5-30-2018

Planet-driven Spiral Arms in Protoplanetary Disks. I. Formation Mechanism

Jaehan Bae
Carnegie Institution for Science, jbae@carnegiescience.edu

Zhaohuan Zhu
University of Nevada, Las Vegas, zhaohuan.zhu@unlv.edu

CrossMark

Planet-driven Spiral Arms in Protoplanetary Disks. I. Formation Mechanism

Jaehan Bae[1,2,4] and Zhaohuan Zhu[3]
[1] Department of Terrestrial Magnetism, Carnegie Institution for Science, 5241 Broad Branch Road, NW, Washington, DC 20015, USA; jbae@carnegiescience.edu
[2] Department of Astronomy, University of Michigan, 1085 S. University Avenue, Ann Arbor, MI 48109, USA
[3] Department of Physics and Astronomy, University of Nevada, Las Vegas, 4505 South Maryland Parkway, Las Vegas, NV 89154, USA

Abstract

Protoplanetary disk simulations show that a single planet can excite more than one spiral arm, possibly explaining the recent observations of multiple spiral arms in some systems. In this paper, we explain the mechanism by which a planet excites multiple spiral arms in a protoplanetary disk. Contrary to previous speculations, the formation of both primary and additional arms can be understood as a linear process when the planet mass is sufficiently small. A planet resonantly interacts with epicyclic oscillations in the disk, launching spiral wave modes around the Lindblad resonances. When a set of wave modes is in phase, they can constructively interfere with each other and create a spiral arm. More than one spiral arm can form because such constructive interference can occur for different sets of wave modes, with the exact number and launching position of the spiral arms being dependent on the planet mass as well as the disk temperature profile. Nonlinear effects become increasingly important as the planet mass increases, resulting in spiral arms with stronger shocks and thus larger pitch angles. This is found to be common for both primary and additional arms. When a planet has a sufficiently large mass ($\gtrsim 3$ thermal masses for $(h/r)_p = 0.1$), only two spiral arms form interior to its orbit. The wave modes that would form a tertiary arm for smaller mass planets merge with the primary arm. Improvements in our understanding of the formation of spiral arms can provide crucial insights into the origin of observed spiral arms in protoplanetary disks.

Key words: hydrodynamics – planet–disk interactions – waves

1. Introduction

A point mass perturber creates disturbances in differentially rotating (e.g., Keplerian) background disks, which appear in the form of wakes spiraling away from the perturber. Such spiral structures are seen in observations and/or numerical simulations of a variety of astrophysical disks, including protoplanetary disks, circumplanetary disks, and disks around various binary systems (e.g., binary stars, binary black holes).

Interestingly, numerical simulations of protoplanetary disks show that one planet can launch multiple spiral arms (e.g., Kley 1999; Dong et al. 2015; Fung & Dong 2015; Juhász et al. 2015; Zhu et al. 2015; Bae et al. 2016a, 2016c, 2017; Lee 2016; Dong & Fung 2017; Hord et al. 2017; Richert et al. 2017). Possibly supporting the idea of a single planet launching multiple spiral arms, recent high-resolution observations revealed multi-armed spirals in protoplanetary disks (Muto et al. 2012; Garufi et al. 2013; Grady et al. 2013; Benisty et al. 2015; Pérez et al. 2016; Stolker et al. 2016; Maire et al. 2017; Tang et al. 2017; Reggiani et al. 2018), while the origin of observed spiral arms is yet to be confirmed.

As a spiral arm steepens into a shock, its angular momentum is transferred to the background disk gas, opening a gap (Rafikov 2002b). When a single planet excites multiple spiral arms, it can create multiple gaps and pressure bumps in between (Bae et al. 2017; Dong et al. 2017). The resulting structure in the gas disk can produce corresponding, enhanced features in dust emission and scattering, such that a single planet can be responsible for more than one main gap. This mechanism possibly explains some of the multiple concentric gaps and rings seen in young protoplanetary disks (e.g., ALMA Partnership et al. 2015; Andrews et al. 2016; Ginski et al. 2016; Isella et al. 2016; Tsukagoshi et al. 2016; Walsh et al. 2016;

Zhang et al. 2016; van Boekel et al. 2017). These structures can also affect subsequent planet formation in such ringed/gapped disks by collecting solid particles preferentially in local gas pressure maxima (i.e., rings) through aerodynamic drag.

Beyond the effects on dust and gas rings and gaps, spiral shocks can transport angular momentum and dissipate energy in various astrophysical disks including protoplanetary disks (Goodman & Rafikov 2001; Ju et al. 2016, 2017; Rafikov 2016; Zhu et al. 2016; Ryan & MacFadyen 2017). Not surprisingly, this capability is not limited to the primary spiral arm—the one directly attached to the planet—but the secondary spiral arm is also known to be able to contribute to angular momentum transport (Arzamasskiy & Rafikov 2018). If additional spiral arms exist (e.g., tertiary, quaternary, etc.) and they form through the same mechanism as the primary arm, angular momentum transport is also expected for the additional arms.

The characteristics of the observed spiral arms can be used to constrain the masses of unseen planets. As we will show in a companion paper (Bae & Zhu 2017, hereafter Paper II), the number of spiral arms varies as a function of the planet's mass. In addition, it is known that the arm-to-arm separation increases as a function of a planet's mass (Fung & Dong 2015; Zhu et al. 2015; Lee 2016, Paper II).

Despite the importance and growing number of numerical/observational studies showing multi-armed spirals in protoplanetary disks, the mechanism by which a planet excites multiple arms has not been fully understood. In Ogilvie & Lubow (2002, hereafter OL02), the formation of the primary arm was explained as the result of constructive interference among a set of wave modes having different azimuthal wavenumbers. The gravitational potential of a point mass perturber Φ_p at a position (r, ϕ) and time t can be decomposed into a Fourier series, a sum of individual azimuthal modes having azimuthal wavenumbers $m = 0, 1, 2, \ldots, \infty$:
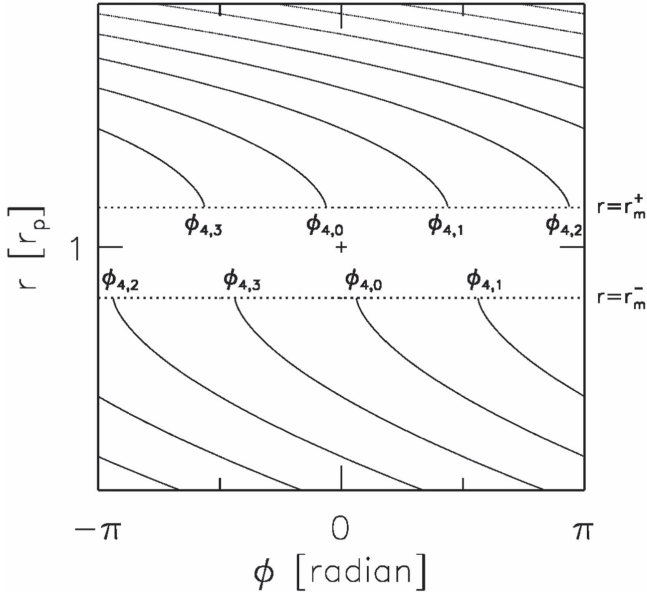
Figure 1. A schematic diagram showing the phases of individual wave modes with $m = 4$, as an example. The two horizontal dotted lines indicate the inner and outer Lindblad resonances $r = r_m^-$ and $r = r_m^+$, around which radii the wave modes launch. The perturber's position in the disk ($\phi = 0$, $r = 1\, r_p$) is marked with a cross symbol. The disk rotates left to right (increasing ϕ) in this diagram.

$\Phi_p(r, \phi, t) = \sum_{m=0}^{\infty} \Phi_m(r)\exp[im(\phi - \Omega_p t)]$. Through the resonance between the rigid rotation of a perturbing potential and the epicyclic motions of the disk material, the mth Fourier component of the potential launches m wave modes that are evenly spaced in the azimuth (Goldreich & Tremaine 1978a, 1978b, 1979, see also the review by Shu 2016). Throughout this paper, we use $n = 0, 1, \ldots, m - 1$ to represent each wave mode excited by the mth Fourier component. Among the m wave modes launched by an arbitrary mth azimuthal component, OL02 considered the one that originates from the perturber position (hereafter $n = 0$ components; see Figure 1 for illustration). Using linear wave theory, OL02 calculated the phases of $n = 0$ components having different azimuthal wavenumbers and confirmed with numerical simulations that these wave modes add coherently, creating a primary spiral arm.

In this paper, we show that additional spiral arms form in a similar way to the primary arm: through constructive interference among appropriate sets of wave modes having different m. More specifically, in the inner disk, we show that the wave modes excited at the Lindblad resonance with an azimuthal shift of $2\pi/m$ from the $n = 0$ component (i.e., $n = 1$ components) create a secondary spiral arm. Similarly, a tertiary spiral arm forms via constructive interference among the wave modes excited at the Lindblad resonance with an azimuthal shift of $4\pi/m$ from the $n = 0$ component (i.e., $n = 2$ components), and so on and so forth when possible. In the outer disk, it is the $n = m - 1$ components in each mth azimuthal component that form a secondary spiral arm. As we will show throughout this paper, this mechanism explains the characteristics of spiral arms known from previous numerical simulations very well.

This paper is organized as follows. In Section 2, we first compute the phases of individual wave modes excited by a planet using linear density wave theory and show that certain sets of wave modes can be in phase. We then carry out a suite of two-dimensional numerical simulations in Section 3, and verify that the sets of individual wave modes predicted by linear theory add constructively on to each other, creating spiral arms. In Section 4, we highlight some important nonlinear effects, based on numerical simulations with a range of planet mass covering three orders of magnitude. We summarize our findings and conclude in Section 5. In Paper II, we present a parameter study varying the disk temperature and planet mass and implications of the present work.

2. Expectation from Linear Wave Theory

A rigidly rotating point mass perturber in a differentially rotating disk excites density waves at the vicinity of Lindblad resonances (Goldreich & Tremaine 1979). The perturbation driven by the mth azimuthal Fourier component of the perturber potential in a two-dimensional (r, ϕ) plane can be written as

$$X(r, \phi, t) = \tilde{X}(r)\exp^{i[\int k(r')dr + m(\phi - \Omega_p t)]}, \quad (1)$$

where $\tilde{X}(r)$ is the amplitude of the perturbation, $k(r)$ is the radial wave vector, Ω_p is the orbital frequency of the perturber, and t represents time. The radial wavenumber k can be related to the azimuthal wavenumber m and the background disk properties through the WKBJ dispersion relation, which can be written in a two-dimensional disk as

$$m^2(\Omega - \Omega_p)^2 = \kappa^2 + c_s^2 k^2 \quad (2)$$

in the absence of self-gravity and dissipation processes. In the dispersion relation, Ω is the local orbital angular frequency, $\kappa^2 \equiv (1/r^3)d(r^2\Omega)^2/dr$ is the square of the epicyclic frequency, and c_s is the local sound speed.

As a wave propagates, its phase varies over radius. The phase of a wave with an azimuthal wavenumber m at an arbitrary radius r, $\phi_m(r)$, can be obtained by integrating $d\phi/dr = -k/m$:

$$\phi_m(r) = \phi_m(r_0) + \int_{r_0}^{r} \frac{d\phi}{dr'}dr', \quad (3)$$

where $\phi_m(r_0)$ is the phase at $r = r_0$. For a Keplerian disk ($\Omega \propto r^{-3/2}$, $\kappa = \Omega$), one can re-write the dispersion relation in Equation (2) as follows:

$$\frac{k}{m} = \frac{\Omega}{c_s} \left| \left(1 - \frac{r^{3/2}}{r_p^{3/2}}\right)^2 - \frac{1}{m^2} \right|^{1/2}. \quad (4)$$

Inserting Equation (4) into Equation (3), with $d\phi/dr = -k/m$, we obtain

$$\phi_m(r) = \phi_m(r_0) - \int_{r_0}^{r} \frac{\Omega(r')}{c_s(r')} \left| \left(1 - \frac{r'^{3/2}}{r_p^{3/2}}\right)^2 - \frac{1}{m^2} \right|^{1/2} dr'. \quad (5)$$

Density waves driven by a point mass perturber launch around the Lindblad resonance $r_m^{\pm} = (1 \pm 1/m)^{2/3}r_p$, propagating inward in the inner disk and outward in the outer disk (Goldreich & Tremaine 1978b, 1979), where r_p is the radius of the perturber's circular orbit. Far from the resonance, the

phases of m wave modes are

$$\phi_{m,n}(r) = -\text{sgn}(r - r_p)\frac{\pi}{4m} + 2\pi\frac{n}{m}, \qquad (6)$$

where $n = 0, 1, \ldots, m - 1$, as can be inferred from the asymptotic behavior of the Airy function (Ward 1986; OL02).[5] Using Equations (5) and (6), we now obtain the phases of individual wave modes with any given combination of m and n:

$$\phi_{m,n}(r) = -\text{sgn}(r - r_p)\frac{\pi}{4m} + 2\pi\frac{n}{m}$$
$$- \int_{r_m^\pm}^{r} \frac{\Omega(r')}{c_s(r')} \left| \left(1 - \frac{r'^{3/2}}{r_p^{3/2}}\right)^2 - \frac{1}{m^2} \right|^{1/2} dr'. \quad (7)$$

Throughout this paper, we call Equation (7) the phase equation.

We note that the phase equation consists of a constant component (the first two terms) that determines the launching position of wave modes in the azimuth at the Lindblad resonance, and a radially varying component (the third term) that determines how tightly the wave modes are wrapped as they propagate away from the Lindblad resonance. The tightness of the wave modes depends on the azimuthal wavenumber m, as well as the background disk temperature (c_s), and rotation profiles (Ω). The epicyclic term is often ignored in the literature, such that the dependence of the wave propagation on m is neglected. This may be a minor effect in many cases; however, the m-dependency in wave propagation is what enables the formation of multiple spiral arms by a single perturber, and therefore the epicyclic term has to be included.

To help visualize the wave excitation and propagation, we present a schematic diagram showing the phases of individual wave modes with $m = 4$ as an example in Figure 1 .

2.1. Primary Spiral Arm Formation

In OL02, the formation of the primary spiral arm driven by a planet in a protoplanetary disk was explained as the result of constructive interference among wave modes having different azimuthal wavenumbers. The constructive interference considered was for $n = 0$ components of each mth azimuthal mode. Here, we follow OL02 and briefly summarize the findings since it will help one to understand the formation of additional spiral arms that will be explained in the following section. For the example presented in this section, we adopt a temperature profile that is decreasing as a function of radius following $T \propto r^{-1/2}$ ($c_s \propto r^{-1/4}$). In addition, we limit our attention to a Keplerian rotation profile and adopt a disk aspect ratio at $r = r_p$ of $(h/r)_p = (c_s/v_\phi)_p = 0.1$ such that the sound speed is much smaller than the rotation speed.

For $n = 0$ components, ϕ_m is independent of m in the large m limit as can be seen from the phase equation. This implies that waves with different azimuthal wavenumbers can have the same phase so constructive interference among the waves may be possible. In Figure 2, we present the phases of $n = 0$ components of wave modes having azimuthal wavenumbers

$m = 1$–20, calculated with the phase equation. The phases are growing positively/negatively in the inner/outer disk, meaning that these wave modes are trailing waves. The fact that the phases become greater than 2π as they propagate to the inner disk, or smaller than -2π in the outer disk, indicates that these wave modes can wind up multiple times before they reach the disk inner/outer boundary.

In the right panels of Figure 2, we present the relative phases of $n = 0$ wave modes ($\phi_{m,0}$) to the phase of the $m = \infty$ wave mode (ϕ_∞), so that the phase difference among the wave modes can be more clearly seen. As can be seen from the figure, wave modes are nearly in phase when they launch; this is why the primary arm forms directly attached to the perturber. However, because wave modes with a small m are less tightly wound than those with a large m, as inferred from the phase equation, small m modes are left behind/ahead in the inner/outer disk as they propagate.

The perturbation driven by a point mass perturber in a disk is dominated by an azimuthal wavenumber $m \approx (1/2)(h/r)_p^{-1}$ (Goldreich & Tremaine 1980). In order for the wave modes with different m to be coherently added, they have to be within the wave crest generated by the dominating mode, $\Delta\phi \approx 2\pi(h/r)_p$. This azimuthal width is presented with "I"-shaped symbols in Figure 2. As can be inferred from the figure, the constructive interference can fail for small m wave modes far from the planet.

2.2. Formation of Additional Spiral Arms

Extending the primary spiral arm formation scenario outlined in the previous section, we propose that the nth components of each azimuthal mode, where n is now non-zero, can become in phase as they propagate and form additional spiral arms through constructive interference. While $n = 0$ components launch almost in phase as seen in Figure 2, other n components launch with non-negligible phase differences. For example, if one would draw $\phi_{m,1}$ for different m in Figure 1, small m modes will launch with larger initial phases than large m modes. However, small m modes are less tightly wound than large m modes so it is possible that small m modes are caught up by large m modes as the wave modes propagate. As in Section 2.1, we compute the phases of different wave modes and examine whether or not constructive interference will be possible. We first focus on the inner disk in Section 2.2.1 and then move on to the outer disk in Section 2.2.2.

2.2.1. Inner Disk

In Figure 3, we plot the phases of $n = 1$ components for $m = 2$–20 azimuthal modes. As shown, small m modes launch at a larger azimuthal angle, but large m modes catch up with the small m modes in phase because the small m modes are less tightly wound. In this specific example, $m = 2$–20 modes become in phase ($\Delta\phi \lesssim 2\pi(h/r)_p$) at $r \sim 0.3\,r_p$. The same can happen for $n = 2$ components; however, small m modes in this case will launch at even larger initial azimuthal angles compared with $n = 1$ components, so the wave modes have to travel further in order to become in phase. For $n = 2$ components, $m = 3$–20 modes become in phase at $r \sim 0.1\,r_p$. For the disk considered here, $n > 2$ components are unlikely to become in phase before they reach the disk's inner boundary.

---

[5] One may use a phase offset term $\pi/(3m)$ in Equation (6), which is the offset at exact Lindblad resonance locations (Ward 1986), instead of $\pi/(4m)$. The differences between the two offset values are small ($\pi/(12m)$), especially when $m \gg 1$ modes are considered, and we find that the formation mechanism of the spiral arms is not affected by the choice of the offset value (i.e., $\pi/(4m)$ versus $\pi/(3m)$).
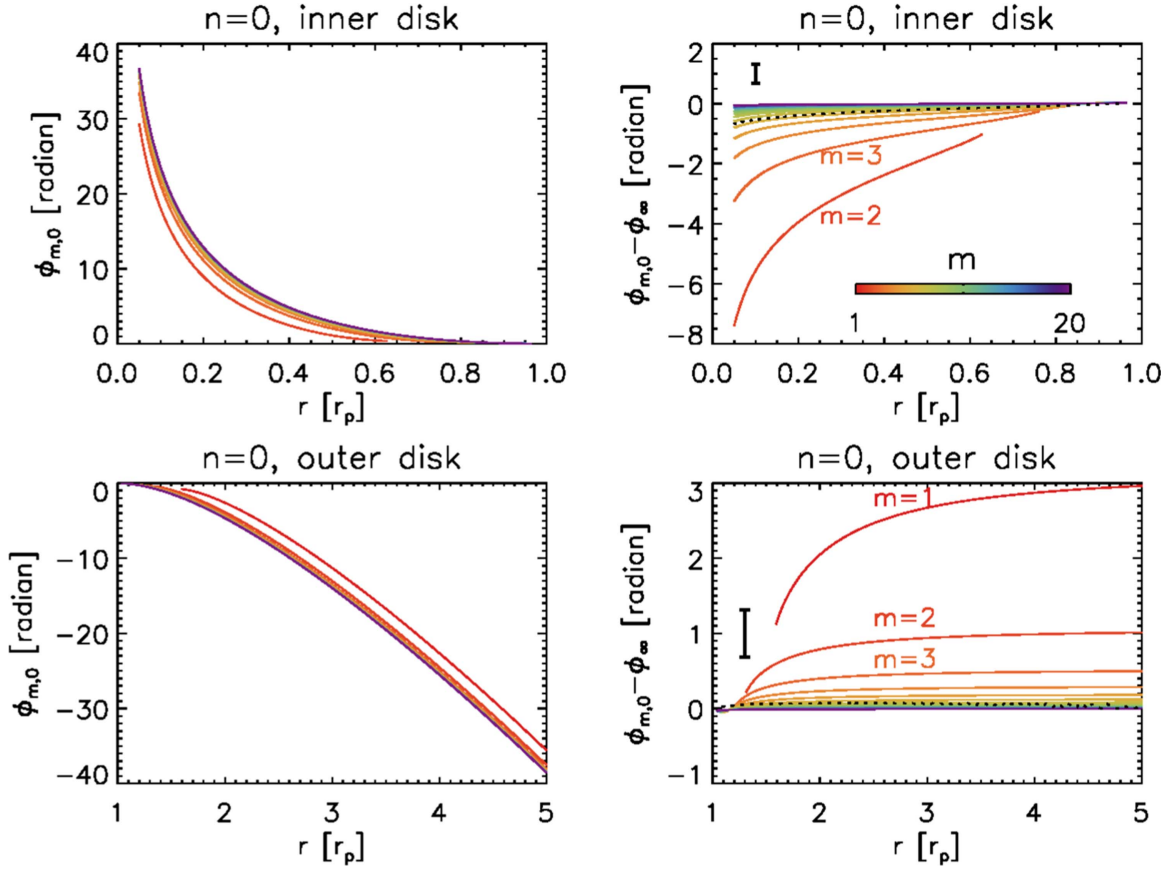
Figure 2. Phases of $n = 0$ wave modes with different azimuthal wavenumbers from $m = 1$ (red) to 20 (purple): the left panels show the actual phase values while the right panels show the relative phase values to $m = \infty$ mode. The upper panels present phases in the inner disk ($r < r_p$) while the lower panels present phases in the outer disk ($r > r_p$). Note that there is no $m = 1$ mode in the inner disk because its inner Lindblad resonance is located at $r = 0$. The "I"-shaped marks in the right panels show the azimuthal width $\Delta\phi \approx 2\pi(h/r)_p$ within which different modes have to be located to participate in the constructive interference. The dotted curves in the right panels show the phase of the primary arm in the numerical simulation with full perturber potential (see Section 3).

2.2.2. Outer Disk

We now turn our attention to the outer disk. We examine $n = m - 1$ and $n = m - 2$ components instead of $n = 1$ and 2 components, since the wave modes considered here are trailing waves and, again, small m modes are less tightly wound than large m modes. In Figure 4, we present the phases of $n = m - 1$ and $n = m - 2$ components. As seen in the figure, the phase differences among the different m modes initially decrease, but remain nearly constant beyond $r \sim 3\,r_p$. It hence appears that small m modes are not able to catch up to large m modes.

While we present the phases out to $r = 5\,r_p$ only in Figure 4, constructive interference for $n = m - 1$ and $n = m - 2$ components beyond the radius is unlikely. This can be inferred from the phase equation. When $r \gg r_p$, the last term in the phase equation simplifies to

$$\int_{r_m^+}^{r} \frac{\Omega(r')}{c_s(r')} \left| \left(1 - \frac{r'^{3/2}}{r_p^{3/2}} \right)^2 - \frac{1}{m^2} \right|^{1/2} dr'$$

$$\simeq \int_{r_m^+}^{r} \frac{\Omega(r')}{c_s(r')} \left(\frac{r'}{r_p} \right)^{3/2} dr', \qquad (8)$$

and thus has no m dependence. This means that when different m modes launch at different azimuthal angles and they are not in phase before $r \gg r_p$, they will not be in phase in the outer disk. For the disk considered here, it is thus expected that only

one arm forms in the outer disk through constructive interference among $n = 0$ components. On the other hand, more than one outer spiral arm can form when multiple sets of wave modes become in phase before $r \gg r_p$, which can occur in colder disks (see Paper II).

3. Numerical Simulations: Verifying the Linear Theory Prediction

In Section 2, we showed that appropriate sets of wave modes having different azimuthal wavenumbers can be in phase from their launching points ($n = 0$ component) or as they propagate (non-zero nth components). In this section, we carry out numerical simulations to verify that constructive interference among the sets of wave modes predicted by linear theory can indeed occur, generating spiral arms. We consider three models for this purpose.

Model 1. We carry out 20 calculations, each of which includes a single mth Fourier-decomposed potential of a planet, where $m = 1, 2, ..., 20$. We then construct a single surface density output Σ by summing the perturbed density from each of the single mode calculations: $\Sigma = \Sigma_{\rm init} + \sum_{m=1}^{20}(\Sigma_m - \Sigma_{\rm init})$, where $\Sigma_{\rm init}$ is the initial, unperturbed surface density, and Σ_m is the surface density obtained in a simulation with only the mth Fourier potential included.
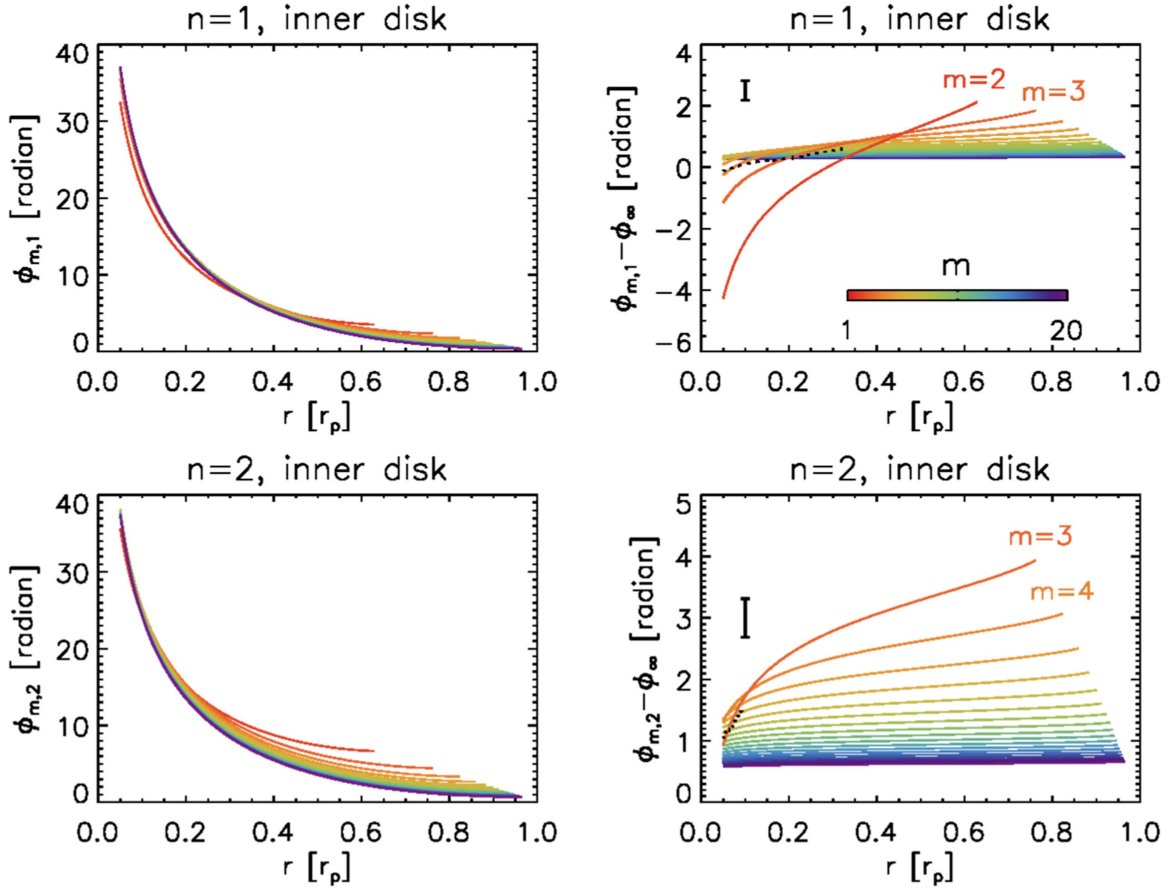
Figure 3. Phases of $n = 1$ and $n = 2$ wave modes in the inner disk ($r < r_p$), having different azimuthal wavenumbers from $m = 1$ (red) to 20 (purple): the left panels show the actual phase values while the right panels show the relative phase values to the $m = \infty$ mode. Note that the shape of the phase curves for a given m are identical regardless of n, but only the launching point is shifted in the azimuth for different ns. The "I"-shaped marks in the right panels show the azimuthal width $\Delta\phi \approx 2\pi(h/r)_p$ within which different modes have to be located for the constructive interference to occur. The dotted curves in the right panels show the phase of the secondary and tertiary arms in the numerical simulation with full perturber potential (see Section 3).

Model 2. We carry out one calculation in which $m = 1$–20 Fourier-decomposed azimuthal components of the planet potential are included.

Model 3. We carry out one calculation with the full planet potential.

By comparing Models 1 and 2 with Model 3, we will be able to test whether linear addition (i.e., superposition) of individual waves explains the main features of the model with the full potential. The comparison between Models 1 and 2 will allow us to examine if any nonlinear mode–mode interactions exist. If there is no interaction between the different azimuthal modes, we expect that Models 1 and 2 will produce identical results. Finally, by comparing Model 2 with Model 3, we will be able to see the contribution from large m modes ($m > 20$) in generating spiral arms.

3.1. Numerical Methods

We solve the hydrodynamic equations for mass and momentum conservation in the two-dimensional polar coordinates (r, ϕ) using FARGO 3D (Benítez-Llambay & Masset 2016):

$$\frac{\partial \Sigma}{\partial t} + \nabla \cdot (\Sigma v) = 0, \tag{9}$$

$$\Sigma\left(\frac{\partial v}{\partial t} + v \cdot \nabla v\right) = -\nabla P - \Sigma \nabla (\Phi_* + \Phi_p). \tag{10}$$

In the above equations, Σ is the surface density, v is the velocity, $P = \Sigma c_s^2$ is the pressure where c_s is the isothermal sound speed, $\Phi_* = -GM_*/r$ is the gravitational potential of the central star, and Φ_p is the potential of the planet. The potential of the planet is

$$\Phi_p(r, \phi, t) = -\frac{GM_p}{(|r - r_p|^2 + s^2)^{1/2}}, \tag{11}$$

where M_p is the planet mass, r and r_p are the radius vectors of the center of the grid cells in question and of the planet, and $s = 0.6\,h_p$ is the smoothing length. In this work, we ignore the indirect term that arises due to the offset between the central star and the origin of the coordinate system. The "full planet potential" in Equation (11) is used for Model 3.

Assuming a circular planetary orbit, the potential in Equation (11) can be expanded into a Fourier series:

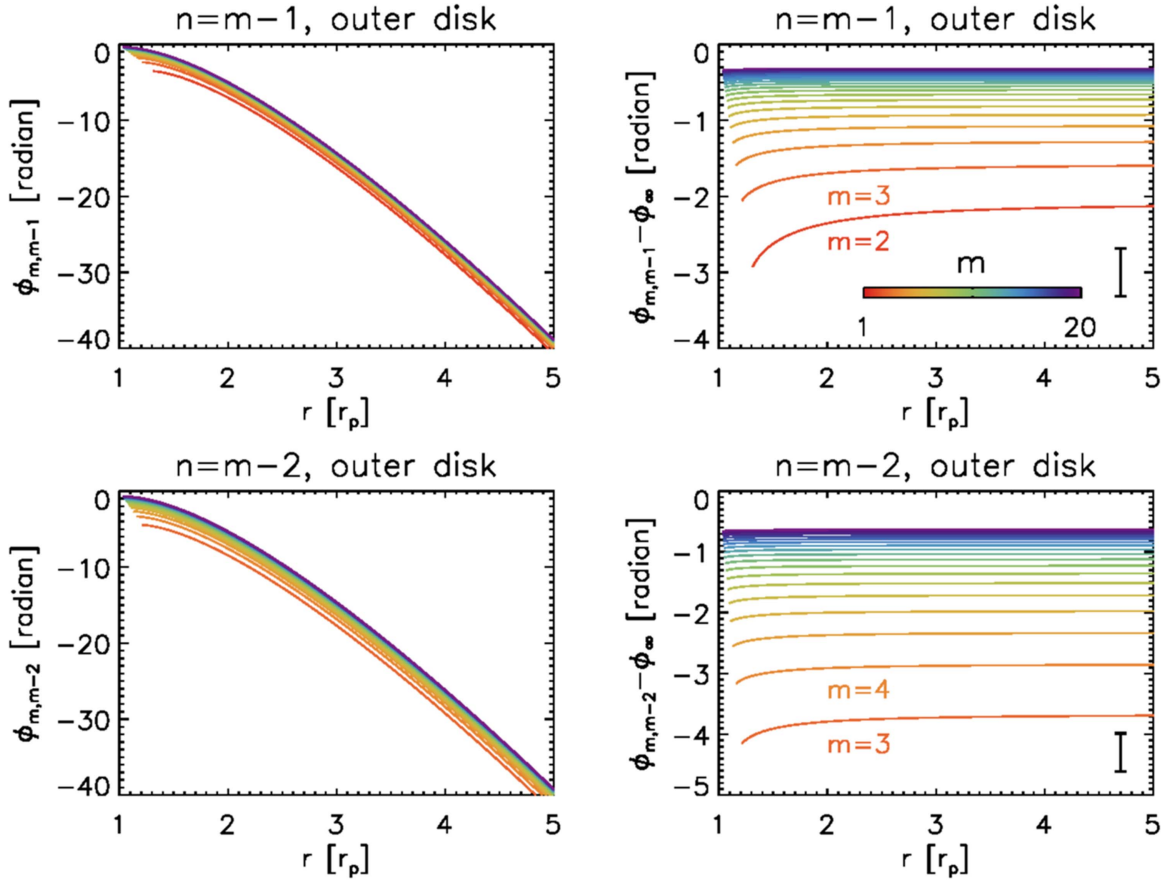$$\Phi_p(r, \phi, t) = \sum_{m=0}^{\infty} \Phi_m(r) \exp[im(\phi - \Omega_p t)]. \tag{12}$$

Figure 4. Same as Figure 3, but for $n = m -1$ and $n = m -2$ wave modes in the outer disk ($r > r_p$).

Here,

$$\Phi_m(r) = -(2 - \delta_{m0})\frac{GM_p}{2r_p}b_{1/2}^m(\beta), \quad (13)$$

where δ_{ij} is the Kronecker delta, $\beta = r/r_p$, and $b_{1/2}^m(\beta)$ is the Laplace coefficient defined as (Brouwer & Clemence 1961)

$$b_{1/2}^m(\beta) \equiv \frac{2}{\pi}\int_0^\pi \frac{\cos(m\phi)}{(1 - 2\beta\cos\phi + \beta^2 + s^2)^{1/2}}d\phi. \quad (14)$$

The smoothing length s is included in the denominator of the right-hand side of Equation (14) in order for the summation of the Fourier-decomposed potential in Equation (12) to be consistent with the full potential in Equation (11). The Fourier-decomposed potential in Equation (12) is used for Models 1 and 2, with ms chosen following the model description.

We use the planet mass of $M_p = 0.01 M_{th}$, where $M_{th} \equiv c_s^3/\Omega G = M_*(h/r)_p^3$ is the so-called thermal mass (Lin & Papaloizou 1993; Goodman & Rafikov 2001), at which the Hill radius is comparable to the disk scale height. Assuming a solar-mass star, $0.01 M_{th}$ is about 3 Earth masses. When $M_p \ll M_{th}$, the excitation and the initial propagation of the density waves from the planet is known to be well approximated in the linear regime. Goodman & Rafikov (2001) predict that spiral arms driven by a $0.01 M_{th}$ planet steepen into shocks ~6 scale heights away from the planet (see Equation (30) of their paper). When we compare the phases of the spiral arms driven by a $0.01 M_{th}$ planet with the ones driven by a $0.001 M_{th}$ planet in Paper II, we find that the difference in

the phases of the spiral arms are negligible, not only within the ±6 scale height regions around the planet, but in the entire disk. This suggests that, although spiral arms driven by a $0.01 M_{th}$ planet can steepen into shocks, nonlinear effects are negligible. As shown in the following section, density waves excited by a $0.01 M_{th}$ planet in numerical simulations indeed show excellent agreement with the linear theory predictions.

Our initial disk has power-law surface density and temperature distributions: $\Sigma_{init}(r) = \Sigma_p(r/r_p)^{-1}$ and $T(r) = T_p(r/r_p)^{-1/2}$, where Σ_p and T_p are the surface density and temperature at the location of the planet $r = r_p$. We choose T_p such that $(h/r)_p = 0.1$, to be consistent with the disk model used in Section 2. The simulation domain extends from $r_{in} = 0.05 r_p$ to $r_{out} = 5 r_p$ in radius and from 0 to 2π in azimuth. We adopt 4096 logarithmically spaced grid cells in the radial direction and 5580 uniformly spaced grid cells in the azimuthal directions, with which $\Delta r{:}r\Delta\phi \simeq 1{:}1$. At the radial boundaries, we adopt a wave-damping zone (de Val-Borro et al. 2006) to suppress wave reflection. No kinematic viscosity is added in the simulations.

3.2. Simulation Results

We first present the perturbed density distributions, $\delta\Sigma_m/\Sigma_{init}$, where $\delta\Sigma_m = \Sigma_m - \Sigma_{init}$, from individual mode calculations (Model 1) in Figure 5. While Figure 5 includes results from the $m = 1$–4 mode runs only, we note that the discussion below applies to all the individual wave mode runs with $m = 1$–20.

Most importantly, the excitation and propagation of the density waves in the numerical simulations show excellent agreement with the linear theory. Each azimuthal component of
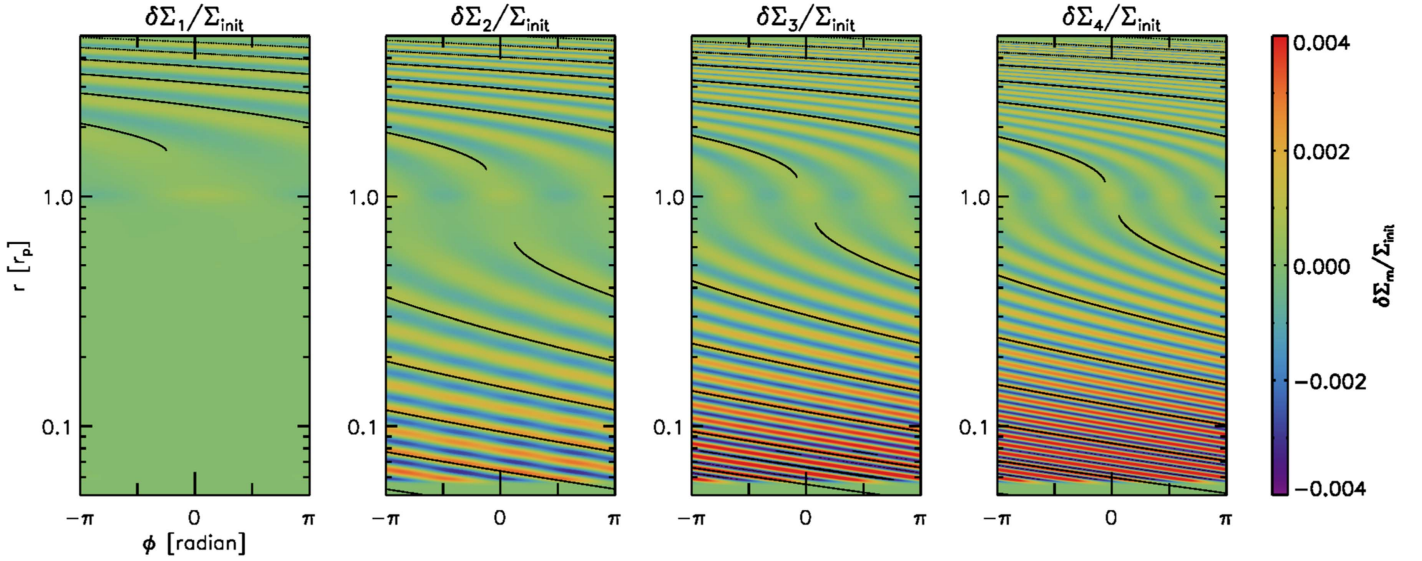
Figure 5. Two-dimensional distributions of the perturbed surface density $\delta\Sigma_m/\Sigma_{init}$ from individual mode runs with (left to right) $m = 1$–4, where $\delta\Sigma_m = \Sigma_m - \Sigma_{init}$. In each panel, the phase of $n = 0$ component calculated using Equation (7) is plotted with a black curve. Note the excellent agreement between the linear theory and numerical simulation.

the Fourier-decomposed potential excites m wave modes at the inner and outer Lindblad resonance; $m = 1$ mode does not excite any waves in the inner disk because the inner Lindblad resonance is located at $r = 0$. The perturbation from individual wave modes is <1% over the entire simulation domain, supporting the fact that these waves are in a linear regime. The amplitude of the perturbation in all the individual mode runs increases as the waves propagate, which is also in a good agreement with the expectation for linear waves (Rafikov 2002a).

In Figure 6, we display the perturbed surface density distributions $\delta\Sigma/\Sigma_{init}$ for Model 1, 2, and 3. The azimuthal distributions of $\delta\Sigma/\Sigma_{init}$ at $r = 1.5, 0.6, 0.3$, and $0.1\,r_p$ are presented in Figure 7 for more quantitative comparison among the models. All three models form three spiral arms in the inner disk and one spiral arm in the outer disk. The primary arm is directly attached to the planet, spiraling away from it. In the inner disk, the secondary and tertiary arms excite around $r \sim 0.3\,r_p$ and $r \sim 0.1\,r_p$, respectively. Note that these radial positions are in good agreement with the predictions made based on the phase argument in Section 2. The secondary and tertiary arms become narrower in azimuth and produce stronger perturbations as they propagate, indicating that the individual waves participating in the formation of these arms become closer in phase, so the constructive interference become more effective. This is also consistent with the linear theory prediction (see e.g., Figure 3).

Comparing the models, we find that both Models 1 and 2 reproduce the full potential model (Model 3) fairly well. The major difference seen in Model 3 is that the primary arm is sharper and produces a larger perturbation close to the planet (e.g., $r = 0.6$ and $1.5\,r_p$ in Figure 7). This is because the Laplace coefficients (Equation (14)) that determine the strengths of the perturbation driven by the individual azimuthal modes decline slowly with increasing m when $\beta = r/r_p$ is close to unity. The contribution from the azimuthal modes with $m > 20$ is therefore not negligible near $r = r_p$. At the radius a spiral arm excites (e.g., $0.3\,r_p$ for the secondary and $0.1\,r_p$ for the tertiary), and we see that all of the three models agree with each

other very well, suggesting that the excitation of the additional spiral arms is a linear process. As spiral arms propagate, however, we see that Models 2 and 3 deviate from Model 1. For example, the phase of the primary arm in Models 2 and 3 is offset in phase from the primary arm phase in Model 1 at $r = 0.1$ and $0.3\,r_p$. Also, at $r = 0.1\,r_p$, the secondary arm breaks up into finer azimuthal scales in Models 2 and 3. This suggests that there could potentially be nonlinear mode coupling (e.g., Lee 2016) even at this low level of perturbations.

To further ensure that it is the $n = 0$, $n = 1$, and $n = 2$ components from different azimuthal modes that generate the primary, secondary, and tertiary arms, we subtract each nth component from Model 1 one at a time when constructing the final surface density output. More specifically, we return the surface density associated with the nth component in each individual azimuthal mode calculation to the unperturbed value as illustrated in Figure 8.

In Figure 9, we present the perturbed density distributions from Model 1 along with those from models without $n = 0$ components, $n = 1$ components ($n = m - 1$ components in the outer disk), and $n = 2$ components ($n = m - 2$ components in the outer disk). For more quantitative comparison among the models, we present the azimuthal distributions of the perturbed density at $r = 0.1\,r_p$ in Figure 10. As seen in the figures, the primary arm does not form when the $n = 0$ components are subtracted, the secondary arm does not form when the $n = 1$ components are subtracted, and the tertiary arm does not form when the $n = 2$ components are subtracted. Also, we note that removing the $n = 0$ components does not affect the secondary and tertiary arms. Likewise, removing the $n = 1$ or $n = 2$ components only affects the secondary or tertiary arms.

Previous studies pointed out that a negative density perturbation appears before the secondary spiral arm forms (e.g., Arzamasskiy & Rafikov 2018). We also find such a negative density perturbation in our simulations: as shown in Figures 6 and 7, the negative density perturbation just on the right side of the primary arm develops at $r \sim 0.6\,r_p$ and
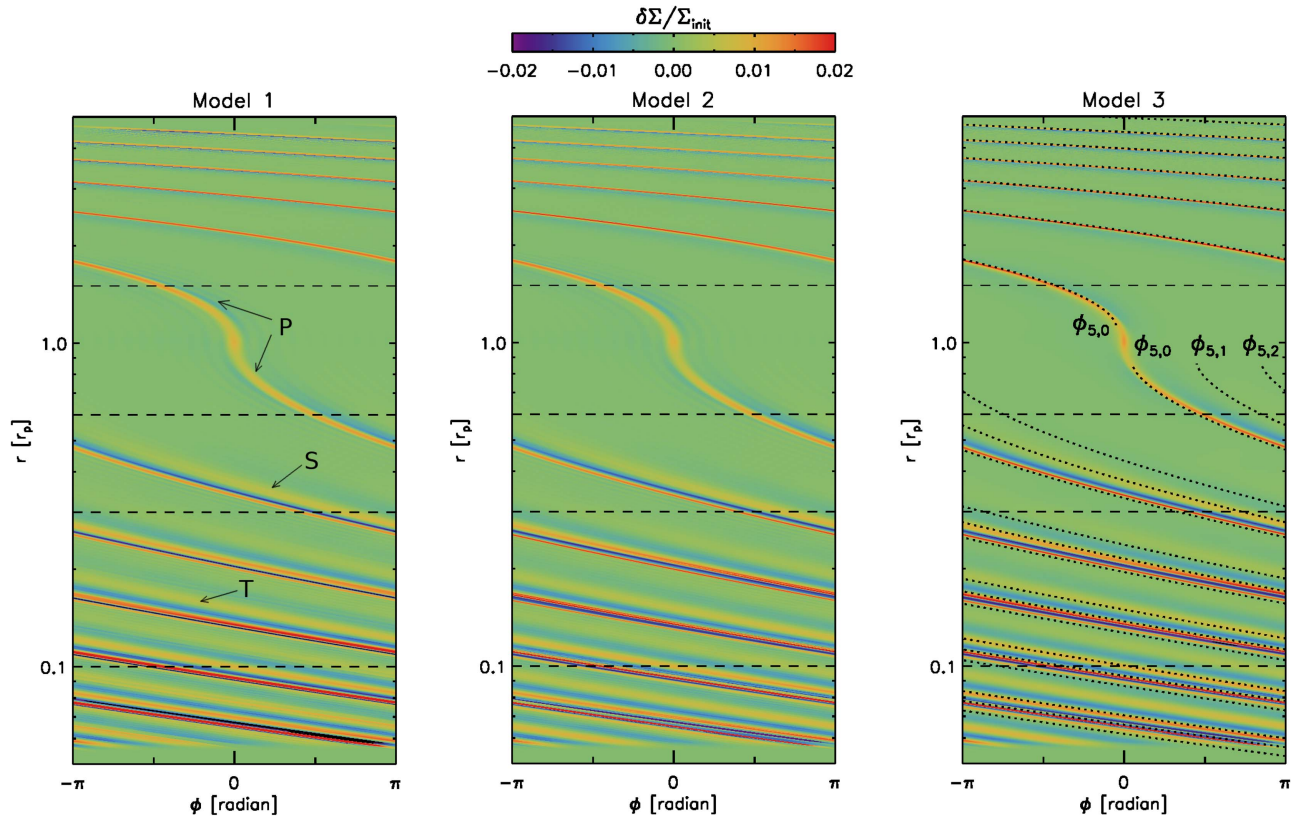
Figure 6. Two-dimensional distributions of the perturbed surface density $\delta\Sigma/\Sigma_{init}$ for (from left to right) Models 1, 2, and 3. The horizontal lines indicate (from top to bottom) $r = 1.5, 0.6, 0.3$, and $0.1\ r_p$, for which radii we present the one-dimensional perturbed density distributions along the azimuth in Figure 7. The primary, secondary, and tertiary arms are labeled with "P," "S," and "T," respectively, in the left panel. In the right panel, the dotted curves labeled as $\phi_{5,0}$, $\phi_{5,1}$, and $\phi_{5,2}$ present the phases of the three spiral arms predicted with the linear theory.
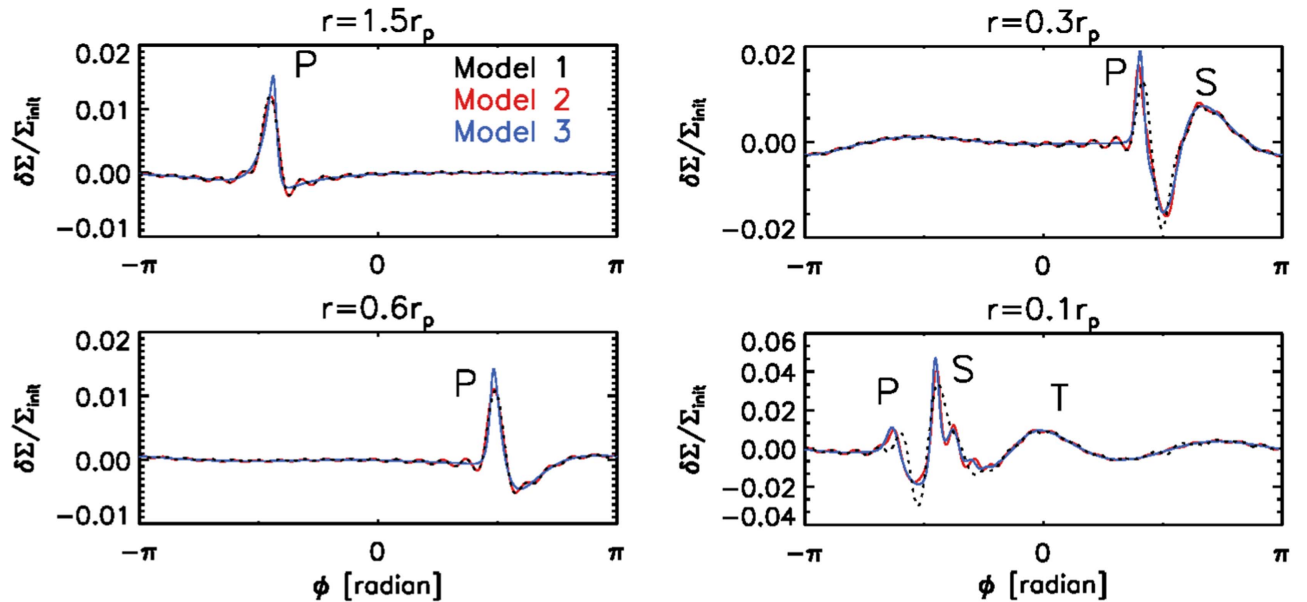

Figure 7. Black dotted curves show the perturbed density distributions $\delta\Sigma/\Sigma_{init}$ from Model 1, whereas the red and blue curves show the results from Models 2 and 3. The primary, secondary, and tertiary arms are labeled with "P," "S," and "T," respectively.

deepens inward before the secondary arm launches. In the constructive interference scenario we explain here it is obvious that, after $n = 0$ components (i.e,. wave crests) form the primary arm, the wave troughs between $n = 0$ and $n = 1$ wave crests have to be in phase before the $n = 1$ wave crests become in phase to form the secondary arm. Similarly, the

wave troughs between $n = 1$ and $n = 2$ wave crests become in phase before $n = 2$ wave crests form the tertiary arm, and this is what forms the negative density perturbation between the secondary and tertiary arms. In short, a negative density perturbation between spiral arms can be understood as constructive interference among wave troughs, as opposed to
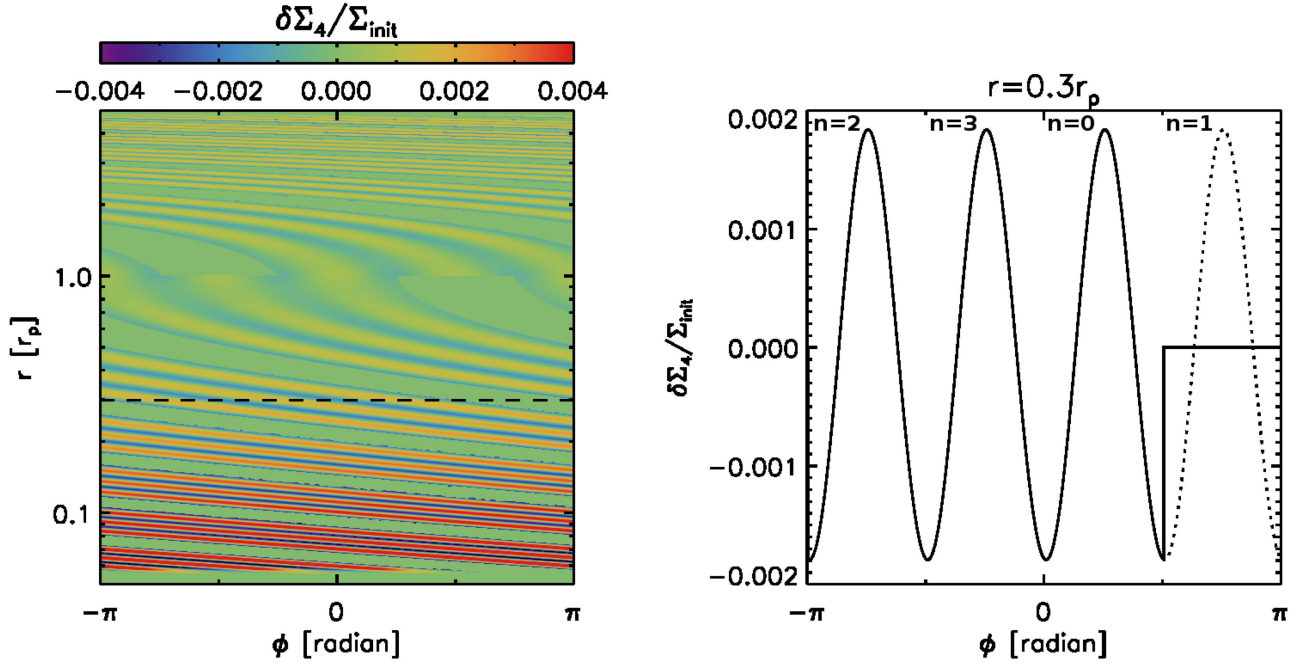
Figure 8. Example showing the subtraction of selected nth components. In this example, $n = 1$ component in the inner disk and $n = m - 1$ component in the outer disk are subtracted from $m = 4$ mode. (Left) The two-dimensional distribution of the perturbed surface density $\delta\Sigma_4/\Sigma_{\rm init}$ after the subtraction. The horizontal dashed line indicates $r = 0.6\,r_p$. (Right) $\delta\Sigma_4/\Sigma_{\rm init}$ along the azimuth at $r = 0.6\,r_p$, before the subtraction with a dotted curve and after the subtraction with a solid curve.

constructive interference among wave crests, which forms a positive density perturbation (i.e., spiral arms).

Since the linear approach (i.e., superposition of individual wave modes) explains the formation and propagation of spiral arms well, we can make use of the linear wave theory to predict the phases of spiral arms. In the right panel of Figure 6, we present the phases of $m = (1/2)(h/r)_p^{-1} = 5$ mode for $n = 0$, 1, and 2 components: $\phi_{5,0}$, $\phi_{5,1}$, and $\phi_{5,2}$. As shown, the phases of spiral arms predicted by the linear theory agree well with the phases of spiral arms in the numerical simulation.

4. Nonlinear Evolution of Spiral Arms

As shown in the previous section, the linear wave theory explains the formation and propagation of spiral arms reasonably well for a sufficiently low-mass planet (i.e., $0.01\,M_{\rm th}$). As the planet mass grows, however, nonlinear effects are expected to play an increasingly important role. In order to investigate the nonlinear effects, we ran a set of simulations with various planet masses of $M_p = 0.1$, 0.3, 1, 3, and $10\,M_{\rm th}$ (0.1, 0.3, 1, 3, and 10 Jupiter mass assuming a solar-mass star), adopting the full planet potential as in Equation (11). All the other numerical setups, except the planet mass, remain the same as explained in Section 3.

In each of the simulations we introduce the planet at the beginning of the calculation with its full mass, instead of growing the planet mass over an extended period of time. We take this approach because planets with $M_p \gtrsim 1\,M_{\rm th}$ open a gap around their orbit. The gap edges then become unstable to the growth of the Rossby wave instability (RWI; Lovelace et al. 1999; Li et al. 2000, 2001), launching spiral waves that have different pattern speeds from the planet-driven spiral arms. For the planet masses considered here, we find that the RWI develops over about ten or more orbital times. By having the full planet mass from the beginning, spiral arms launched

by the planet fully develop in the entire disk well before the RWI develops. This approach thus allows us to avoid the interference from RWI-driven spiral waves.

With the background disk profile assumed here, the planet excites two or three spiral arms in the inner disk depending on its mass. In the outer disk, on the other hand, the planet excites only one spiral arm independently on the planet mass. To determine the phases of spiral arms from the simulations, we find the local maximum of the density perturbation in the azimuth as we follow each spiral arm along the radius. The phases of the primary, secondary, and tertiary arms for different planet masses are presented in Figure 11. In the figure, we also indicate the radial locations at which the secondary and tertiary spiral arms start to shock the disk gas. In order to diagnose the shock location, we compute the potential vorticity (PV) $\zeta \equiv (\nabla \times \mathbf{v})/\Sigma$. The idea is that the PV experiences a jump at the shock front (Li et al. 2005; Dong et al. 2011; Bae et al. 2017). In Figure 12, we plot the perturbed surface density distributions along the azimuth at some selected radii to show the level of perturbation driven by spiral arms and the morphology of the spiral arm front. We highlight some important aspects of spiral arm formation and propagation below.

First, the phases of the secondary and tertiary arms during their excitation and initial propagation agree reasonably well with the linear wave theory prediction. Looking at the phases of the secondary arms in Figure 11 first, one can see that they are very closely located to each other in the azimuth at $r = 0.4$–$0.6\,r_p$ for such a broad range of planet mass, and moreover follow the linear prediction ($\phi_{5,1}$) very well. This is also clear in Figure 12: at $r = 0.4\,r_p$, the secondary spiral arms are located close to the linear theory prediction. Note also that secondary arms from planets with masses of $\geqslant 0.3\,M_{\rm th}$ have already evolved into shocks at this radius, as the steep density gradient as well as the shock locations presented in Figure 11 suggest. The fact that spiral arms follow the linear theory well
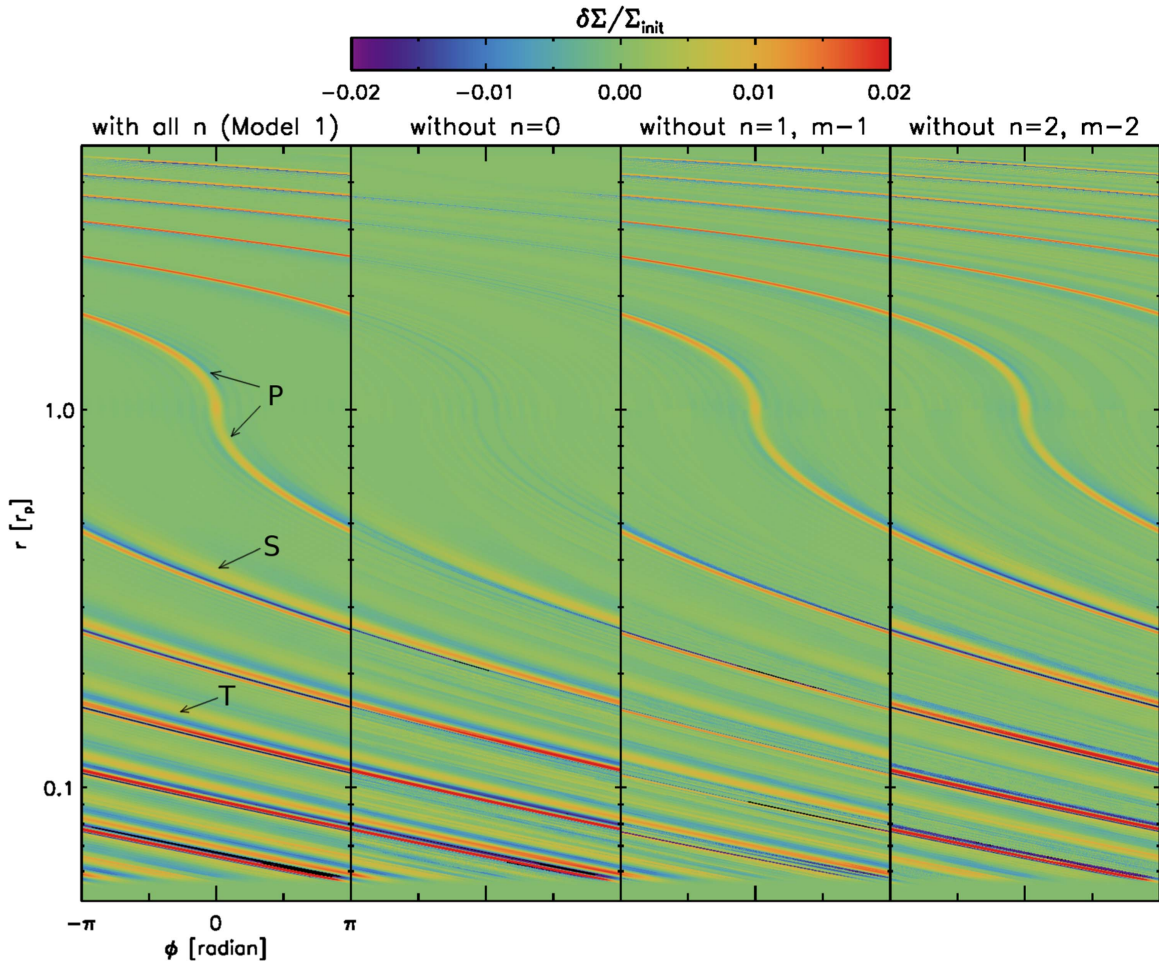
Figure 9. (Left) The two-dimensional distributions of the perturbed surface density $\delta\Sigma/\Sigma_{init}$ from Model 1. The primary, secondary, and tertiary arms are labeled with "P," "S," and "T," respectively. In the other three panels, the distributions of $\delta\Sigma/\Sigma_{init}$ (left middle) without $n = 0$ components, (right middle) without $n = 1$ components at $r < r_p$ and $n = m - 1$ components at $r > r_p$, and (right) without $n = 2$ components at $r < r_p$ and $n = m - 2$ components at $r > r_p$ are presented. Note that the primary arm does not form when the $n = 0$ components are subtracted, the secondary arm does not form when the $n = 1$ components are subtracted, and the tertiary arm does not form when the $n = 2$ components are subtracted.
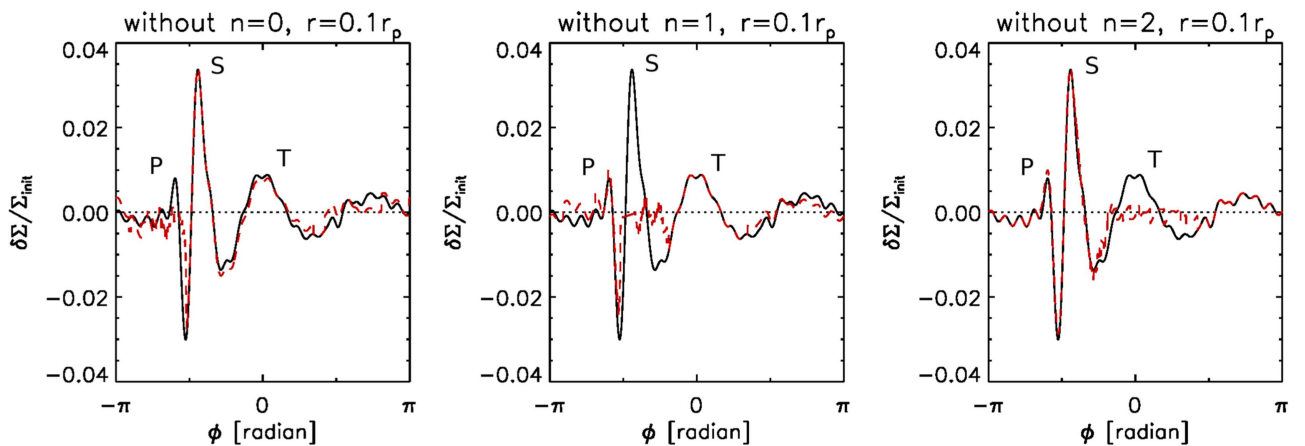

Figure 10. One-dimensional plots of $\delta\Sigma/\Sigma_{init}$ along the azimuth at $r = 0.1$ r_p. The black curves present $\delta\Sigma/\Sigma_{init}$ before the subtraction, while the red curves present $\delta\Sigma/\Sigma_{init}$ after the subtraction: (left) without the $n = 0$ components, (middle) without the $n = 1$ components, and (right) without the $n = 2$ components. Note that removing certain n components does not affect formation of other spiral arms. The primary, secondary, and tertiary arms are labeled with "P," "S," and "T," respectively.

after they start to shock the disk gas supports the notion that shocks need to propagate some distance before they deviate from the linear theory (Goodman & Rafikov 2001). Similarly, the tertiary arms are closely located in the azimuth at

$r = 0.15$–$0.2\,r_p$, and the linear theory predicts the phases of the tertiary arms very well at the radii.

Second, spiral arms are more open with a larger planet mass. Spiral arms deviate from the linear theory prediction after their
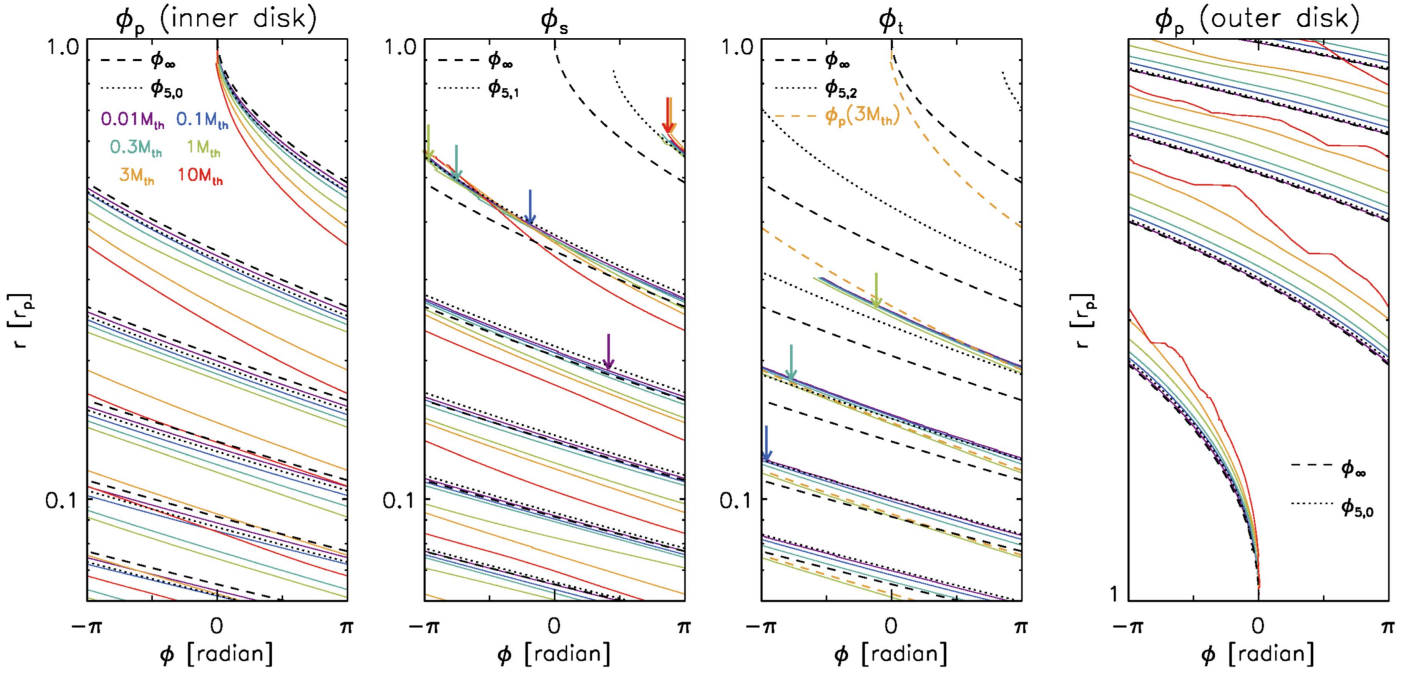
Figure 11. Phases (from left to right) of the primary arm (ϕ_p) in the inner disk, secondary arm (ϕ_s), tertiary arm (ϕ_t), and primary arm in the outer disk, for planet masses of 0.01, 0.1, 0.3, 1, 3, and 10 M_{th}. The dotted curves represent the phases for $m = (1/2)(h/r)_p^{-1} = 5$ mode ($\phi_{5,n}$) calculated with the phase equation, while the dashed curves represent the phase for $m = \infty$ mode (ϕ_∞). The arrows in the two middle panels show where each spiral arm starts to shock the disk gas, diagnosed based on the potential vorticity jump. In the middle right panel, the yellow dashed curve presents the primary arm phase for the $M_p = 3\,M_{th}$ case (see text). The distortion seen in the outer primary spiral arm for 10 M_{th} planet is due to too strong shocks at the arm front.

initial propagation, which ends sooner (i.e., at larger radii in the inner disk and at smaller radii in the outer disk) for larger planet masses (Goodman & Rafikov 2001; see also Figure 11). When a spiral arm nonlinearly steepens into a shock it travels at a faster speed, resulting in a less tightly wound shape than the linear theory prediction. The speed of the shock expansion is proportional to the amplitude of the shock, so spiral arms excited by a planet with more mass propagate at faster speeds and thus appear to be more open (Zhu et al. 2015). In Figure 13, we present the measured pitch angle of spiral arms driven by 0.1, 1, and 10 M_{th} planets from numerical simulations, along with linear theory predictions. As expected, the pitch angles measured in simulations with 1 and 10 M_{th} planets are larger than the linear theory prediction, while the linear theory and simulation agree well with each other for a 0.1 M_{th} planet. Compared at a given radius, the more massive the planet is, the more open a spiral arm is in general. This trend is commonly seen, not only for the primary arm but also for additional arms. We confirm this trend in disks with other $(h/r)_p$ values from the parameter study carried out in Paper II (see Figure 5 from the aforementioned paper).

Third, only two spiral arms form in the inner disk for sufficiently large planet masses ($\gtrsim 3\,M_{th}$). We conjecture that this is because the primary arm nonlinearly propagates and merges with the wave modes that would form a tertiary arm for smaller mass planets. In the third panel of Figure 11, we over-plot the phase of the primary arm driven by a $3\,M_{th}$ planet. As shown, the primary arm becomes more open as it nonlinearly propagates to smaller radii and eventually overlaps with the linear phase of the tertiary arm at $r \sim 0.2\,r_p$. At $r = 0.3\,r_p$ in Figure 12, the primary arm and the tertiary arm are well separated in the azimuth for a $1\,M_{th}$ planet. For a $3\,M_{th}$ planet, on the other hand, the primary arm has a broad wing-like density enhancement on the left side of the "N"-shaped shock,

which is likely formed by the waves that would constructively interfere with each other to form a tertiary arm at the azimuth. The fact that the primary arm produces a comparable magnitude of perturbation to the secondary arm with $M_p = 3$ and $10\,M_{th}$ at $r = 0.2$ and $0.3\,r_p$—whereas the secondary arm is generally stronger for lower-mass planets at these radii because the primary arm weakens due to less efficient constructive interference—also supports the idea of tertiary arm-forming waves merging with the primary arm.

Last, interference between spiral arms may occur. In Figure 14 we present the two-dimensional density distribution from the $M_p = 1\,M_{th}$ model. Also presented with a black curve in the figure is the predicted primary spiral arm front position based on nonlinear shock expansion theory (Goodman & Rafikov 2001; Rafikov 2002a), following the procedure detailed in Section 4 of Zhu et al. (2015). Note that the shock expansion theory predicts the primary arm phase reasonably well at $r \gtrsim 0.25\,r_p$, but fails inward of the radius where the primary arm gradually approaches the secondary arm in the azimuth. The primary arm then returns back to the predicted position at $r \lesssim 0.08\,r_p$. Interestingly, we find that this is not transient, but a long-lasting and stationary feature. We propose that one possibility for this is the interference between the primary and secondary spiral arms. The perturbation driven by the primary arm is expected to gradually decrease since it dissipates while propagating. However, the primary arm gains in strength inward of $\sim 0.2\,r_p$ as it approaches the secondary arm, while the secondary arm loses its strength over the same radii, possibly because some low azimuthal modes that become out of phase from the secondary arm are added to the primary arm and/or because of nonlinear mode–mode interaction. This supports the hypothesis that the deviation of the primary arm from the shock expansion theory at $0.08\,r_p \lesssim r \lesssim 0.25\,r_p$ is due to the interference from the secondary arm. In Paper II, we
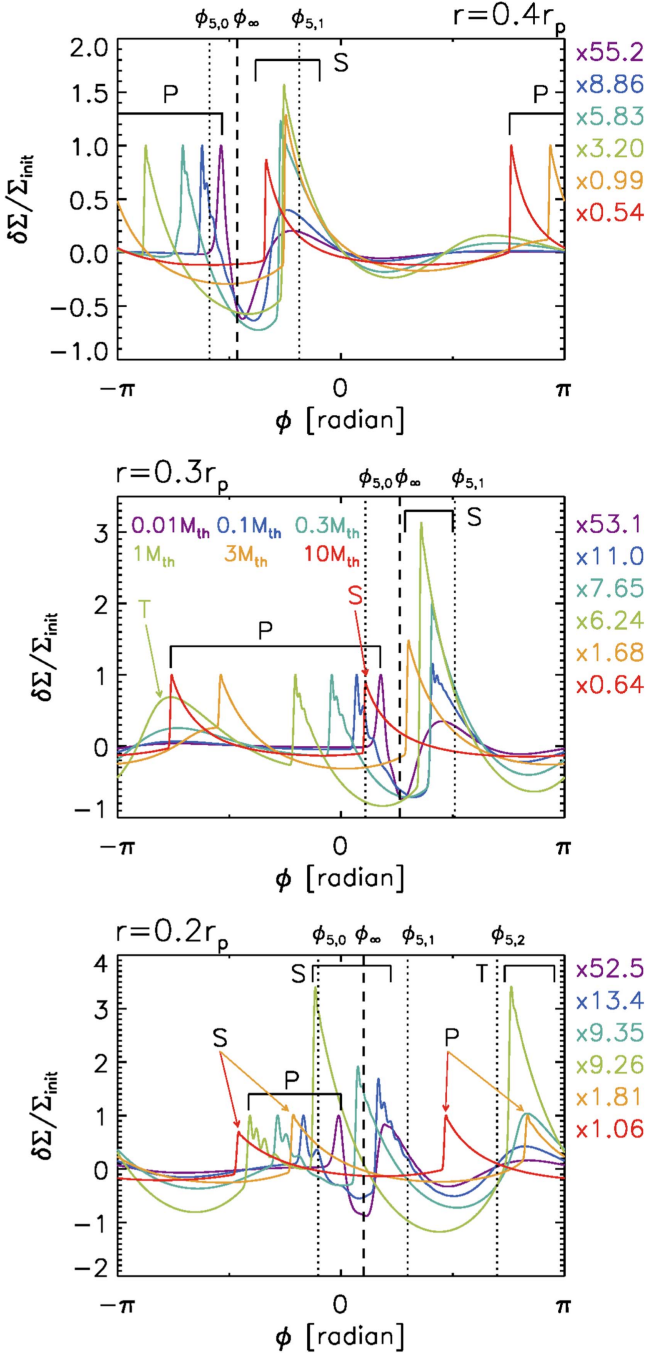
Figure 12. Perturbed surface density distributions $\delta\Sigma/\Sigma_{init}$ along the azimuth for various planet masses at (top) $r = 0.4\,r_p$, (middle) $r = 0.3\,r_p$, and (bottom) $r = 0.2\,r_p$. For the purposes of visualization, each curve is scaled by a factor presented on the right side of each panel such that the peak of the primary arm at the radius has $\delta\Sigma/\Sigma_{init} = 1$. The primary, secondary, and tertiary arms are indicated by "P," "S," and "T," respectively. The dotted vertical lines present $\phi_{5,0}$, $\phi_{5,1}$, and $\phi_{5,2}$ at each radius, while the dashed vertical line presents the phase of $m = \infty$ mode ϕ_∞ at the radius.

Figure 13. Pitch angle of the (circle) primary, (triangle) secondary, and (cross) tertiary arm for (blue) 0.1 M_{th}, (green) 1 M_{th}, and (red) 10 M_{th} planets. The black solid curve presents the pitch angle calculated with $m = \infty$ in the phase equation assuming $(h/r)_p = 0.1$, whereas the dashed curve presents the pitch angle calculated with $m = 5$.

Figure 14. Two-dimensional distribution of the perturbed surface density $\delta\Sigma/\Sigma_{init}$ from the $M_p = 1M_{th}$ model. The black curve presents the predicted shock front obtained with a nonlinear shock expansion theory (see text). The black and red arrows on the right indicate the primary and secondary arms, respectively. Note that the primary arm deviates from the nonlinear shock expansion theory prediction at $\sim0.25\,r_p$ approaching the secondary, and then returns back to the predicted phase at $\sim0.08\,r_p$.

find that interference between spiral arms in colder disks can even result in a merging of the arms, presumably made possible because spiral arms in colder disks launch with smaller azimuthal separations. While we see potential evidence of interference between spiral arms, it is unclear at the moment under which conditions such interference occurs. Interference between spiral arms is an interesting phenomenon to study, but
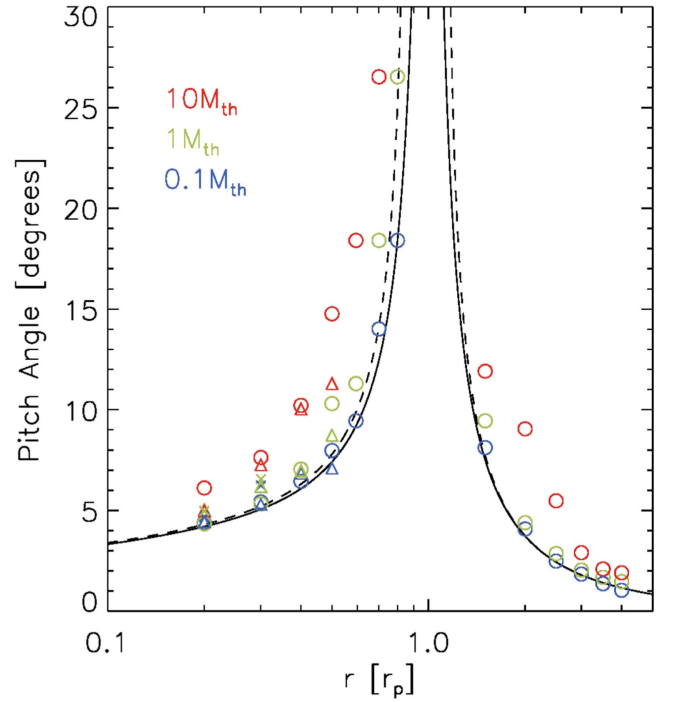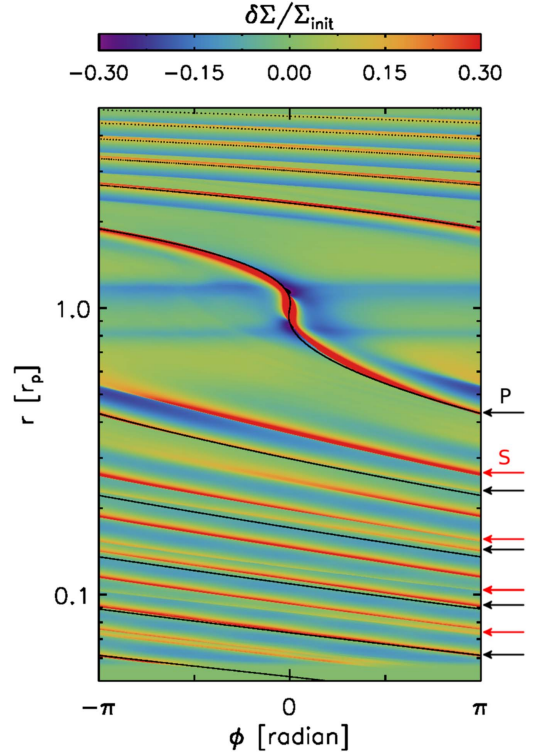
it is beyond the scope of the paper and thus a more thorough investigation is deferred to a future paper.

One thing that does not change regardless of planet mass (and also disk temperature; see Paper II) is that small m modes are less tightly wound than large m modes. This property of waves suggests that an additional spiral arm always forms ahead of the previous arm in the azimuth in the inner disk and behind the previous arm in the azimuth in the outer disk. Also, an additional spiral arm always forms farther away from the planet in radius, because the wave modes have to travel a longer distance to be in phase.

5. Summary and Conclusion

We have shown how a planet excites multiple spiral arms in the underlying protoplanetary disk. Using linear wave theory we first calculated the phases of individual wave modes excited by a planet and showed that appropriate sets of wave modes having different azimuthal wavenumbers can be in phase, from their launching points (in the case of the primary arm) or as they propagate (in the case of additional arms). By carrying out a suite of two-dimensional hydrodynamic simulations, we then verified that the sets of wave modes predicted by the linear theory add constructively on to each other and form spiral arms.

As the planet mass grows, nonlinear effects play an increasingly important role. We investigated the nonlinear effects by carrying out numerical simulations with various planet masses from 1% of a thermal mass to 10 thermal masses.

Our main findings are:

1. The formation of spiral arms—both primary and additional arms—is a linear process: constructive interference among appropriate sets of wave modes having different azimuthal wavenumbers.
2. A planet excites m evenly spaced wave modes at the mth Lindblad resonance, where $m = 1, 2, 3, \ldots, \infty$ is the azimuthal wavenumber. Among the wave modes the $n = 0$ components, which are those originating from the planet's location, add constructively on to each other and form the primary arm (Sections 2 and 3), confirming the mechanism presented in OL02.
3. Additional spiral arms form in a similar manner to the primary arm, but through constructive interference among non-zero nth components. Non-zero nth components excite out of phase at the Lindblad resonance, in contrast to $n = 0$ components, but constructive interference among the wave modes is possible because wave propagation is dependent upon their azimuthal wavenumber (Equation (7)).
4. The phases of spiral arms follow the dominating azimuthal mode with $m \approx (1/2)(h/r)_p^{-1}$ reasonably well, until they steepen into shocks and depart from the linear regime. We provide a generalized analytic formula in Equation (7), which can be used for the primary arm but also additional arms.
5. In the outer disk, the propagation of wave modes becomes independent on the azimuthal wavenumber m when $r \gg r_p$. Additional spiral arms can thus form in the outer disk only if wave modes become in phase before $r \sim$ a few $\times r_p$ (Section 2.2.2).
6. Spiral arms excited by a planet with more mass propagate at faster speeds and thus appear to be more open,

consistent with previous studies (e.g., Zhu et al. 2015). This trend is seen to be in common with the primary arm and also for additional arms (Figure 13).
7. Only two spiral arms form for sufficiently large planet masses ($M_p \gtrsim 3 M_{\rm th}$). The wave modes that would form a tertiary arm for smaller mass planets merge with the primary arm (Figures 11 and 12).

To conclude, the multiple spiral arm formation mechanism presented in this paper can explain many characteristics of planet-driven spiral arms known from previous studies. An obvious extension is to examine this scenario in three-dimensions, particularly when the background disk is vertically stratified. Because of vertical gravity and/or buoyancy, the wave modes in such disks will behave differently as they depart from the disk midplane. Spiral arms thus may not have a coherent vertical structure, possibly explaining the curvature of spiral arms seen in three-dimensional numerical simulations (e.g., Zhu et al. 2015; Bae et al. 2016b). While we focused on the case in which the primary object (i.e., star) has a much larger mass than the perturbing companion (i.e., planet), the spiral arm formation mechanism presented here can also be applied to the systems where the companion body has a comparable mass to the primary (e.g., disks around dwarf novae or binary black holes), or a much larger mass than the primary (e.g., circumplanetary disks). In the cases when the companion body has a mass comparable to or greater than the primary mass, it is very likely that the companion launches largely open two-armed spirals in the disk around the primary. We present applications of the present work in Paper II and discuss whether various characteristics of observed spiral arms can be used to constrain the masses of yet unseen planets and their positions within their disks.

ORCID iDs

Jaehan Bae ⓘ https://orcid.org/0000-0001-7258-770X
Zhaohuan Zhu ⓘ https://orcid.org/0000-0003-3616-6822

References

ALMA Partnership, Brogan, C. L., Pérez, L. M., et al. 2015, ApJL, 808, L3
Andrews, S., Wilner, D. J., Zhu, Z., et al. 2016, ApJL, 820, L40
Arzamasskiy, L., & Rafikov, R. R. 2018, ApJ, 854, 84
Bae, J., Nelson, R. P., & Hartmann, L. 2016a, ApJ, 833, 126

Bae, J., Nelson, R. P., Hartmann, L., & Richard, S. 2016b, ApJ, 829, 13
Bae, J., & Zhu, Z. 2017, ApJ, 859, 119 (Paper II)
Bae, J., Zhu, Z., & Hartmann, L. 2016c, ApJ, 819, 134
Bae, J., Zhu, Z., & Hartmann, L. 2017, ApJ, 850, 201
Benisty, M., Juhász, A., Boccaletti, A., et al. 2015, A&A, 578, L6
Benítez-Llambay, P., & Masset, F. 2016, ApJS, 223, 11
Brouwer, D., & Clemence, G. M. 1961, Methods of Celestial Mechanics (New York: Academic)
de Val-Borro, M., Edgar, R. G., Artymowicz, P., et al. 2006, MNRAS, 370, 529
Dong, R., & Fung, J. 2017, ApJL, 835, L38
Dong, R., Li, S., Chiang, E., & Li, H. 2017, ApJ, 843, 127
Dong, R., Rafikov, R. R., & Stone, J. M. 2011, ApJ, 741, 57
Dong, R., Zhu, Z., Rafikov, R. R., & Stone, J. M. 2015, ApJL, 809, L05
Fung, J., & Dong, R. 2015, ApJL, 815, L21
Garufi, A., Quanz, S. P., Avenhaus, H., et al. 2013, A&A, 560, A105
Ginski, C., Stolker, T., Pinilla, P., et al. 2016, A&A, 595, A112
Goldreich, P., & Tremaine, S. 1978a, Icar, 34, 240
Goldreich, P., & Tremaine, S. 1978b, ApJ, 222, 850
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
Hord, B., Lyra, W., Flock, M., et al. 2017, ApJ, 849, 164
Isella, A., Guidi, G., Testi, L., et al. 2016, PhRvL, 117, 251101
Ju, W., Stone, J. M., & Zhu, Z. 2016, ApJ, 823, 81
Ju, W., Stone, J. M., & Zhu, Z. 2017, ApJ, 841, 29
Juhász, A., Benisty, M., Pohl, A., et al. 2015, MNRAS, 451, 1147

Kley, W. 1999, MNRAS, 303, 696
Lee, W.-K. 2016, ApJ, 832, 166
Li, H., Colgate, S. A., Wendroff, B., & Liska, R. 2001, ApJ, 551, 874
Li, H., Finn, J. M., Lovelace, R. V. E., & Colgate, S. A. 2000, ApJ, 533, 1023
Li, H., Li, S., Koller, J., et al. 2005, ApJ, 624, 1003
Lin, D. N. C., & Papaloizou, J. C. B. 1993, in Protostars and Planets III, ed. E. H. Levy & J. I. Lunine (Tucson, AZ: Univ. Arizona Press), 749
Lovelace, R. V. E., Li, H., Colgate, S. A., & Nelson, A. F. 1999, ApJ, 513, 805
Maire, A.-L., Stolker, T., Messina, S., et al. 2017, A&A, 601, A134
Muto, T., Grady, C. A., Hashimoto, J., et al. 2012, ApJL, 748, L22
Ogilvie, G. I., & Lubow, S. H. 2002, MNRAS, 330, 950, (OL02)
Pérez, L. M., Carpenter, J. M., Andrews, S. M., et al. 2016, Sci, 353, 1519
Rafikov, R. R. 2002a, ApJ, 569, 997
Rafikov, R. R. 2002b, ApJ, 572, 566
Rafikov, R. R. 2016, ApJ, 831, 122
Reggiani, M., Christiaens, V., Absil, O., et al. 2018, A&A, 611, 74
Richert, A. J., Lyra, W., Boley, A., et al. 2017, ApJ, 804, 95
Ryan, G., & MacFadyen, A. 2017, ApJ, 835, 199
Shu, F. H. 2016, ARA&A, 54, 667
Stolker, T., Dominik, C., Avenhaus, H., et al. 2016, A&A, 595, 113
Tang, Y.-W., Guilloteau, S., Dutrey, A., et al. 2017, ApJ, 840, 32
Tsukagoshi, T., Nomura, H., Muto, T., et al. 2016, ApJL, 829, L35
van Boekel, R., Henning, Th., Menu, J., et al. 2017, ApJ, 837, 132
Walsh, C., Juhász, A., Meeus, G., et al. 2016, ApJ, 831, 200
Ward, W. R. 1986, Icar, 67, 164
Zhang, K., Bergin, E. A., Blake, G. A., et al. 2016, ApJL, 818, L16
Zhu, Z., Dong, R., Stone, J. M., & Rafikov, R. R. 2015, ApJ, 813, 88
Zhu, Z., Ju, W., & Stone, J. M. 2016, ApJ, 832, 193