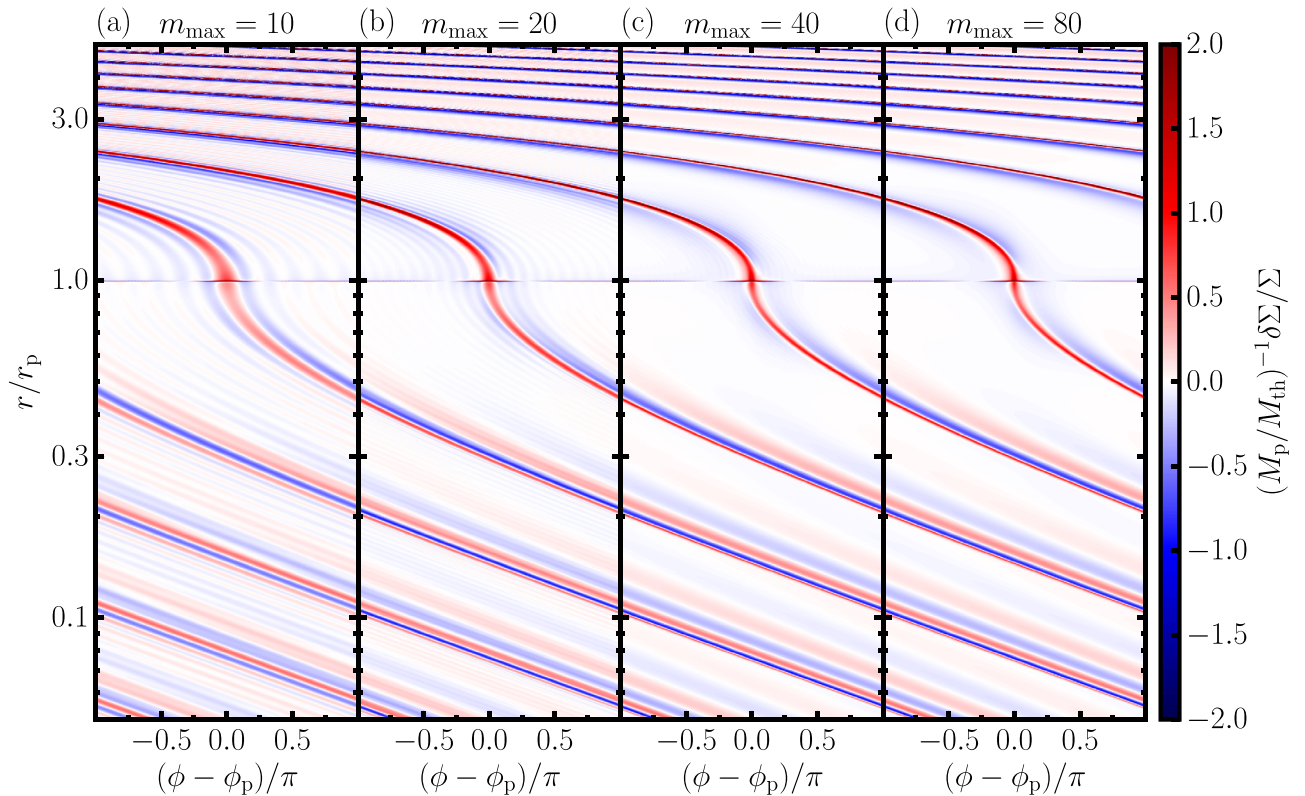**Figure 16.** Two-dimensional map in polar coordinates (r, ϕ) of the fractional surface density perturbation $\delta\Sigma/\Sigma$ (scaled by the ratio of the planet mass, M_p, to the thermal mass, $M_{th} = h_p^3 M_*$), produced using different values of m_{max}, the number of azimuthal modes synthesized. The case shown here uses the fiducial parameters.

low $m_{max} = (1-2)h_p^{-1}$ (panels (a) and (b)). Therefore, in all of our calculations, we choose $m_{max} \approx 8m_{cut}$, for which Figure 16(d) demonstrates convergence.

ORCID iDs

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

References

Arzamasskiy, L., & Rafikov, R. R. 2018, ApJ, 854, 84
Bae, J., & Zhu, Z. 2018a, ApJ, 859, 118
Bae, J., & Zhu, Z. 2018b, ApJ, 859, 119
Bae, J., Zhu, Z., & Hartmann, L. 2017, ApJ, 850, 201
Baruteau, C., & Masset, F. 2008, ApJ, 672, 1054
Benisty, M., Juhasz, A., Boccaletti, A., et al. 2015, A&A, 578, L6
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., & Fung, J. 2017, ApJ, 835, 38
Dong, R., Fung, J., & Chiang, E. 2016a, ApJ, 826, 75
Dong, R., Hall, C., Rice, K., & Chiang, E. 2015a, ApJL, 812, L32
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., Zhu, Z., Fung, J., et al. 2016b, ApJL, 816, L12
Dong, R., Zhu, Z., Rafikov, R. R., & Stone, J. M. 2015b, ApJL, 809, L5
Fung, J., & Dong, R. 2015, ApJL, 815, L21
Garufi, A., Quanz, S. P., Avenhaus, H., et al. 2013, A&A, 560, A105
Goldreich, P., Goodman, J., & Narayan, R. 1986, MNRAS, 221, 339
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
Grady, C. A., Muto, T., Hashimoto, J., et al. 2013, ApJ, 762, 48
Hennebelle, P., Lesur, G., & Fromang, S. 2017, A&A, 599, A86
Korycansky, D. G., & Pollack, J. B. 1993, Icar, 102, 150
Lee, W.-K. 2016, ApJ, 832, 166
Lesur, G., Hennebelle, P., & Fromang, S. 2015, A&A, 582, L9
Lin, M.-K. 2015, MNRAS, 448, 3806
Lin, M.-K., & Papaloizou, J. C. B. 2011, MNRAS, 415, 1445
Maire, A.-L., Stolker, T., Messina, S., et al. 2017, A&A, 601, A134
Montesinos, M., Perez, S., Casassus, S., et al. 2016, ApJL, 823, L8
Muto, T., Grady, C. A., Hashimoto, J., et al. 2012, ApJL, 748, L22
Ogilvie, G. I., & Lubow, S. H. 2002, MNRAS, 330, 950
Ostriker, E. C., Shu, F. H., & Adams, F. C. 1992, ApJ, 399, 192
Paardekooper, S.-J., Lesur, G., & Papaloizou, J. C. B. 2010, ApJ, 725, 146
Petrovich, C., & Rafikov, R. R. 2012, ApJ, 758, 33
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
Stolker, T., Dominik, C., Avenhaus, H., et al. 2016, A&A, 595, A113
Tsang, D. 2014, ApJ, 782, 112
Tsang, D., & Lai, D. 2008, MNRAS, 387, 446
Wagner, K., Apai, D., Kasper, M., & Robberto, M. 2015, ApJL, 813, L2
Wagner, K., Dong, R., Sheehan, P., et al. 2018, ApJ, 854, 130
Ward, W. R. 1997, Icar, 126, 261
Zhu, Z., Dong, R., Stone, J. M., & Rafikov, R. R. 2015, ApJ, 813, 88