Gaps and Rings in Protoplanetary Disks with Realistic Thermodynamics: The Critical Role of In-plane Radiation Transport

Ryan Miranda\textsuperscript{1} and Roman R. Rafikov\textsuperscript{1,2}

\textsuperscript{1} Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA; miranda@ias.edu
\textsuperscript{2} Centre for Mathematical Sciences, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

Received 2020 July 27; revised 2020 September 30; accepted 2020 October 5; published 2020 November 26

Abstract

Many protoplanetary disks exhibit annular gaps in dust emission, which may be produced by planets. Simulations of planet–disk interaction aimed at interpreting these observations often treat the disk thermodynamics in an overly simplified manner, which does not properly capture the dynamics of planet-driven density waves driving gap formation. Here we explore substructure formation in disks using analytical calculations and hydrodynamical simulations that include a physically motivated prescription for radiative effects associated with planet-induced density waves. For the first time, our treatment accounts not only for cooling from the disk surface but also for radiation transport along the disk midplane. We show that this in-plane cooling, with a characteristic timescale typically an order of magnitude shorter than the one due to surface cooling, plays a critical role in density wave propagation and dissipation (we provide a simple estimate of this timescale). We also show that viscosity, at the levels expected in protoplanetary disks ($\alpha \lesssim 10^{-3}$), has a negligible effect on density wave dynamics. Using synthetic maps of dust continuum emission, we find that the multiplicity and shape of the gaps produced by planets are sensitive to the physical parameters—disk temperature, mass, and opacity—that determine the damping of density waves. Planets orbiting at $\lesssim 20$ au produce the most diverse variety of gap/ring structures, although significant variation is also found for planets at $\gtrsim 50$ au. By improving the treatment of the physics governing planet–disk coupling, our results present new ways of probing the planetary interpretation of annular substructures in disks.

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

1. Introduction

Recent observations of protoplanetary disks in dust continuum emission using the Atacama Large Millimeter/submillimeter Array (ALMA) have revealed a variety of disk substructures, among them numerous axisymmetric gaps and rings (Huang et al. 2018). A number of mechanisms have been invoked to explain the formation of these substructures. However, planet–disk interaction appears to provide the most promising quantitative description of these features (e.g., Zhang et al. 2018). Recent indirect detections of planets embedded in such disks using gas kinematics (Pinte et al. 2019, 2020) appear to lend support to the planet hypothesis.

In the planet–disk interaction scenario, one or more planets within the disk excite density waves in the surrounding gas via their gravity (Goldreich & Tremaine 1979). Dissipation of the waves leads to the transfer of their angular momentum to the bulk disk material, resulting in the formation of axisymmetric structures in the disk surface density. The resulting pressure maxima are capable of trapping dust, leading to the appearance of rings and gaps in dust continuum emission.

In the classical picture of gap opening by massive planets (e.g., Lin & Papaloizou 1986), damping of the planet-driven waves produces a wide gap around the planetary orbit. For lower-mass planets, at least initially and provided that the disk viscosity is low enough, a planet is expected to produce a pair of narrow gaps situated on either side of its orbit (Rafikov 2002b; Duffell & MacFadyen 2012; Zhu et al. 2013), as a result of nonlinear wave dissipation via shock formation (Goodman & Rafikov 2001; Rafikov 2002a). Recent numerical simulations have shown that the formation of several additional gaps interior to the orbit of the planet is also possible (Bae et al. 2017; Dong et al. 2017, 2018). Such gaps are associated with the splitting of a planet-driven density wave into multiple spiral arms in the inner disk (which is a purely linear process; see Arzamasskiy & Rafikov 2018 and Miranda & Rafikov 2019a) and the subsequent shocking and dissipation of each arm due to nonlinear effects (Bae et al. 2017; Bae & Zhu 2018). Therefore, the multiple rings and gaps seen in many disks may require only a single planet to produce. For example, Zhang et al. (2018) demonstrated that the positions of the five gaps in the AS 209 disk can potentially be accounted for by a single planet.

Until recently, numerical efforts to understand the formation of gaps and rings by planets have largely relied upon a rather simple, locally isothermal treatment of disk thermodynamics, in which the disk temperature is assumed to be a fixed function of the distance from the central star. Recently, we described the shortcomings of this approximation and argued in favor of the more realistic treatment of thermal physics in numerical simulations (Miranda & Rafikov 2019b).

The passage of a density wave generates inhomogeneous compressional heating and cooling of the disk fluid, taking it locally out of equilibrium with the heating (e.g., stellar irradiation) and cooling (e.g., radiative losses from the disk surface) processes operating in an unperturbed disk. Radiative transport associated with these temperature perturbations acts to reestablish thermodynamic equilibrium, thus reducing the restoring force due to pressure and resulting in wave damping. In Miranda & Rafikov (2020) we presented the first study examining the role of such cooling processes in planet–disk coupling.
interaction using linear perturbation theory, demonstrating their strong impact on wave propagation and wave-driven disk evolution. Most importantly, when cooling is neither very slow nor very rapid, it leads to the linear damping of density waves, which is strongest when \( \beta \), the ratio of the cooling timescale to the local orbital timescale, is comparable to the disk aspect ratio \( h \) (see Section 3.4.3 for more details). In this case the multiple-gap structure expected to be produced by a planet in a low-viscosity disk is suppressed, in favor of a single wide gap. Another important result of this work is that the cooling timescale must be very short, typically \( 10^2 - 10^3 \) times shorter than the orbital timescale, for the conventional locally isothermal approximation to provide an accurate description of density wave dynamics.

These findings have been recently corroborated by several numerical studies. In particular, Zhang & Zhu (2020) and Ziampras et al. (2020b) also found the damping of density waves and the suppression of a multiple-gap structure for intermediate cooling timescales, whereas Facchini et al. (2020) noted the importance of cooling for predicting the number and positions of rings in their modeling of the LkCa 15 disk.

In Miranda & Rafikov (2020), we used a cooling prescription in which the dimensionless cooling timescale \( \beta \) is constant across the disk; Zhang & Zhu (2020) and Facchini et al. (2020) took the same approach. While appropriate for a theoretical study of the effects of cooling, this is not a good approximation for real disks. In this paper, we go beyond the constant-\( \beta \) assumption and provide a realistic, radially dependent estimate of cooling due to radiative losses from the surface of the disk, as has also been done in Ziampras et al. (2020b).

However, surface cooling, which relies on vertical energy transport toward the disk surface, is only one way in which radiative effects can affect planet-driven density waves. Localized disk regions experiencing transient compressional heating/cooling by the wave can also radiatively exchange energy in the horizontal direction, i.e., parallel to the disk midplane. Far from the planet, density waves evolve into tightly wrapped, sharp structures with large radial temperature gradients, as shown in Figure 1, where a perturbation due to a transient cooling of the spiral arms along the midplane may be much more efficient than cooling from the disk surface. The role of such in-plane cooling, as we refer to it in this work, has not been explored in detail previously in the context of planet–disk interaction.

In our present study we fully account for the effects of both surface and in-plane cooling on the propagation of density waves and wave-driven disk evolution. Including in-plane cooling is a rather nontrivial exercise, as it is sensitive to the spatial structure of the perturbation, i.e., its wavelength. Here we develop a realistic in-plane cooling prescription that accounts for such details and use it both for semi-analytical calculations and in 2D hydrodynamical simulations (we also propose an approximate analytical expression for the cooling timescale, which can be used for simple estimates; see Section 3.4.1). Our calculations demonstrate that in-plane cooling plays the dominant role in the thermodynamics of density waves and thus cannot be ignored.

Using this approach, we explore the sensitivity of the gap and ring structures produced by a planet to various parameters of the underlying disk—including temperature, surface density, and opacity—that affect the efficiency of cooling. We do this using the results of hydrodynamical simulations, from which we produce simulated dust continuum emission maps.

The plan for this paper is as follows. We first present our basic setup in Section 2. In Section 3, we derive a prescription for the cooling timescale appropriate for planet-driven density waves. We describe the setup and details of our numerical

![Figure 1](image-url)
simulations in Section 4. We present our main results, the gas surface density profiles and dust emission maps for disks with realistic cooling for a variety of disk models, in Section 5. In particular, in Section 5.3.6 we explore the role of viscosity in density wave propagation and disk evolution. In Section 6, we compare the results of our simulations to the results of locally isothermal simulations. We discuss and contextualize our results in Section 7 and summarize and conclude in Section 8.

2. Basic Setup

We consider a thin, two-dimensional gaseous disk around a star of mass $M_*$ that is described in polar coordinates $(r, \phi)$ by the surface density $\Sigma$, height-integrated pressure $P = c_s^2 \Sigma$ (where $c_s = (k_B T/\mu)^{1/2}$ is the (isothermal) sound speed), radial velocity $u_r$, and azimuthal velocity $u_\phi = r \Omega$. The rotation rate $\Omega$ differs from the Keplerian rate $\Omega_K = (GM_*/r^3)^{1/2}$ by $\mathcal{O}(h^2)$, where $h = H/r$ is the disk aspect ratio and $H = c_s/\sqrt{\Omega}$ is the pressure scale height. We neglect the disk self-gravity (its effect was explored by Zhang & Zhu 2020). We primarily consider a disk with a negligible viscosity, although we also carry out a subset of simulations with explicit viscosity, which we parameterize via the conventional dimensionless $\alpha$-parameter (Shakura & Sunyaev 1973).

The disk temperature follows a power-law profile with $T(r) \propto r^{-q}$. It is specified indirectly through the aspect ratio $h(r)$, parameterized by its value at $r = 50$ au, according to

$$h(r) = h_{50 \text{ au}} r^{(1-q)/2},$$

where $r_{50} = r/(50 \text{ au})$. The surface density profile is similarly parameterized in terms of its value at 50 au and the power-law index $p$:

$$\Sigma(r) = \Sigma_{50 \text{ au}} r^{-p}.$$  \hspace{1cm} (2)

To compute the thermodynamic properties of the disk, we adopt the opacity (due to dust grains) of Bell & Lin (1994), appropriate for the cold ($T < 170$ K) outer regions of protoplanetary disks:

$$\kappa_d = \kappa_{d,0} T^2 = 0.38 M_{\odot}^2 \kappa_{d,0} \rho_0 h_{0.1}^{-4/3} r_{50}^{-2} \text{ cm}^2 \text{ g}^{-1},$$

where $M_{\odot} = M_*/M_\odot$ and $h_{0.1} = h(r)/0.1$. We also introduce $\kappa_{d,0}$, the constant $\kappa_{d,0}$ scaled by its fiducial value $2 \times 10^{-4} \text{ cm}^2 \text{ g}^{-1} \text{ K}^{-2}$, which is used to scale the overall opacity of the disk material (which may depend on metallicity). As a result of our opacity law, the optical depth (from the midplane to the surface) of the disk is

$$\tau = \frac{1}{2} \kappa_d \Sigma = 1.9 M_{\odot}^2 \kappa_{d,0} \rho_0 h_{0.1}^{-4/3} \Sigma_{10} r_{50}^{-2},$$

where $\Sigma_{10} = \Sigma(r)/(10 \text{ g cm}^{-2})$. In the numerical estimates above, we assume a mean molecular weight $\mu = 2m_p$.

The temperature and surface density profiles, along with the opacity $\kappa_d$, fully define our disk model for the purposes of computing the cooling timescale. The disk model is therefore fully specified by six parameters: $h_{50 \text{ au}}, q, \Sigma_{50 \text{ au}}, p, \kappa_{d,0},$ and $\alpha$. In describing our results we consider 12 different disk models, which are summarized in Table 1. The fiducial disk model has $h_{50 \text{ au}} = 0.1, q = 1/2, \Sigma_{50 \text{ au}} = 10 \text{ g cm}^{-2}, p = 1, \kappa_{d,0} = 1$, and $\alpha = 0$. The disk mass is varied (by varying $\Sigma_{50 \text{ au}}$) in the low-mass and high-mass models. The temperature (through $h_{50 \text{ au}}$) is varied in the hot and cold models. The opacity is varied by means of $\kappa_{d,0}$ in the low-opacity and high-opacity models. The temperature and surface density power-law indices $q$ and $p$ are varied in the $T \propto r^{-1/4}$ and $\Sigma \propto r^{-1/2}$ models, respectively. The effects of a nonzero viscosity are explored in the $\alpha = 10^{-4}$ and $10^{-3}$ models.

For each disk model, we consider planets with a range of different masses $M_p$ and orbital radii $r_p$. For $r_p$ we choose 10, 20, 50, and 100 au, spanning the range of radii at which most of the rings and gaps in protoplanetary disks are observed (Huang et al. 2018). Note that for each disk model, $h(r)$ and $\Sigma(r)$ are defined as functions of $r$, not $r/r_p$. The aspect ratio and surface density at the location of the planet $h_p$ and $\Sigma_p$ are therefore $h_{50 \text{ au}} = h_{50 \text{ au}}^{(1-q)/2}$ and $\Sigma_{50 \text{ au}} = \Sigma_{50 \text{ au}}^{p}$ (where $r_p_{50 \text{ au}} = r_p/50$ au), respectively.

The planet mass $M_p$ is parameterized in terms of the thermal mass (Goodman & Rafikov 2001),

$$M_{th} = h_p^3 M_p = 1M_J \left( \frac{M_p}{M_J} \right) \left( \frac{h_p}{0.1} \right)^3.$$  \hspace{1cm} (5)

The ratio $M_p/M_{th}$ describes the nonlinearity of disk response. Note that since we choose the planet mass to be a fixed fraction of $M_{th}$, the planet mass varies with $r_p$, as $M_p = h_{50 \text{ au}}^3 3^{(1-q)/2}$, for a fixed $M_p/M_{th}$.

3. Cooling Prescription

In this section we provide a detailed description of the procedure that we use to compute the thermal relaxation timescale for density waves propagating through protoplanetary disks. In Appendix A we derive the following 2D form of the energy equation for a disk perturbed by the passage of a density wave, accounting for radiative effects:

$$\frac{de}{dt} + P \frac{d}{dt} \left( \frac{1}{\Sigma} \right) = \left( \frac{\partial e}{\partial t} \right)_\text{surf} + \left( \frac{\partial e}{\partial t} \right)_\text{mid},$$

$$\left( \frac{\partial e}{\partial t} \right)_\text{surf} = -2 \sigma T^4 \kappa \varepsilon_r,$$  \hspace{1cm} (7)

$$\left( \frac{\partial e}{\partial t} \right)_\text{mid} = - \nabla \cdot \mathbf{F}.$$  \hspace{1cm} (8)

Here $e = c_v T$ and $P = (\gamma - 1) c_v e$ are the 2D (properly vertically averaged) internal energy (per unit mass) and pressure, and $\varepsilon_r$ is the wave-driven perturbation of $e$ from its equilibrium value $e_0$. The two terms in the right-hand side of (6) represent the energy source/sink terms\textsuperscript{3} associated with the transport of radiation caused by the passage of a density wave. The one given by Equation (7) accounts for radiative losses from the disk surface, with $T_{eff}$ being the effective temperature of the disk in equilibrium. The term given by Equation (8) accounts for the transport of radiation along the disk midplane, with $\mathbf{F}$ being the density wave–induced perturbation of the (vertically integrated) component of the radiative flux along the plane of the disk.

Using this form of the energy equation, we first provide a calculation of the cooling timescale due to radiative losses from the surface of the disk (Section 3.1), followed by the cooling timescale associated with radiative energy transfer along the disk midplane (Section 3.2).
The surface cooling term (7) can be directly rewritten in the form

$$\frac{1}{\Omega} \left( \frac{\partial \delta e}{\partial t} \right)_{\text{surf}} = -\frac{\delta e}{\beta_{\text{surf}}}$$

(9)

Here we introduce the cooling timescale due to thermal continuum dust emission from the disk surface:

$$t_{\text{surf}} = \frac{\Sigma e}{2\epsilon T_{\text{eff}}^4},$$

(10)

and the dimensionless cooling timescale $\beta_{\text{surf}} = \Omega t_{\text{surf}}$ (in the following we will often omit the word “dimensionless” and simply refer to $\beta$ as a cooling timescale). The factor of two in the denominator accounts for the fact that the disk cools from both its upper and lower surfaces. In the above, we do not distinguish between $\epsilon$ and $\epsilon_0$, since the cooling term (9) is already linear in a perturbed variable $\delta \epsilon$.

The effective temperature $T_{\text{eff}}$ is related to the midplane temperature $T$ according to Hubeny (1990):

$$T_{\text{eff}}^4 = \tau_{\text{eff}}^{-1} T^4,$$

(11)

where

$$\tau_{\text{eff}} = \frac{3\tau}{8} + \frac{\sqrt{3}}{4} + \frac{1}{4\tau}$$

(12)

is the effective optical depth. Note that the cooling timescale (10) combined with the relations (11) and (12) should be valid not only for self-luminous disks but also for externally irradiated ones (at least approximately). In particular, our $t_{\text{surf}}$ would reduce (up to factors of order unity) to the cooling time $t_{\text{cool}}$ derived in Zhu et al. (2015), namely their Equation (8), in which one needs to set $T_e = T_{\text{surf}}$. However, the precise form of $t_{\text{surf}}$ is not especially important, since, as we show later, surface cooling is not the dominant factor in density wave thermodynamics. With this consideration, we have

$$\beta_{\text{surf}} = \beta_{\text{surf},0}(\tau),$$

(13)

where

$$\beta_{\text{surf},0} = \frac{c \nu \Omega}{4 \kappa \Delta \sigma T^5} = 8.7 \times 10^{-3} M_{\text{p},1}^{-9/2} \kappa_{0,0}^{-1} h_{0,1}^{-10} r_{50}^{7/2}$$

(14)

is the surface cooling timescale in the optically thin regime ($\tau \ll 1$) and

$$f(\tau) = 1 + \sqrt{3(\tau + \frac{3}{2}\tau^2)}.$$ 

(15)

The numerical estimate in Equation (14) assumes an adiabatic index $\gamma = 7/5$. Note that $h(r) \propto r^{(1-\gamma)/2}$, so the radial dependence of $\beta_{\text{surf},0}$ is not determined solely by the $r_{50}^{7/2}$ term. For a fiducial temperature profile with $q = 1/2$, $\beta_{\text{surf},0} \propto r$.

Note that $f(\tau) \approx 1$ for $\tau \ll 1$ and that $f(\tau) > 1$ always. Therefore, $\beta_{\text{surf},0}$ is a lower limit for the surface cooling timescale.

### 3.2. In-plane Cooling

In addition to cooling from the disk surface, thermal relaxation occurs through radiative energy transfer along the midplane of the disk. Its contribution to the energy evolution is described by Equation (8), in which we need to specify the explicit form of the (vertically integrated) flux perturbation $\delta F$. The form of this perturbation, and hence of the resulting cooling term, depends on the ratio $L_e/L_{\text{rad}}$, where $L_e = |\delta \epsilon|/|\nabla \delta \epsilon|$ is the perturbation lengthscale and

$$L_{\text{rad}} = \frac{1}{\kappa \rho} = 3.3 M_{\text{p},1}^{-2} \kappa_{0,0}^{-1} h_{0,1}^{-3} \tau_{10}^{-1} r_{50}^{-3}$$

(16)

is the radiative lengthscale or photon mean free path length. In the above estimate, we take a 3D density $\rho$ equal to the
midplane density $\rho_{\text{mid}} = \Sigma/(\sqrt{2}\pi H)$ (assuming the disk is vertically isothermal).

In general, computing $\delta F$ from the full radiation transfer equation is highly nontrivial but can be bypassed by using an approximate theory, e.g., flux-limited diffusion (FLD; Levermore & Pomraning 1981; Kley 1989). However, the behavior of $\delta F$ can be understood in two important limits: the diffusion limit $l_r \gg l_{\text{rad}}$ and the streaming limit $l_r \ll l_{\text{rad}}$.

### 3.2.1. Diffusion Limit

For $l_r \gg l_{\text{rad}}$, we show in Appendix B that the vertically integrated flux perturbation takes the form

$$\delta F = -\Sigma\eta \nabla \delta e, \quad (17)$$

where

$$\eta = \frac{16\sigma T^3}{3\kappa_4 \rho^2 c_v} \quad \text{(18)}$$

is the radiative diffusion coefficient (here we again assume $\rho = \rho_{\text{mid}}$). As a result of specifying $\delta F$ via Equation (17), the in-plane cooling term (Equation (8)) reads

$$\left( \frac{\partial}{\partial t} \right)_{\text{mid}} = \frac{1}{\Sigma} \nabla \cdot (\Sigma \eta \nabla \delta e). \quad (19)$$

One can see that in this regime, the in-plane cooling contribution is indeed diffusive, depending on the specific form of $\delta e$ through its second-order spatial derivative. This raises a serious problem for trying to understand the density wave dynamics in the well-defined limit of linear perturbations, an approximation that has been successfully used for this purpose in the past (Goldreich & Tremaine 1980; Miranda & Rafikov 2019a, 2020).

In linear theory, we assume an expansion of $\delta e$ in Fourier harmonics:

$$\delta e(r, \phi) = \sum_m \delta e_m(r) \exp(im\phi), \quad (20)$$

where $\delta e_m(r)$ is a complex quantity describing the amplitude and phase of the perturbation with azimuthal number $m$. As a result of this expansion, Equation (19) then describes the evolution of each harmonic $\delta e_m$, rather than the total perturbation $\delta e$.

An analogous Fourier expansion (20) is assumed for all other perturbed fluid variables. As demonstrated in Miranda & Rafikov (2020), one can then manipulate the fluid continuity and momentum equations into a single master equation describing the global behavior of linear perturbations. If the cooling term depends linearly on $\delta e$, as in Equation (9), then this master equation ends up being a second-order ordinary differential equation (ODE) in the radial coordinate, which has been studied in detail in Miranda & Rafikov (2020).

However, with the in-plane cooling contribution in the form (19) the perturbations of, e.g., pressure and surface density, would necessarily involve second-order radial derivatives of the energy (or enthalpy) perturbation. Propagating this dependence through the linear perturbation analysis therefore results in a fourth-order master ODE rather than a second-order equation as in Miranda & Rafikov (2020). This significantly increases the complexity of the analysis and numerical solutions of the master equation.

To avoid this complication in our present work and to fully benefit from the mathematical framework and physical understanding developed in Miranda & Rafikov (2020), we resort to a different, approximate approach. Instead of taking full spatial derivatives in the right-hand side of Equation (19), we evaluate them using the local Wentzel–Kramers–Brillouin (WKB) approximation, assuming $l_{r,m} \ll r$ (which is a good approximation, as demonstrated by Figure 1). Here $l_{r,m} = [\delta e_m]/[\nabla \delta e_m]$ is the lengthscale associated with a particular Fourier harmonic of the perturbation. As a result of making this assumption, Equation (19) for a Fourier harmonic of the energy perturbation $\delta e_m$ takes the simple form

$$\left( \frac{\partial}{\partial t} \right)_{\text{mid},m} \approx -\frac{\delta e_m}{t_{\text{diff},m}}, \quad (21)$$

where

$$t_{\text{diff},m} = \frac{l_{r,m}^2}{\eta} \quad \text{(22)}$$

is the in-plane cooling timescale for the $m$th Fourier harmonic of the perturbation in the diffusive limit. Note that $t_{\text{diff},m}$ explicitly depends on the radial structure of the perturbation due to the wave, i.e., on the wavenumber $m$. Nevertheless, we can still deal with cooling in the form (21) directly using the framework of Miranda & Rafikov (2020), which allows one to describe density wave propagation using a single second-order ODE.4.

### 3.2.2. Streaming Limit and Interpolation

In the opposite streaming limit, for $l_r \ll l_{\text{rad}}$, we show in Appendix B that the vertically integrated flux perturbation approximately satisfies

$$\nabla \cdot \delta F = \frac{16\kappa_4 \Sigma \sigma T^3}{3c_v} \delta e. \quad (23)$$

Using definitions (16) and (18), we can then transform Equation (8) into

$$\left( \frac{\partial}{\partial t} \right)_{\text{mid}} \approx -\frac{\delta e}{t_{\text{stream}}}, \quad (24)$$

where

$$t_{\text{stream}} = \frac{l_{\text{rad}}^2}{3\eta} \quad \text{(25)}$$

is the timescale for in-plane cooling in the streaming limit (Lin & Youdin 2015; Malygin et al. 2017). Note that $t_{\text{stream}}$ is independent of the perturbation lengthscale.

In the general situation, following Lin & Youdin (2015), we express the midplane cooling timescale for the $m$th Fourier harmonic of the perturbation as $t_{\text{mid},m} \approx t_{\text{stream}} + t_{\text{diff},m}$, which captures the correct behavior in both the diffusive and streaming limits. The dimensionless timescale for in-plane cooling...
cooling $\beta_{\text{mid},m} = \Omega_{\text{mid},m}$ is then, using Equations (22) and (25),

$$\beta_{\text{mid},m} \approx \frac{\Omega}{\eta} \left( \frac{l^2_{e,m} + \frac{1}{3} l^2_{\text{rad}}}{} \right).$$  \hspace{1cm} (26)

Once again, the subscript $m$ denotes that the in-plane cooling timescale $\beta_{\text{mid},m}$ explicitly depends on the azimuthal wavenumber of the perturbation through the lengthscale $l_{e,m}$, which we derive next.

### 3.2.3. Perturbation Lengthscale

For a density wave with azimuthal number $m$, the (inverse) perturbation lengthscale is

$$l_{e,m}^{-1} = \left( k_m^2 + \frac{m^2}{r^2} \right)^{\frac{1}{2}} \approx k_m,$$

where $k_m(r)$ is the radial wavenumber for the $m$th Fourier harmonic (this notation is consistent with Miranda & Rafikov 2019a). The second (approximate) equality above follows from the fact that the radial wavelength $k_m^{-1}$ is shorter than the azimuthal wavelength $m/r$ by a factor of $\approx h$ after the wave has propagated far enough away from the planet. Indeed, Figure 1(b) shows that $\delta r \sim \frac{r}{\delta \phi} \sim H$ only within about one scale height of the planet (where in-plane cooling is not dominant anyway). At larger distances, beyond about $2H$ from the planet, the radial width becomes much smaller, with $\delta r \approx 0.2H$. On the other hand, the azimuthal width $\delta \phi$ remains roughly constant, so that $\delta r/(r\delta \phi) \ll 1$. The spiral arm is therefore much narrower in the radial direction than in the azimuthal direction for $|r - r_p| \gtrless H$, and so $l_{e,m}^{-1}$ can be approximated by $k_m$.

For planet-driven density waves, the radial wavenumber is well approximated by its expression in the WKB limit,

$$k^2_m = k^2_m - \frac{1}{H^2},$$

where

$$k_m = m \frac{\Omega_p}{H} \frac{\Omega - 1}{\Omega}$$

is the approximate wavenumber for $m \gg 1$ (e.g., Ogilvie & Lubow 2002; Miranda & Rafikov 2019a). From Equation (28) we see that $k^2_m > 0$ for $r > r_c$ and $r < r_c$, where $r_c$ is the location of the outer/inner Lindblad resonance (LR), defined by $k_m(r_c) = 0$. Between the LRs, density waves are evanescent, with $k^2_m < 0$. Estimating the perturbation lengthscale as $l_{e,m} \approx k_m^{-1}$ using Equation (28) would then result in an unphysical negative timescale for diffusive in-plane cooling (Equation (22)) in the evanescent zone.

However, note that $k^2_m$ (Equation 29) is positive everywhere. Furthermore, it differs from the actual (WKB) $k^2_m$ (Equation 28) only by $O(m^{-2})$ far from the LRs. For planet-driven waves, which are dominated by harmonics with $m \sim h_p^{-1}$ (Miranda & Rafikov 2019a), this difference is $O(h_p^2)$ (also note that the $m = 1$ harmonic is very weak). Therefore, $k_m$ is a good approximation for $k_m$ outside of the LRs. We will henceforth approximate the perturbation lengthscale of planet-driven density waves as $l_{e,m} \approx k^{-1}_m$.

### 3.3. Combined Cooling Timescale

Taking into account both cooling from the disk surface with timescale $\beta_{\text{surf}}$ (Equation (13)) and in-plane cooling with timescale $\beta_{\text{mid},m}$ (Equation (26)), the cooling/thermal relaxation part of the evolution equation for a Fourier harmonic of the specific internal energy perturbation $\delta e_m$ reads

$$\frac{1}{\Omega} \left( \frac{\partial \delta e_m}{\partial t} \right)_{\text{cool}} = -\frac{\delta e_m}{\beta_{\text{surf}}} - \frac{\delta e_m}{\beta_{\text{mid},m}}.$$  \hspace{1cm} (30)

We can rewrite this as

$$\frac{1}{\Omega} \left( \frac{\partial \delta e_m}{\partial t} \right)_{\text{cool}} = -\frac{\delta e_m}{\beta_m},$$

where we define the cooling timescale resulting from the combined effects of both surface and in-plane cooling,

$$\beta_m = \left( \frac{1}{\beta_{\text{surf}}} + \frac{1}{\beta_{\text{mid},m}} \right)^{-1}.$$  \hspace{1cm} (32)

Making use of the fact that $\beta_{\text{surf}}/l_{\text{rad}}^2 = 4/3$ (see Equations (14), (16), and (18)), the cooling timescale (Equation (32)) can be expressed as

$$\beta_m = \beta_{\text{surf}} \left[ 1 + \frac{4\beta(\tau)}{1 + 3(k_m l_{\text{rad}})^2} \right]^{-1},$$

where

$$k_m l_{\text{rad}} = \left( \frac{\pi}{2} \right)^{\frac{1}{2}} \frac{\Omega_p}{\Omega} \frac{\Omega - 1}{\tau} \frac{m}{\tau}$$

is the ratio of the radiative lengthscale (Equation 16) to the (approximate) radial wavelength of the density wave Fourier harmonic (Equation 29). Equations (26) and (34) indicate that the transition between the diffusive and streaming limits for the in-plane cooling (taking place when $k_m l_{\text{rad}} \sim 1$) occurs at $\tau \approx m$ in the inner disk, provided that $|\Omega - \Omega_p| \sim \Omega$. Also note that $\beta_m$ is, in general, a function of the orbital distance $r_p$ of the planet through the factor $\Omega_p/\Omega$, which also ensures that $\beta_m(r_p) = \beta_{\text{surf}}$.

Equation (33) represents our final working expression for the cooling timescale associated with a particular Fourier harmonic of the density wave perturbation.

#### 3.3.1. Limiting Behaviors

The behavior of the cooling timescale (Equation 33) depends on two key parameters: the optical depth $\tau$ and the ratio between the radiative and perturbation lengthscales $k_m l_{\text{rad}}$. Except within the immediate vicinity of the planet, the value of $k_m l_{\text{rad}}$ is mostly determined by $m/\tau$ (see Equation (34)). Therefore, $\beta_m$ is largely determined by $\tau$. Its behavior can be broken down into several different regimes, delineated by whether $\tau$ is $\gg m$, $m$, $1$, or $\ll 1$.

For $\tau \gg m$, in the diffusive cooling limit, we have

$$\beta_m \approx \beta_{\text{surf}} \left[ 1 + \pi m^2 \left( \frac{\Omega_p}{\Omega} - 1 \right)^2 \right]^{-1},$$

(35)
or roughly \( \beta_m \sim \beta_{\text{surf}}/m^2 \) for \( |\Omega - \Omega_p| \sim \Omega \). This reduction of the cooling timescale by a factor of \( m^2 \) as compared to the surface cooling timescale \( \beta_{\text{surf}} \) is easy to understand by examining Equation (29) and noticing that the characteristic radial scale of the wave-like temperature perturbation far from the planet is \( \sim m^{-1}H \). This is considerably shorter than the vertical disk thickness \( H \) (see Figure 1), so that the radiation needs to diffuse through to reach the disk surface, leading to in-plane cooling being a factor of \( m^2 \) faster (in the diffusive regime) than surface cooling.

For \( \tau \ll m \), in the streaming limit, we have

\[
\beta_m \approx \frac{\beta_{\text{surf}}}{1 + 4f(\tau)} \approx \frac{\beta_{\text{surf}}}{1 + 4f(\tau)},
\]

which is independent of \( m \).

Furthermore, in the conventional optically thin limit \( \tau \ll 1 \), we have \( f(\tau) \approx 1 \) and \( \beta_{\text{surf}} \approx \beta_{\text{surf},0} \), so that

\[
\beta_m \approx \frac{1}{5} \beta_{\text{surf}} \approx \frac{1}{5} \beta_{\text{surf},0}.
\]

This is the result of an equivalence (to within a constant factor) between the timescales for cooling from the disk surface (Equation (13)) in the optically thin limit and for cooling in the midplane (Equation (26)) in the streaming limit. The factor of one-fifth indicates that the combined effect of surface and in-plane optically thin cooling is faster cooling than either of their effects alone. Whereas a more careful, fully 3D treatment of the radiation transfer would likely find the numerical prefactor to deviate from one-fifth, the difference is not fundamental. It is a consequence of assumptions made about other approximate prefactors in our calculation of the cooling timescale.

Finally, when \( \xi_m \approx 0 \)—which occurs only at \( r = r_p \), regardless of the optical depth—we have \( \beta_m = \beta_{\text{surf}} \). Therefore, \( \beta_{\text{surf}} \) sets an upper limit for \( \beta_m \), i.e., we always have \( \beta_m \leq \beta_{\text{surf}} \).

### 3.3.2. Radial Profiles of \( \beta_m \)

Figure 2(a) shows profiles of \( \beta_m \) (Equation (33)) for a range of azimuthal numbers \( m \) for the density waves launched by a planet at \( 50 \) au using our fiducial disk model (see Section 2 and Table 1 for details). These serve to illustrate the different aspects of the behavior of \( \beta_m \) outlined above.

The dashed line in Figure 2(a) shows the profile of the surface cooling timescale \( \beta_{\text{surf}} \) (Equation (13)). Each \( \beta_m \) (differently colored solid line) is equal to \( \beta_{\text{surf}} \) at the planetary orbital radius \( r_p = 50 \) au. Everywhere else, \( \beta_m < \beta_{\text{surf}} \), as expected. This roughly indicates that surface cooling may play a role in the excitation of density waves (by determining the total torque absorbed by the wave; see Miranda & Rafikov 2020), which takes place within \((1-2)H \) from the planet where \( mL(\Omega - \Omega_p)/\Omega \lesssim 1 \). However, midplane cooling dominates the subsequent propagation and evolution of the waves.

Far from the planet, the behavior of the midplane cooling depends on the optical depth \( \tau \), which is shown in Figure 2(b). For each \( \beta_m \) shown in Figure 2(a), a filled point indicates the radius at which \( \tau = m \), which marks a transition in the behavior of the in-plane cooling from diffusive (for \( \tau > m \)) to streaming (\( \tau < m \)). In the diffusive region, to the left of the \( \tau = m \) radius, \( \beta_m \) is described by Equation (35) and hence is smaller than \( \beta_{\text{surf}} \) by \( \sim m^2 \). In the streaming region, to the right of the \( \tau = m \) radius, \( \beta_m \) is described by Equation (36) and hence should be independent of \( m \). In practice, however, in the model presented in Figure 2, the interval \( 1 \leq \tau \lesssim m \) falls in the radial range where \( \Omega \) is rather close to \( \Omega_p \). As a result, \( \xi_m \) ends up being not very large (see Equation (33)), and some sensitivity to \( m \) gets retained in the expression (34) for \( \beta_m \), as we see in Figure 2(a).

At sufficiently large radii in the outer disk (\( r \gtrsim 100 \) au for the disk model in Figure 2), \( \tau \ll 1 \), and hence \( \tau \ll m \) for all \( m \). In this region, \( \beta_m \) is in the streaming regime and is therefore independent of \( m \), for all \( m \) (see Equations (36) and (37)). Hence all of the \( \beta_m \) converge to the same profile, as they are all proportional to \( \beta_{\text{surf}} \) with the same constant of proportionality of one-fifth for \( \tau \ll 1 \).

We see that the different \( \beta_m \) span a wide range of values in the model shown in Figure 2(a), from \( 10 \) for \( m = 1 \) in the inner disk to \( \lesssim 0.01 \) (for all \( m \)) in the outer disk. This indicates that the effect of cooling varies greatly in different regions of the disk and for different perturbation harmonics. The large spread in \( \beta_m \) in the inner disk suggests that understanding the role of cooling in planet-driven density waves made up of many Fourier harmonics may be highly nontrivial.

### 3.4. Effective Cooling Timescale

Equation (33) describes the thermal relaxation timescale for a single Fourier harmonic with azimuthal number \( m \). In general, this timescale is different for each harmonic. The specific internal energy perturbation \( \delta e_m \) for each Fourier harmonic of a

![Figure 2](image-url)
density wave is then damped on its own individual timescale \( \beta_m \). In the numerical simulations in this work, we evolve each harmonic independently in this fashion, even when the density wave is nonlinear (see Section 4.2 and Appendix C). However, to facilitate the interpretation of our results it is useful to define an effective cooling timescale \( t_{\text{eff}} \) or dimensionless \( \beta_{\text{eff}} = \Omega t_{\text{eff}} \), which gives the approximate relaxation timescale for the perturbation of the total internal energy, which is made up of many harmonics.

First, using Parseval’s theorem, we relate the total internal energy perturbation \( \delta e \) to its Fourier harmonics \( \delta e_m \):

\[
\frac{1}{2\pi} \oint |\delta e|^2 d\phi = \sum_m |\delta e_m|^2.
\]  

Differentiating (38) and making use of the cooling law for the individual harmonics (Equation (31)), we have

\[
\frac{1}{4\pi \Omega} \frac{\partial}{\partial t} \oint |\delta e|^2 d\phi = -\beta_{\text{eff}}^{-1} \sum_m |\delta e_m|^2,
\]  

where we define the effective cooling timescale:

\[
\beta_{\text{eff}}^{-1} = \sum m \beta_m^{-1} |\delta e_m|^2 \left( \sum_m |\delta e_m|^2 \right)^{-1}.
\]  

Using Equation (38) once again, we can rewrite Equation (39) as

\[
\oint |\delta e| \left( \frac{1}{\Omega} \frac{\partial |\delta e|}{\partial t} + \frac{|\delta e|}{\beta_{\text{eff}}^*} \right) d\phi = 0,
\]  

which shows that \( \beta_{\text{eff}} \) indeed plays the role of a dimensionless cooling timescale for the total energy perturbation.

Notice that \( \beta_{\text{eff}}(r) \) is a weighted harmonic mean of the cooling timescales \( \beta_m \), with weights given by the squared Fourier amplitudes \( |\delta e_m(r)|^2 \). In order to calculate it, we need the specific form of \( \delta e_m(r) \) for all harmonics of interest. As a benchmark, we determine these using numerical solutions of the linear perturbation equations (see Section 4.1) previously developed in Miranda & Rafikov (2020), which are a well-defined way of fixing \( \delta e_m(r) \). For massive planets, \( \delta e_m(r) \) is modified by nonlinear effects (which are fully captured in our numerical simulations). Nonetheless, computing \( \beta_{\text{eff}} \) using the linear \( \delta e_m(r) \) provides a useful diagnostic for the role of cooling in the behavior of planet-driven density waves and reduces it to the true effective cooling timescale in the limit of \( M_p << M_\text{th} \).

We will not discuss in detail the dependence of the linear \( \delta e_m \) on \( r \). The most salient features of the behavior of \( \delta e_m(r) \) can be roughly understood by examining the radial profiles of the \( P_m(r) \)—the contributions of the different Fourier harmonics to the angular momentum flux (AMF)—derived in Miranda & Rafikov (2020). While connecting \( P_m(r) \) profiles to \( \delta e_m(r) \) it should be remembered that \( \delta e_m \) is a linear function of the enthalpy perturbation \( \delta h_m \), whereas \( P_m(r) \) depends on \( \delta h_m \) quadratically. Also, the results in Miranda & Rafikov (2020) were obtained for a radially constant and \( m \)-independent cooling time \( \beta \), which is clearly different from the \( \beta_m \) given by Equation (33). Nevertheless, one can still use the \( P_m(r) \) profiles in Miranda & Rafikov (2020) to get a qualitative idea of how the radial variation of \( \delta e_m(r) \) might impact the calculation of \( \beta_{\text{eff}}(r) \).

The effective cooling timescale \( \beta_{\text{eff}} \) for different disk models (see Table 1) is shown in Figure 3. Profiles of \( \beta_{\text{eff}}(r) \) for waves driven by planets orbiting at four different radii \( r_p \), from 10 au to 100 au, are shown by the solid colored curves for each disk model.

We see that \( \beta_{\text{eff}} \) is equal to \( \beta_{\text{surf}} \) (dashed black curve in each panel of Figure 3) at \( r = r_p \) and is strictly less than \( \beta_{\text{surf}} \) everywhere else. The factor by which \( \beta_{\text{eff}} \) is smaller than \( \beta_{\text{surf}} \) depends on whether in-plane cooling operates in the diffusive or streaming limit. Far from the planet in the outer, optically thin part of the disk, \( \beta_{\text{eff}} \) converges to the behavior given by Equation (37), which is independent of \( m \). Far inside of the planetary orbit, in the optically thick part of the disk, the \( \beta_{\text{eff}}(r) \) curves tend to show a universal behavior independent\(^5 \) of \( r_p \), which is due to \( \Omega \gg \Omega_p \) there, making \( \beta_m \) independent of \( r_p \) (see Equation (35)).

3.4.1. Simple Estimate for \( \beta_{\text{eff}} \)

For a single Fourier harmonic, the character of the in-plane cooling depends on \( \tau \) and \( m \), but the density waves are composed of multiple such harmonics. The density waves launched by the planet are initially dominated by perturbations with an azimuthal wavenumber \( m = m_\ast \approx (2\pi a_p)^{-1} \) (Miranda & Rafikov 2019a). In the fiducial disk model, \( m_\ast \approx 4 \sim 7 \) for \( r_p \) in the range 10–100 au (it is roughly in this range for our other disk models as well). To provide an idea of the behavior of \( \beta_{\text{eff}}(r) \), we can assume for simplicity that \( \delta e_{m_\ast} \) —the \( m = m_\ast \) harmonic of internal energy—is the only nonzero \( \delta e_m \) in Equation (40). In other words, we approximate \( \beta_{\text{eff}} \) with \( \beta_{m_\ast} \), defined as the value of \( \beta_m \) (see Equation (33)) with \( k_m \ell_{\text{rad}} \) computed by setting \( m = m_\ast \) in Equation (34).

The dotted curves in Figure 3 show the behavior of \( \beta_{m_\ast} \) defined in such a way, and we see that in many cases \( \beta_{m_\ast} \) provides a reasonably good estimate for the full \( \beta_{\text{eff}} \). Typically the two timescales differ by at most a factor of a few,\(^6 \) which is small compared to the 3–4 orders of magnitude over which \( \beta_{\text{eff}} \) intrinsically varies.

The approximation \( \beta_{\text{eff}} \approx \beta_{m_\ast} \) simplifies our interpretation of the behavior of \( \beta_{\text{eff}} \) and allows one to make simple estimates of the role of in-plane cooling. When \( \tau/m_\ast \gg 1 \), in-plane cooling is mostly diffusive, and far from the planet \( \beta_{m_\ast} \) reduces to

\[
\beta_{m_\ast} \approx \frac{4}{\pi} \frac{h_p^2 \beta_{\text{surf}}}{\Omega_p} \left( \frac{\Omega - 1}{\Omega} \right)^2
\]  

(see Equation (35)). In other words, in this limit the cooling timescale \( \beta_{m_\ast} \) for a mode with \( m \approx m_\ast \) is shorter than \( \beta_{\text{surf}} \) by a factor \( m_\ast^{-2} \sim h_p^2 \). Thus, \( \beta_{m_\ast} \) is at least an order of magnitude smaller than \( \beta_{\text{surf}} \) in this regime, in agreement with Figure 3.

On the other hand, when \( \tau/m_\ast \lesssim 1 \), in-plane cooling operates in the streaming regime; hence \( \beta_{m_\ast} \) is given by Equation (36). The location at which \( \tau = m_\ast \), where the transition between these two regimes occurs, is indicated by the arrows at the top of each panel in Figure 3.

\(^5 \) This is not true in the diffusive limit for \( r > r_p \) since \( \Omega_p \beta \Omega \) is not small there.

\(^6 \) More substantial deviations of \( \beta_{m_\ast}(r) \) from \( \beta_{m_\ast}(r) \) visible in Figures 3(c), (e), (g) are caused by the strong radiative damping of the \( m = m_\ast \) harmonic of the perturbation, which invalidates our assumption of this Fourier mode dominating \( \delta e_{m_\ast} \).
3.4.2. Dependence of $\beta_{\text{eff}}$ on Disk Parameters

Next we discuss how the effective cooling timescale $\beta_{\text{eff}}$ changes as we vary the parameters of the disk model away from their values in the fiducial model.

An examination of Figure 3 reveals that $\beta_{\text{eff}}$ is very sensitive to the disk temperature (or aspect ratio $h$; see panels (g), (h)). For the fiducial disk model ($h_{50,\text{au}} = 0.1$), the minimum value of $\beta_{\text{eff}}$ is $\approx 10^{-3}$, while for the cold model ($h_{50,\text{au}} = 0.07$), the minimum value is $\approx 0.05$, and for the hot model ($h_{50,\text{au}} = 0.15$), the minimum value drops to $\approx 10^{-4}$. This dramatic variation stems from the fact that the minimum of $\beta_{\text{eff}}$ is typically attained at the transition to the free-streaming regime of the in-plane cooling (at $\tau \approx m_*$). In the free-streaming limit of in-plane cooling, according to Equation (36), $\beta_{\text{eff}}$ depends primarily on the timescale for surface cooling in the optically thin limit $\beta_{\text{surf},0}$ (with only weak additional dependence on $\tau$), which has a very steep scaling with the disk aspect ratio $h$, $\beta_{\text{surf},0} \propto h^{-10}$ (see Equation (14)). Thus, a factor of two variation in $h$ corresponds to a factor of $10^{10}$ variation of $\beta_{\text{eff}}$, as we see in Figures 3(g), (h). In the diffusive, optically thick limit, the temperature dependence is weaker with $\beta_{\text{surf}} \sim \beta_{\text{surf},0} r^{-2} \propto h^{-2}$ (see Equation (35)).

Varying the disk mass (normalization of $\Sigma$) also has an effect on the profile of $\beta_{\text{eff}}$. Varying the disk mass shifts the $\tau = m_*$ radius relative to the fiducial model. Also, there is an effect on $\beta_{\text{eff}} \propto \beta_{\text{surf}}$ through the $\Sigma$ dependence of $\tau$ in $\beta_{\text{surf}} \propto f(\tau)$. In the optically thick region (inside of the $\tau = m_*$ radius), $\beta_{\text{eff}} \propto \tau^{-2} \propto \Sigma^2$, while in the optically thin region it is independent of $\Sigma$ (as $f(\tau) \rightarrow 1$). Hence, in the low-mass model (Figure 3(d)), the $\tau = m_*$ radius is shifted inward relative to the fiducial model, and the cooling timescale is reduced inside of this radius. Conversely, in the high-mass model (Figure 3(e)), the $\tau = m_*$ radius is shifted outward, and the cooling timescale is increased inside of this radius. This has the effect of increasing $\beta_{\text{surf}}$ in most of the disk ($r \lesssim 100$ au).

Varying the opacity affects $\beta_{\text{eff}}$ similarly to varying the disk mass. Note that $\beta_{\text{eff}} \propto \beta_{\text{surf}}$ scales as $\Sigma_d^{1/4}$ in the optically thin region, whereas it is proportional to $\Sigma_{d,0}$ in the optically thick limit. In the low-opacity model (Figure 3(b)), the $\tau = m_*$ radius is shifted inward relative to the fiducial model (since $\tau \propto \Sigma_{d,0}$). The cooling timescale is correspondingly reduced inside of this radius and increased outside of it. On the other hand, in the high-opacity model (Figure 3(c)), the $\tau = m_*$ radius is shifted outward, to about 100 au. This has the effect of making most of the disk optically thick, hence increasing $\beta_{\text{eff}}$ in this region.

Finally, varying the profile of either the surface density or the temperature (Figures 3(f), (i)) has a fairly minor effect on...
Dimensionless cooling time $\beta$

$\sim 10^{-3} - 10^{-2}$ $\sim 10$ $\beta$

Regime of density wave propagation

| Locally isothermal | Radiatively damped | Adiabatic |
|--------------------|-------------------|------------|

Linear AMF behavior

$F_1 \sim T(r)$

- $F_1$ strongly decays near planetary orbit
- $F_1 = \text{const}$

Morphology of disk substructures induced by low mass planets

| Gap at the planetary orbit, multiple gaps/rings in the inner disk | Gap at the planetary orbit, multiple gaps/rings in the inner disk |

Figure 4. Schematic classification of different cooling regimes, parameterized by the dimensionless cooling timescale $\beta$, and their impact on the propagation of planet-induced density waves and the production of different types of annular structures in protoplanetary disks. See Section 3.4.3 for details.

$\beta_{\text{eff}}$. Only the slope of $\beta_{\text{eff}}(r)$ in the optically thin part of the disk changes as $T(r)$ varies, in agreement with Equation (14).

3.4.3. Implications for Disk Evolution

In order to better understand the implications of the $\beta_{\text{eff}}$ profiles in Figure 3 for the disk evolution results presented further in Section 5, we briefly review the impact of cooling on the dynamics of planet-driven density waves and the associated global evolution of the disk, as described in Miranda & Rafikov (2020). Note that these results are based on a constant-$\beta$ framework.

The evolution of density waves falls into one of three regimes based on the value of $\beta$, which can be understood through the behavior of the AMF of linear (low-amplitude) waves. These regimes are schematically summarized in Figure 4. For slow cooling, with $\beta \gtrsim 10$, linear waves propagate adiabatically, conserving their AMF. For fast cooling, with $\beta \lesssim 0.01$,8 waves propagate in the locally isothermal regime, in which the AMF scales with the local disk temperature (Miranda & Rafikov 2019b, 2020). Since typically the temperature decreases with $r$, this means that inward-propagating waves become stronger as they travel inward in this regime, and weaker as they travel outward. Finally, for intermediate values of $\beta \sim (10^{-2} - 10^{-1})$, density waves experience strong linear damping as a result of cooling, losing AMF as they propagate. In this regime perturbations with azimuthal wavenumber $m$ experience the strongest damping when $\beta \sim 1/m$.

Gaps and rings emerge in protoplanetary disks as a result of the dissipation of planet-driven density waves, which transfer their angular momentum to the disk material. In both the adiabatic and locally isothermal regimes, the waves dissipate by developing into shocks. This leads to a series of narrow gaps in the disk interior to the planet, each corresponding to the shock locations of different spiral arms into which the density wave splits (Bae et al. 2017; Miranda & Rafikov 2019a). The structures produced in the adiabatic and locally isothermal regimes are similar in shape, but deeper gaps are formed in the locally isothermal case (Miranda & Rafikov 2019b). For intermediate cooling timescales, strong radiative damping transfers the wave AMF to the disk material close to the planetary orbit. As a result, only a single wide gap forms around the orbit of the planet, rather than the series of narrow gaps emerging in adiabatic or locally isothermal disks (Miranda & Rafikov 2020; Ziampras et al. 2020b; Zhang & Zhu 2020).

Combining this understanding with the profiles of $\beta_{\text{eff}}$ in Figure 3, we would expect density waves to experience strong local radiative damping in our fiducial disk model for planets with small orbital radii ($r_p = 10$ and 20 au), for which $\beta_{\text{eff}} \sim 10^{-1} - 1$ for $r < r_p$. On the other hand, for larger $r_p = 50$ and 100 au, $\beta_{\text{eff}} \lesssim 10^{-2}$ is small enough that waves may propagate far from the planet in the nearly locally isothermal regime, experiencing minimal cooling-related damping.

This qualitative conclusion can also be drawn for most of the other disk models shown in Figure 3, with one exception. In the cold disk model (Figure 3(e)), $\beta_{\text{eff}} \sim (0.1-10)$ everywhere for all the different planetary orbital radii considered. Therefore, radiative damping is almost always dominant for this disk model, which should lead to a single gap near the planet and not much structure in the inner disk.

We note that these expectations would be very different if in-plane cooling were neglected. Then cooling would operate on a longer timescale, falling in the range $\beta_{\text{surf}} \sim (0.1-10)$ (dashed black curve in each panel of Figure 3) for most of the disk in all the models considered. In this intermediate cooling regime, radiative damping would play a crucial role in wave dissipation and disk evolution almost universally.

We emphasize that these interpretations depend on the detailed, nontrivial profile of $\beta_{\text{eff}}$, which itself serves only to give an approximate description of the role of cooling in the density waves. It is critical to consider the fact that each Fourier harmonic $\delta \sigma_m$ is thermally relaxed on its own individual timescale $\beta_m$, which is done next.

4. Numerical Setup and Methods

Now we describe the numerical setup and methods used in our calculations of the disk evolution due to planet-driven waves with radiative cooling. We compute a variety of evolutionary and observational diagnostics for each of the disk–planet models listed in Table 1. The disk temperature and surface density profiles are given by Equations (1) and (2), which represent the initial conditions in our numerical simulations. The dust opacity due to small dust grains is described by the opacity law (3). Both cooling from the disk surface and in-plane cooling are considered in all models except for the surface cooling only model, for which in-plane cooling is turned off (see the last column of Table 1). This allows us to assess the importance of properly including in-plane cooling for understanding density wave dynamics.
4.1. Linear Calculations

As part of our calculations, we compute the linear response of the disk to an orbiting planet. This is used for several different purposes, including calculating the effective cooling timescale (for which we compute the linear $\delta r_m$, see Section 3.4), validating the treatment of cooling in our numerical simulations (see Appendix C.1), and understanding the role of cooling in density wave dissipation (decay of the AMF; see Section 5.2). We forgo a detailed description of these linear solutions here and refer the reader to Miranda & Rafikov (2020) for a description of the equations that are solved and to Miranda & Rafikov (2019a) for the numerical solution method.

The key modification to the approach described in these previous works is the specific form of the cooling timescale $\beta_m(r)$, which is given by Equation (33) for each Fourier harmonic of the perturbation. The formalism of Miranda & Rafikov (2020) is easily generalized from the constant-$\beta$ formulation used in that work, as the perturbation equations for each harmonic derived in that study permit an arbitrary radial profile for the cooling timescale.

4.2. Hydrodynamical Simulations

We also carry out a suite of numerical simulations of planet–disk interaction using FARGO3D (Benítez-Llambay & Masset 2016). We use a logarithmic (in $r$) grid with a radial extent of $r_{in} = 0.08r_p$ to $r_{out} = 4r_p$. Wave damping (de Val-Borro et al. 2006) is applied for $r < 0.1r_p$ and $r > 3.6r_p$. When presenting the results of our simulations, we exclude the inner damping zone, so that $0.1r_p$ can be regarded as the effective location of the inner boundary.

The fiducial grid resolution is $N_r \times N_\phi = 1276 \times 2048$, resulting in $32(h_p/0.1)$ grid cells per scale height at $r_p$. In simulations with a thinner or thicker disk, i.e., a smaller or larger value of $h_{50, in}$ (see Table 1), we use a finer or coarser grid, so that the resolution in terms of grid cells per scale height is maintained for a given $r_p$.

A softening length of $0.6H_p$ is applied to the gravitational potential of the planet. Its mass is gradually increased from zero to $M_p$ over the first 10 orbits to reduce transient effects. In each simulation we evolve the disk for 500 orbits.

The key ingredient of our simulations is their inclusion of the radiative effects on density wave propagation. Since the code employed in this study does not include an explicit treatment of radiation transfer, we resort to an effective approach based on the theory developed in Section 3. The details of our implementation of cooling effects are described in Appendix C. There we also test our cooling module (see Appendix C.1) by comparing the results of simulations with cooling for low-mass planets ($M_p = 0.01M_\text{Jup}$) with semi-analytical linear calculations (see Section 4.1), finding good agreement. This validates our implementation of cooling and allows us to extend its use for simulating disks with more massive planets.

For each set of disk/planet parameters, we run two simulations, one with an ideal equation of state (EoS) with $\gamma = 7/5$ and cooling (see Appendix C) and another with a locally isothermal EoS for comparison. We compare the results of the simulations with different thermodynamics, assessing the validity of the locally isothermal approximation for each set of parameters (see Section 6). Note that there are two parameters associated with cooling, $\Sigma_{50, \text{in}}$ and $\delta d_{0,0}$, that are not involved in the locally isothermal simulations. Cooling simulations that differ only by the values of one or both of these parameters can therefore be compared to the same locally isothermal simulation.

4.3. Dust Dynamics and Submillimeter Emission

To facilitate comparison with observations, we produce maps of submillimeter dust emission for each simulation using the procedure described in Appendix D. To make the maps we follow the radial motion of large dust grains in the disk as it is perturbed by the dissipation of planet-driven density waves. This is done by post-processing our hydrodynamical simulations, using the approximate 1D method described in Appendix D.1.

Our post-processing approach allows us to explore different dust particle sizes with minimal computational cost. The results of a single hydro simulation can be used to quickly generate dust distributions for a variety of particle sizes. Since we perform a systematic exploration of the parameters that affect disk cooling, with a large number (~100) of simulations, the additional computational cost associated with a more self-consistent 2D gas + dust treatment would be prohibitive.

Having determined the radial distribution of the dust, we then compute the maps of submillimeter thermal emission as described in Appendix D.3. In most of the maps presented in this paper, we focus on a single particle size (1 mm). In Section 5.4, we provide a limited exploration of the effect of varying the particle size. We find that a single dust size is sufficient for characterizing the critical cooling-related effects described in this work.

5. Results

We now present our simulation results, focusing first on the gas and dust evolution (Section 5.1) and AMF behavior (Section 5.2) in the fiducial disk model and then moving on to other disk models (Section 5.3). Our main focus will be on structures that form in the disk interior to the planetary orbit, which past simulations have suggested to be their dominant location.

5.1. Disk Structure: Fiducial Disk Model

The disk structure produced by the planet in the fiducial disk model is illustrated in Figures 5 and 6. The solid black lines in Figure 5 show the profiles of the azimuthally averaged gas surface density perturbation, $\delta \Sigma = \Sigma(r) - \Sigma_0(r)$, relative to the initial surface density profile $\Sigma_0(r)$, in runs with cooling at 500 orbits, for planets with different masses and orbital radii. The corresponding dust continuum emission maps are shown in the right halves of the panels in Figure 6. Note that the results for the corresponding locally isothermal simulations are shown as red dashed lines in Figure 5 and in the left halves of the panels in Figure 6; see Section 6 for a detailed discussion of these results.

The gas surface density profiles in Figure 5 exhibit a varying number of gaps (surface density depletions or minima), as well as rings (surface density enhancements or maxima) between the gaps, depending on $M_p$ and $r_p$. The exact number of gaps is somewhat ill-defined, since adjacent minima of $\Sigma$ can to some extent merge into a single gap. In some cases, low-amplitude gap-like structures may be embedded in a larger gap structure. These caveats aside, the number of gaps ranges from one to as many as three or four. Alternatively, we can describe the disks...
as having less structure when there are fewer gaps, and more structure when there are more gaps, rather than explicitly quantifying the number of gaps.

As a general trend, smaller planet masses or smaller orbital radii result in fewer gaps or less structured disks, while larger planet masses or larger orbital radii result in a larger number of gaps or a more structured disk. The multiplicity of gaps is affected by which dissipation mechanism—cooling or shocks—dominates the evolution of the density waves launched by the planet. Typically, cooling-dominated dissipation produces fewer gaps, and shock-dominated dissipation produces more gaps. For low-mass planets in the fiducial disk model, cooling is dominant for planets with smaller orbital radii, and nonlinear dissipation is dominant for larger orbital radii (see Section 5.2).

For a 0.1\(M_{\text{th}}\) planet, perturbations to the gas surface density at 500 orbits (Figures 5(a)–(d)) are very weak, at the level of a few percent. Correspondingly, the resulting perturbations to the brightness temperature of the dust emission (Figures 6(a)–(d)) are also weak. They are just barely perceptible relative to the global radial variation of \(T_b\). For the more massive planets (\(M_p = 0.3M_{\text{th}}\) and \(1M_{\text{th}}\)), the perturbations are significant enough to produce order-unity variations in \(T_b\), modifying the global morphology of the emission.

The dust emission maps largely reflect the same trends illustrated by the gas surface density profiles, with more pronounced secondary gaps (gaps that are not co-orbital with the planet) for larger planet masses and orbital radii. The most notable cases are the ones with \(r_p = 10\) au and \(M_p = 0.1M_{\text{th}}\) and \(0.3M_{\text{th}}\), for which secondary gaps are completely suppressed and there is only a single wide gap. This is because \(\beta_{\text{eff}} \sim 1\) at this \(r_p\) (see Figure 3(a)), ensuring rapid deposition of the wave angular momentum into the disk material close to the planetary orbit.

Figure 5. Azimuthally averaged gas surface density perturbation \(\delta \Sigma = \Sigma - \Sigma_0(r)\) relative to the initial surface density \(\Sigma_0(r)\) for planets with different masses \(M_p\) (in terms of the thermal mass \(M_{\text{th}}\); see Equation (5); different columns) and orbital radii \(r_p\) (different rows) at 500 orbits. The solid black lines correspond to the results of simulations with cooling, and the dashed red lines to the results of locally isothermal simulations. The discrepancy score (see Section 6.2) for the pair of profiles is given in the lower right corner of each panel.
In disks with planets at larger orbital radii, the dust ends up concentrated into thinner rings, with wider, more empty gaps between them. The reason for this is the lower overall gas surface density at larger disk radii, resulting in more mobile dust (large values of $St$) for a given dust particle size. Dust is therefore able to more efficiently drift into pressure maxima (rings) and evacuate from regions between them (gaps).

5.2. AMF

The AMF of the planet-driven density waves $F_J$ provides an important diagnostic for understanding the wave-driven evolution of the disk (Miranda & Rafikov 2019a, 2019b). This information is very useful for interpreting the results presented in Section 5.1. Profiles of the AMF for planets with different
masses and orbital radii for the fiducial disk model are shown in Figure 7. The AMF is expressed in terms of the characteristic scale \( F_{J,0} \) (see Equation (43)) for planets with different masses and orbital radii for the fiducial disk model. The black dashed line shows the AMF for a low-mass planet computed using linear theory. The solid lines show the AMF from numerical simulations with planets of different masses, ranging from a very low-mass planet, \( M_p = 0.01M_\odot \) (red curves), for which the disk response is approximately linear, to a massive planet with \( M_p = 1M_\odot \) (magenta curves), for which it is strongly nonlinear. The dotted curve in each panel shows the linear AMF profile for an adiabatic (i.e., very slowly cooling) disk.

\[ F_{J,0} = \left( \frac{M_p}{M_\odot} \right)^2 \rho_p^{-3} \Sigma_p \rho_p^4 \Omega_p^2, \]  

Equation (43)

associated with the total one-sided torque driving the density waves at LRs.

We first examine the profiles of the linear AMF (black dashed curves in Figure 7), corresponding to the evolution of density waves produced by a very low-mass planet. In the absence of cooling, when the disk is adiabatic, the AMF is constant far from the planet (see the gray dotted lines in Figure 7). When cooling is present, the AMF falls off—or, if cooling is sufficiently rapid, grows (in the inner disk)—with distance from the planet. The difference between the black and gray curves is indicative of the extent to which cooling affects the wave evolution.

For \( r_p = 10 \) and 20 au (Figures 7(a), (b)), the linear AMF falls off sharply with distance from the planet, indicating that there is strong dissipation due to cooling. Indeed, Figure 3(a) shows that in these cases \( \beta_{eff} \sim 0.1-1 \) in the vicinity of the planet. Strong wave damping is typical for a cooling timescale in this range (see Section 3.4.3).

For planets at larger orbital radii, the linear AMF is either nearly constant (\( r_p = 50 \) au; Figure 7(c)) or slightly grows with distance from the planet (\( r_p = 100 \) au; Figure 7(d)) in the inner disk. This behavior is consistent with the fact that \( \beta_{eff} \sim 10^{-2} \) near the planet (see Figure 3(a)), corresponding to the nearly locally isothermal limit. The AMF decreases in the outer disk, as this is the only behavior that can be produced by cooling in the outer disk, no matter how short the cooling timescale is.

The AMF from a numerical simulation with \( M_p = 0.01M_\odot \) follows the linear AMF very well for all of the cases in Figure 7. This validates our numerical implementation of cooling described in Appendix C. We have also verified the convergence of these results with respect to resolution using higher-resolution runs \((N_\phi = 4096)\).

Comparing the AMF profile for more massive planets to the corresponding linear AMF profile allows us to quantify the importance of nonlinear dissipation in the evolution of planet-driven density waves compared to cooling. For \( r_p = 10 \) and 20 au, the numerical AMF is similar to the linear AMF even for \( M_p \geq 0.1M_\odot \). This indicates that linear dissipation due to cooling, rather than nonlinear dissipation, is the dominant driver of AMF evolution. But for \( r_p = 50 \) and 100 au, this is not the case. Instead, the AMF decays substantially faster with distance from the massive planet than the linear AMF, with larger deviations for more massive planets, indicating that nonlinear evolution plays a significant role in AMF evolution.

Therefore, cooling plays a more substantial role in wave propagation when massive planets orbit at small radii than when they orbit at larger \( r_p \).

As a rule of thumb, when radiative damping dominates the AMF evolution, the disk substructure is dominated by a single gap, but when nonlinear dissipation is dominant, multiple gaps tend to form (Miranda & Rafikov 2020). Our results demonstrate that cooling is dominant when (i) the (effective) cooling timescale is neither short enough nor long enough for linear waves to propagate without significant linear damping and (ii) the planet mass is small enough for the waves it produces to be quasi-linear (although for small \( r_p \) cooling dominates even when \( M_p \sim M_\odot \), which can be seen in the insensitivity of \( F_J(r) \) curves to \( M_p \) variation in Figure 7(a)). Planets with smaller masses produce density waves with lower amplitudes, which are efficiently damped by radiation for intermediate values of the cooling timescale \((0.01 \lesssim \beta_{eff} \lesssim 1)\). The cooling timescale is roughly in this range in the inner disk \((r_p \lesssim 50 \) au; see Figure 3(a)). But in the outer disk, \( \beta_{eff} \lesssim 0.01 \), so that density waves are not damped by cooling very efficiently (interior to the orbit of the planet) and propagate nearly in the locally isothermal limit. And for more massive
planets (at any orbital radius), the larger amplitude of the density waves enhances the importance of nonlinear dissipation. The AMF profiles (Figure 7) are therefore consistent with the variation of the gap multiplicity with $M_p$ and $r_p$.

Figure 7 also demonstrates that the “initial amplitude” of the AMF, i.e., its value just outside the wave excitation zone, varies with $r_p$. This variation can be quantified by how much the initial AMF with cooling is different from the corresponding value in an adiabatic disk. For $r_p = 10$ au (Figure 7(a)), the initial AMF is slightly larger in an adiabatic disk than in a disk with cooling. In this case, the density waves have already experienced significant damping before leaving the wave excitation zone. However, for larger orbital radii (Figures 7(b)–(d)), the initial AMF is larger in the presence of cooling than in an adiabatic disk. As described in Miranda & Rafikov (2020), this is the result of the increased torque exerted on the disk when the cooling timescale is short (and the response of the disk is close to locally isothermal). This increased torque is partly responsible for the larger amplitudes of gaps caused by planets at larger radii, as seen in Figure 5.

5.3. Dependence on Disk Parameters

The results for the different disk models listed in Table 1 are shown in Figures 8 and 9 (gas surface density perturbation profiles, analogous to Figure 5) and Figures 10 and 11 (dust emission maps, as in Figure 6). In each panel of these figures, the results (gas surface density or dust emission) are shown for the fiducial disk model, as well as for other models with different parameters (see Table 1). In this subsection we describe the variation of the disk morphology associated with varying each of the disk parameters.

5.3.1. Disk Mass

The effects of varying the disk mass (i.e., the surface density normalization $\Sigma_{50\,\text{au}}$) are illustrated in panels (a) and (g) of Figures 8–11. In the low-mass model, the radial profile of $\Sigma$ is more structured for small planetary orbital radii $r_p = 10$ and 20 au, as a result of the reduction of the cooling timescale (see Figure 3(d)), which brings the wave dynamics closer to the locally isothermal regime in both the inner and outer disks. In the high-mass model, there is also more structure relative to the fiducial model for $r_p = 10$ au. But now this is a result of the cooling timescale being raised into the nearly adiabatic regime in the vicinity of the planet ($\beta_{\text{eff}} \sim 1$–10; see Figure 3(e)), so that damping due to cooling is reduced.

For $r_p = 20$ au and 50 au, the high-mass model produces less structure, i.e., less prominent secondary gaps relative to the fiducial model. For disks with planets at these radii, $\beta_{\text{eff}} \sim 0.01$–0.1, leading to cooling playing a more important role in density wave dissipation. For $r_p = 100$ au, the disk structure is not too different from that in the fiducial model, as in both cases the cooling timescale is short enough to be nearly in the locally isothermal regime.

In addition to the impact on the gas distribution, varying the disk mass makes the dust more or less mobile (by increasing or decreasing $\text{St} \propto \Sigma^{-1}$). This results in more diffuse rings for more massive disks and sharper rings for less massive disks, which is evident in the dust emission maps.

5.3.2. Temperature

The effects of varying the disk temperature (or $h_{50\,\text{au}}$) are shown in panels (b) and (h) of Figures 8–11. In the cold disk model (with $h_{50\,\text{au}} = 0.07$), the planet clears a single gap in gas for $r_p = 10$ and 20 au. The surface density profiles in these cases are qualitatively similar to those in the fiducial model, differing only in the depth and detailed shape of the gap. For $r_p = 50$ and 100 au, the cold model also results in a single strong gap, with only very weak secondary gaps. This is in contrast to the strong multiple-gap structure in the fiducial model. This is a result of $\beta_{\text{eff}} \sim 0.1$–1 throughout the disk for all $r_p$ in the cold model (Figure 3(g)), which leads to density waves being strongly damped by cooling, rather than behaving nearly locally isothermally, and hence being damped by shocks.

In the hot model, the planet produces one gap for $r_p = 10$ au, as in the fiducial and cold models. But for larger orbital radii, there is additional structure within the gap (for $r_p = 20$ au) or stronger secondary features ($r_p = 50$ and 100 au). In these cases, the higher temperature results in a shorter cooling timescale almost everywhere in the disk ($\beta_{\text{eff}} \lesssim 0.01$ for $r_p = 50$ and 100 au; see Figure 3(h)), so that density waves propagate in the locally isothermal regime. Thus, the increase in disk temperature leads to radiative wave damping playing a less important role in disk evolution.

5.3.3. Opacity

Dust opacity is varied in panels (c) and (i) of Figures 8–11. In almost all of the cases with varied opacity, the surface density profile is qualitatively similar to that in the fiducial model. The exception is the low-opacity model with $r_p = 20$ au, in which there is a prominent secondary gap, whereas none is present in the fiducial model. Overall, varying the opacity (by an order of magnitude) has a fairly minor effect on the planet-driven disk structure. This is because the profiles of $\beta_{\text{eff}}$ are not very sensitive to the opacity variation (see Figures 3(b), (c)).

5.3.4. Surface Density and Temperature Profiles

The results for disks with different surface density and temperature profiles are shown in panels (d) and (j) of Figures 8–11. The power-law indices of the surface density and temperature profiles have a fairly minor effect on the structure of the disk. This is especially true for large $r_p = 50$ and 100 au, as the $\delta\Sigma$ profiles for these models do not differ much from those for the fiducial model, except at small radii in the inner disk ($r/r_p \lesssim 0.4$).

For planets with small orbital radii ($r_p = 10$ and 20 au), the $\delta\Sigma$ profile in the $T \propto r^{-1/4}$ model is fairly similar to the one in the fiducial model. However, the case with a shallower surface density profile ($\Sigma \propto r^{-1/2}$) shows additional structure (i.e., secondary gaps) as compared to the fiducial model.

5.3.5. In-plane Cooling

The importance of in-plane cooling is highlighted in panels (f) and (l) of Figures 8–11, where we present the results of hypothetical simulations with in-plane cooling artificially turned off. We find that whether in-plane cooling is included has a significant impact on the gas surface density profile and
dust emission map, in terms of the multiplicity, locations, and amplitudes of rings and gaps, in all cases considered.

For example, for $r_p = 10$ au, in the fiducial model (including in-plane cooling), the disk profile is characterized by a single wide gap. However, without in-plane cooling, there is a more complex structure consisting of multiple gaps. This can be understood by comparing the values of the effective cooling timescale between the two cases (see Figure 3(a)). With in-plane cooling (red curve in Figure 3(a)), $\beta_{\text{eff}} \sim 1$ in the inner disk ($r < r_p$), so that linear damping due to cooling dominates the disk evolution. In the absence of in-plane cooling, $\beta_{\text{eff}}$ is instead equal to $\beta_{\text{surf}}$ (black dashed curve in Figure 3(a)), which is $\sim 10$ in the inner disk. In this case, the planet-driven density wave behaves nearly adiabatically, so its AMF is not damped by cooling. The wave then has a chance to split into multiple spiral arms that eventually damp nonlinearly far from the planet, giving rise to disk substructures. This trend is also present for the case $r_p = 20$ au (Figure 8(l)). Here again the density waves behave almost adiabatically without in-plane cooling, resulting in a more complex gap structure.

However, the situation is different for larger planetary orbital radii, $r_p = 50$ and 100 au. Here the $\delta \Sigma$ profile with only surface

---

**Figure 8.** Profiles of the gas surface density perturbation (as in Figure 5) for different disk models and a $0.3M_0$ planet with an orbital radius $r_p = 10$ au (panels (a)–(f)) and 20 au (panels (g)–(l)).
cooling exhibits less structure than that in the fiducial model, which includes in-plane cooling. In these cases, without in-plane cooling we have $\beta_{\text{eff}} = \beta_{\text{surf}} \sim 0.1$ in the inner disk, while with in-plane cooling, $\beta_{\text{eff}} \sim 0.01$ (see the green and magenta curves in Figure 3(a)). This indicates that, without in-plane cooling, radiative wave damping would be important, leading to less structure in the disk. However, the proper inclusion of in-plane cooling pushes the effective cooling timescale down into the nearly locally isothermal regime. As a result, nonlinear dissipation becomes the primary driver of density wave evolution.

5.3.6. Viscosity

All of the simulations discussed so far are inviscid, with a viscosity parameter $\alpha = 0$. To check how a nonzero viscosity can modify our results, we also perform several runs with the fiducial model setup and a finite $\alpha$. The effect of a finite viscosity on the formation of disk substructures is explored in panels (e) and (k) of Figures 8–11, which illustrate the results for the $\alpha = 10^{-4}$ and $\alpha = 10^{-3}$ models.

For the case of a low viscosity ($\alpha = 10^{-4}$), the surface density profiles are not qualitatively different from those in the fiducial (inviscid) case. The number of gaps and their locations...
are not substantially changed. However, the complex structure inside the primary gap that is present in the fiducial model often gets smeared out with the addition of a low viscosity. Overall, the main effect of this low $\alpha = 10^{-4}$ is to somewhat suppress the depths of the gaps and the height of the rings, which is best seen in Figures 9(e), (k).

But for a higher viscosity ($\alpha = 10^{-3}$), the changes are more dramatic: gap depths are further suppressed (very severely for planets at $r_p = 50$ and 100 au), even though the radial locations remain approximately unchanged. Some hints of the multiple-gap structure for planets with larger planetary orbital radii ($r_p = 50$ and 100 au) are still evident in the gas surface density.
profiles, although with a drastically reduced amplitude. However, these perturbations are so weak that they are not discernible in the dust emission maps, which primarily exhibit a single wide gap structure.

In order to interpret these results, we note that the introduction of a finite viscosity has two distinct effects on the evolution of the disk. First, it leads to the viscous dissipation of density waves (Takeuchi et al. 1996). This dissipation, like the dissipation associated with cooling, is a linear process. Hence in a viscous disk, dissipation generally occurs through three distinct channels — radiative, viscous, and nonlinear. By modifying the wave dissipation, viscosity has the potential to alter the associated gap structure. Second, the viscous spreading of the disk tends to fill in gaps and wipe out rings, thus erasing radial structure. In the

Figure 11. Same as Figure 10 but for $r_p = 50$ au (panels (a)–(f)) and 100 au (panels (g)–(l)).
the linear theory for an inviscid disk. The black dashed curve in each panel corresponds to the AMF predicted by classical scenario of gap opening in a viscous disk, this spreading inhibits gap formation for planets below a threshold mass (Rafikov 2002b).

To disentangle these two ways in which viscosity can affect disk substructures, we examine the effect of a nonzero $\alpha$ on the profile of the wave AMF, which allows us to isolate density wave dissipation from viscous spreading. In Figure 12 we show the profiles of the wave AMF (as in Figure 7), taken at 20 orbits, for a $0.3M_{\text{th}}$ planet for disks with different values of the viscosity parameter $\alpha$. In addition to the cases with $\alpha = 10^{-4}$ and $10^{-3}$, for which the gas profiles and dust emission maps are shown in Figures 8–11, two cases with larger values, $\alpha = 10^{-2}$ and $10^{-1}$, are shown. While these large values of $\alpha$ are not expected to be realized in protoplanetary disks, they are illustrative of the effect of viscous damping on planet-driven density waves.

One can see that the AMF profiles for $\alpha = 10^{-4}$ and $10^{-3}$ show essentially no difference relative to the profiles for $\alpha = 0$. Therefore, we see that viscous damping is completely unimportant for the evolution of the wave AMF—relative to the damping associated with cooling and nonlinear evolution—for $\alpha \lesssim 10^{-3}$. It is only for much larger values of $\alpha (\gtrsim 10^{-2})$ that the AMF profiles begin to differ substantially from the inviscid profiles. We note that the relative importance of viscous damping will vary with the planet mass (which should be explored in future work), as a result of the varying importance of nonlinear damping (see Figure 7), with which it competes. However, at least for the case of $M_p = 0.3M_{\text{th}}$ considered here, we can conclude that the differences in the gas and dust profiles seen in Figures 8–11 are entirely due to the second effect described above—the filling in of gaps due to viscous spreading—rather than to viscous wave damping.

5.4. Dust Particle Size

The dust emission maps presented in Sections 5.1 and 5.3 are computed using a single dust particle size of $s_d = 1$ mm. To examine the effect of varying $s_d$, in Figure 13 we present the emission maps for three different particle sizes (0.1 mm and 1 cm, in addition to the fiducial 1 mm) for several disk model/planetary orbital radius combinations. Note that in Figure 13, the dynamic range of $T_d$, spanning three orders of magnitude, is significantly larger than that in the previous emission maps (Figures 6, 10–11). This is necessary in order to highlight the details of the maps for $s_d = 0.1$ mm and 1 cm, which are significantly fainter than those for $s_d = 1$ mm. This is the result of the lower opacity (at a wavelength of 1.25 mm) for 0.1 mm and 1 cm particles (about 1/4 and 1/13 of the opacity for 1 mm particles, respectively). Figure 13 demonstrates that the ring/gap structure for a particular disk model and planetary orbital radius is more or less consistent across different particle sizes. The number of gaps and rings, as well as their approximate locations, does not vary much with $s_d$. Only the widths of the features vary, with rings becoming narrower and gaps becoming wider (and the contrast of ring/gap pairs increasing correspondingly) for larger particles. Additionally, for larger particles there is a smaller amount of dust that is co-orbital with the planet, as a result of the dust being more mobile and hence less susceptible to trapping in the weak pressure bump near the planet. Aside from these details, there is a great deal of resemblance between the maps for different particle sizes (for the same disk model/planetary orbital radius, i.e., within one panel of Figure 13). This reflects the fact that the distributions of dust with different particle sizes are determined by the same underlying gas distribution in each case.

We also see that qualitative differences between maps for different disk models and planetary orbital radii can be found for all three particle sizes shown. For example, the suppression of the complex multiple-gap structure seen in Figures 13(a) and (b), in favor of the single gap in Figure 13(c), is evident for all of the particle sizes. The emission maps for 1 mm particles presented in previous subsections are therefore not unique in exhibiting varied morphology across different disk models. This variation is not associated with any details of the dust dynamics but with differences in the underlying gas distributions resulting from the effects of cooling.
6. Comparison with Locally Isothermal Simulations

The majority of existing numerical studies addressing the problem of planet–disk interaction utilize the locally isothermal approximation for treating gas thermodynamics. In this section we address the validity of this approximation in the context of our results, fully accounting for the different forms of radiative transport affecting density wave propagation in disks.

6.1. Fiducial Disk Model

We first explore the differences between simulations with cooling and analogous locally isothermal simulations, using the fiducial disk model. In Figure 5 we plot the gas surface density profiles, using solid black lines to show the results for simulations with cooling and dashed red lines to show the results of the corresponding locally isothermal simulations.

In all cases considered in Figure 5, the gas profiles in the locally isothermal simulations exhibit three to five pairs of rings and gaps. In contrast, the gas profiles in the simulations with cooling exhibit a much more diverse range of morphologies. Most notably, for \( r_p = 10 \) and 20 au, the multiple-gap structure seen in the locally isothermal simulations is strongly suppressed in the simulations with cooling. There is a single gap at the location of the planet, while secondary and higher-order gaps are either absent or very weak.

Two general trends can be observed in Figure 5. First, discrepancies between the profiles with cooling and the locally isothermal profiles become less pronounced with an increasing planetary orbital radius \( r_p \). This is because the cooling timescale—as represented by, e.g., \( \beta_{\text{eff}} \) (see Figure 3(a))—is typically shorter for larger disk radii. In particular, for \( r_p = 50 \) and 100 au, \( \beta_{\text{eff}} \sim 0.01 \) at moderate distances from the planet. As a result, the behavior of density waves in these regions is nearly locally isothermal. For planets in the inner part of the disk, \( r_p = 10 \) and 20 au, the longer cooling timescale leads to a strong linear damping of density waves, resulting in a preference for single gaps.

The second trend illustrated in Figure 5 is a rough convergence of the profiles in the cooling and locally isothermal simulations with increasing planet mass \( M_p \). This arises because, for more massive planets, nonlinear dissipation due to shocks (depositing angular momentum locally, close to the planetary orbit) plays an increasingly dominant role in the density wave evolution and the associated disk evolution (see Figure 7). This is in contrast to the potentially important role played by linear dissipation associated with cooling for less massive planets. For a 1\( M_\oplus \) planet, the surface density profiles for the cooling and locally isothermal calculations are very similar to each other for \( r \gtrsim 0.5 r_p \). However, discrepancies between the two still emerge at smaller radii. This is because the high-amplitude waves launched by the planet are initially damped predominantly by nonlinear dissipation, but at some large distance from the planet, their amplitude becomes small enough for cooling to take over and become the dominant driver of the wave evolution.

Dust emission maps computed from simulations with cooling and from locally isothermal simulations are shown in the right and left halves, respectively, of each panel in Figure 6. Differences between them become less apparent for larger planet masses or larger orbital radii. There are significant differences between the two maps for the two smallest orbital radii, \( r_p = 10 \) and 20 au, for planets with masses of \( 0.1 M_\oplus \) and \( 0.3 M_\oplus \) (although they may be hard to see because of the intensity scale of these maps). They are less pronounced at these radii for the case with a 1\( M_\oplus \) planet. For the two largest orbital radii, \( r_p = 50 \) and 100 au, differences between the two emission maps are fairly minor for all three planet masses shown.

The differences between the simulations with cooling and the locally isothermal simulations can be understood in terms of the AMF behavior of the planet-driven waves (see Section 7). The locally isothermal approximation represents the limit in which \( \beta_{\text{eff}} \to 0 \). In locally isothermal disks, the AMF of linear waves is proportional to the disk temperature far from the planet (Miranda & Rafikov 2019b). In the fiducial disk model, this means that \( F_\nu \propto r^{-1/2} \) in a locally isothermal disk. Therefore, the AMF grows as waves propagate toward smaller radii (see Figure 1 of Miranda & Rafikov 2019b), and their effect on the inner disk when they dissipate—the formation of gaps—is enhanced in locally isothermal disks.

This is to be contrasted with the behavior of the linear AMF in disks with cooling (dashed curves in Figure 7), which typically decreases with distance from the planet (Figure 7(d)) is an exception, with the AMF exhibiting a very slight rise toward
the inner disk since cooling is very fast in this case and the thermodynamics approaches the isothermal limit). This implies that when cooling is not rapid enough for the locally isothermal approximation to be applicable, the waves damp predominantly in the vicinity of the planetary orbit and therefore fail to open secondary gaps in the inner disk far from the planet.

6.2. Discrepancy Score and Parameter Exploration

It is useful to understand how the degree of discrepancy between simulations with cooling and locally isothermal simulations depends on the parameters of the disk and planet. However, since our grid of simulations consists of 56 simulations with cooling and 36 locally isothermal simulations, a detailed comparison for all cases (as in Figures 5–6) is impractical and may be difficult to interpret. Instead we choose to quantify the difference in terms of a convenient summary statistic.

We define a “discrepancy score” (DS) that quantifies the difference between $\delta \Sigma_C(r)$, the (azimuthally averaged) gas surface density perturbation resulting from a simulation with cooling, and $\delta \Sigma_{LI}(r)$, the gas surface density perturbation in the corresponding locally isothermal simulation. It is computed according to

$$DS = \frac{\int_{r_1}^{r_2} |\delta \Sigma_C(r) - \delta \Sigma_{LI}(r)| \Sigma_0(r)^{-1}dr}{\int_{r_1}^{r_2} (|\delta \Sigma_C| + |\delta \Sigma_{LI}|) \Sigma_0(r)^{-1}dr}. \tag{44}$$

Here $r_1 = 0.1r_p$ (just outside the inner wave-damping zone) and $r_2 = 2r_p$ (beyond which the disk is typically unperturbed) are the radial boundaries of the comparison region, and $\Sigma_0(r)$ is the initial unperturbed surface density. Notice that DS compares profiles of the fractional surface density perturbation $\delta \Sigma / \Sigma$, as shown for different cases in Figures 5, 8, and 9. Also note that $0 \leq DS \leq 1$. Small values of DS indicate that the two profiles (cooling and locally isothermal) are very similar, while large values indicate that they differ substantially.

Examples of how well DS describes the difference between the simulations with cooling and locally isothermal simulations are shown in Figure 5. In each panel the gas surface density perturbation is shown for both simulations, with the corresponding value of DS given in the lower right corner. Note that the case with $M_p = 0.1M_{th}$ and $r_p = 10$ au has the largest value, $DS = 0.66$. Correspondingly, the two surface density perturbation profiles bear little resemblance to each other. For $M_p = 0.3M_{th}$ with $r_p = 50$ and 100 au, $DS < 0.1$, which is indicative of the close resemblance between the two profiles. Other cases between these two extremes have intermediate values of DS. We see that DS seems to provide a reasonable measure of the discrepancy between the two profiles (although, obviously, a single numeric metric cannot fully capture the richness of the $\Sigma(r)$ behaviors) and that $DS \approx 0.1–0.2$ is the rough threshold separating very similar pairs of profiles from very discrepant pairs.

The validity of the locally isothermal approximation using DS as a metric is explored as a function of the various disk parameters in Figure 14. Generally, for smaller planetary orbital radii ($r_p = 10$ and 20 au) DS is large ($\gtrsim 0.2$), while for larger orbital radii ($r_p = 50$ and 100 au), it is comparatively small ($\lesssim 0.2$). In other words, the locally isothermal limit tends to provide a reasonable approximation of the more realistic thermodynamics for larger orbital radii but a poor approximation for smaller orbital radii.

Moreover, DS tends to exhibit more variation with the different disk parameters for smaller planetary orbital radii and less variation for larger orbital radii. That is, the validity of the locally isothermal assumption for small $r_p$ is more sensitive to the details of the disk model than that for large $r_p$. For example, for $r_p = 10$ and 20 au, DS decreases significantly as $M_p$ is increased (Figure 14(a)), and increases as either $\Sigma_{50}$ (Figure 14(b)), $\Sigma_{d,0}$ (Figure 14(d)), or $p$ (Figure 14(e)) is increased. By comparison,
for \( r_p = 50 \) and 100 au, DS is fairly insensitive to these parameters. However, there are exceptions to these trends. For example, for the coldest disk with \( h_{50 \text{ au}} = 0.07 \) DS is large (\( \approx 0.25–0.4 \)) for all orbital radii considered (Figure 14(c)). Also, DS is quite insensitive to the temperature power-law index \( q \) (see Figure 14(f)) in all cases.

The variation of DS with the viscosity parameter is explored in Figure 14(g). For small orbital radii with large DSs for an inviscid disk (\( \alpha = 0 \)), the score becomes much smaller as viscosity increases. This is a result of viscous spreading filling in the gaps and hence erasing the differences in gap structure resulting from the different thermodynamics (see Section 5.3.6).

The role of in-plane cooling is illustrated in Figure 14(h). Except for the case with \( r_p = 10 \) au, the inclusion of in-plane cooling significantly reduces DS, in comparison to the case in which only surface cooling is considered. For the cases with \( r_p = 50 \) and 100 au, DS is fairly large (\( \approx 0.2–0.3 \)) without in-plane cooling but drops to \( \lesssim 0.1 \) when it is included. In other words, for these large orbital radii, the locally isothermal assumption provides a poor approximation for the real thermodynamics when only surface cooling is considered, but is a better approximation when in-plane cooling is included. This is a result of the substantial reduction in the effective cooling timescale that is brought on by the inclusion of in-plane cooling.

7. Discussion

7.1. The Importance of In-plane Cooling

One of the key findings of this work is that the inclusion of in-plane cooling has a significant impact on the details of density wave propagation and gap opening. In-plane cooling has a dominant effect on wave damping simply because the radial lengthscale of the wave as it propagates rapidly becomes considerably shorter than the vertical disk thickness \( H \)—the lengthscale on which surface cooling operates (see Figure 1). As a result, thermal relaxation of the wave perturbation occurs predominantly along the disk midplane, rather than normally to it. The importance of in-plane cooling was first noted in Goodman & Rafikov (2001), and our detailed calculations provide the first quantitative confirmation of their conclusion.

We find rather generally that for planets at small orbital radii \( (r_p \lesssim 20 \text{ au}) \), inclusion of in-plane cooling leads to a suppression of the multiple-gap structure that is present when only radiative cooling from the disk surface is considered (see Figures 10(f), (l)). In contrast, for planets with larger orbital radii \( (r_p \gtrsim 50 \text{ au}) \), the inclusion of in-plane cooling has essentially the opposite effect on gap opening. Without in-plane cooling, the planet often tends to open a single gap (due to strong linear radiative damping), but with in-plane cooling, it opens multiple gaps. In the latter case, the resulting disk morphology is fairly similar to the one produced in a locally isothermal simulation.

As described in Section 5.3.5, this variation is naturally explained by the effective cooling timescale accounting for in-plane cooling being \( \approx 5–25 \) times shorter than the cooling timescale due to surface cooling alone. This difference is often substantial enough to entirely change the regime in which density waves thermally relax—from flux-preserving, almost adiabatic to strongly damped, or from strongly damped to almost locally isothermal—with obvious implications for the emergence of substructures in the disk.

Given the disk–planet model, it may not be easy to a priori assess the radiative effects on wave dynamics and disk evolution. While the surface cooling rate can be directly computed as a function of distance as described in Section 3.1, the calculation of in-plane cooling is a far more involved procedure that requires understanding the characteristics of the density wave in the Fourier domain (see Section 3.2–3.4). At the same time, a useful approximation for the in-plane cooling timescale behavior can still be provided by \( \beta_{m,\ast} \), defined in Section 3.4.1 by assuming that certain modes (namely \( m = m_{\ast} \)) dominate the wave dynamics. The results of Section 3.4 and Figure 3 show that in many cases \( \beta_{m,\ast}(r) \) reproduces the behavior of the more accurate effective cooling timescale \( \beta_{\text{eff}}(r) \) quite well. For this reason we recommend the use of \( \beta_{m,\ast} \) given by Equations (33) and (34) with \( m = m_{\ast} \) for making simple exploratory analytical estimates of the radiative damping timescale for planet-driven density waves.

More generally, our findings emphasize the need for explicitly including radiative effects in numerical studies of planet–disk interaction, something that has been previously brought up in Miranda & Rafikov (2019b, 2020). Whereas cooling may be fast enough for waves launched by planets in the outer disk to propagate almost in the locally isothermal regime, simulations relying on this thermodynamic simplification may still not be accurate enough to reproduce all details of the observed disk substructures at the quantitative level, as we show in Section 6. Explicit treatment of radiative effects in numerical studies of planet–disk interaction should become a standard approach when interpreting protoplanetary disk observations.

7.2. Observational Implications

Using simulated dust emission maps, we show that the multiplicity, widths, separations, and relative contrasts of planet-driven gaps/rings can exhibit significant variation depending on the details of the underlying disk model (see Figures 10 and 11). The gap structure is sensitive to disk properties such as the temperature, mass, and opacity of the disk, which are often poorly constrained in real observations. Due to the variety of possible morphologies, determining whether annular substructures in disks are signatures of planet–disk interaction should become a standard approach when interpreting protoplanetary disk observations.

Our simulated emission maps do not account for the finite spatial resolution in real observations. We may therefore ask whether the variety of morphologies seen in our emission maps could be probed in ALMA observations, which achieve a spatial resolution of, e.g., about 5 au for disks at a characteristic distance of 100–150 pc (Andrews et al. 2018). We find that the
thermodynamic details of the disk model have the most substantial impact on the structure of planet-driven gaps for planets with small orbital radii, \( \lesssim 20 \) au (see Figure 10). However, at such separations, a spatial resolution of \( \approx 5 \) au constitutes a substantial fraction of the disk radius. On the other hand, for larger separations, \( r_p \gtrsim 50 \) au, where such a resolution allows for finer characterization of the gap structure, there is somewhat less variation in the structures seen in dust emission with respect to the physical disk parameters (see Figure 11).

Despite these limitations, variation of the gap structure with the disk parameters is often significant enough—producing qualitatively different structures—that it should be discernible in real observations. For example, varying the disk temperature has a substantial effect for all the planetary orbital radii we consider. In particular, in our cold disk model, secondary gaps are strongly suppressed relative to the fiducial or hot model. Significant and observationally recognizable differences also arise when the disk mass is varied.

### 7.3. Comparison with Previous Work

In Section 6 we show that the locally isothermal approximation used in many numerical studies of planet–disk interaction (Bae et al. 2017; Dong et al. 2017, 2018; Zhang et al. 2018) can often lead to a biased representation of the planet-driven evolution of the protoplanetary disk. At the same time, a number of authors have studied this problem using hydrodynamical simulations that do include some treatment of cooling/radiative transfer. These treatments vary in their levels of approximation and complexity, and it is instructive to put them in the context of our present results.

In the simplest approximation, one assumes that thermal relaxation occurs with either a (spatially) constant cooling timescale or a constant value of \( \beta \), i.e., a cooling timescale that is a fixed fraction of the orbital timescale (e.g., Miranda & Rafikov 2020). In particular, Zhu et al. (2015) used this approximation to explore the structure of spiral shocks using both 2D and 3D simulations. They noted the modification of the pitch angle of spirals, as a result of the \( \beta \)-dependent sound speed (Miranda & Rafikov 2020); however, the damping of density waves by cooling or the modification of gap opening was not explored. More recently, Zhang & Zhu (2020) employed constant-\( \beta \) cooling to study the structure of planet-driven spirals and gaps. In agreement with Miranda & Rafikov (2020), they found that for intermediate cooling timescales, density waves are strongly damped and that the multiple-gap structures produced by a planet are highly suppressed in favor of a single gap.

At the next level of sophistication, one considers a value for the cooling timescale that is derived from physical processes. Here the cooling timescale will typically vary spatially and have explicit dependence on disk properties such as surface density. This is the approach that we take in this work. Our prescription for the cooling timescale associated with radiative losses from the disk surface (Section 3.1) represents a linearized version of the cooling law of Menou & Goodman (2004), which is often employed in 2D simulations.

However, our simultaneous inclusion of the physically motivated prescription for in-plane cooling is an entirely new development. Our use of the WKB approximation for planet-driven density waves to characterize the in-plane radiative transfer (Equation (26)) is in some ways similar to that of Lin & Youdin (2015), who considered in-plane cooling in their study of vertical shear instability (without including surface cooling). However, the application of this machinery to the problem of planet–disk interaction is unique to our work.

Finally, the most self-consistent treatments of radiative effects in simulations directly compute thermal energy transfer along the midplane of the disk (and normal to it in 3D studies), typically resorting to the FLD approximation (Levermore & Pomraning 1981; Kley 1989). A number of studies have made use of FLD in 2D (e.g., D’Angelo et al. 2003; Kley & Crida 2008; Pierens 2015), as well as in 3D (e.g., Kley et al. 2009; Lega et al. 2014; Tsukamoto et al. 2015), to explore the role of radiative effects on disk structure, planetary torques, and migration. However, they did not examine the role of radiative effects in the propagation and dissipation of density waves.

Recently, Ziampras et al. (2020a) presented 2D numerical simulations of planet–disk interaction, some of which included radiation transport using FLD. They were interested in the effects of planetary density waves on the global temperature profile of the disk and found radiative diffusion to be unimportant in terms of its impact on the global disk structure (presumably because in most of their models planet-induced heating was rather moderate). Subsequently, Ziampras et al. (2020b) investigated the structure of gaps opened by planets, but not including radiative diffusion, i.e., in-plane cooling. Their conclusions are fully consistent with those of Miranda & Rafikov (2020) and our results with only surface cooling. However, since in-plane cooling was not included, its effect on density wave dynamics and the formation of the wave-driven structure was missed.

We also note that fully capturing the effect on in-plane cooling in simulations with computationally expensive direct radiation transfer may be rather challenging. The need to resolve radiation transport across the small radial scale of a tightly wrapped density wave, possibly as short as \( 0.1H \) (see Figure 1), requires very high resolution, which may be difficult to achieve even in 2D.

### 7.4. Approximations Used in This Work

Our study makes a number of simplifying assumptions, which we review next. Our treatment of in-plane cooling represents the main improvement made upon existing work. On length scales longer than the radiative length scale \( l_{\text{rad}} \), radiation is transported along the disk midplane in a diffusive manner. In our semi-analytical calculations (Section 3), rather than explicitly propagating the associated diffusion problem through our mathematical framework (which would result in a fourth-order differential equation for the enthalpy perturbation; see Section 3.2.1), we make use of the local approximation. Perturbations are decomposed into Fourier harmonics, and it is assumed that the lengthscale on which the internal energy of each harmonic varies is small compared to the disk radius. As a result of making this approximation, the diffusion timescale— for each Fourier harmonic—ends up depending on the lengthscale over which each harmonic varies. This lengthscale is then approximated as \( k_m^{-1} \), where \( k_m \) is the radial wavenumber of the mode, which we estimate using the standard WKB dispersion relation for planet-driven waves. This treatment of in-plane cooling may not be accurate close to the planet, where, e.g., \( \beta_{\text{eff}} \) exhibits a cusp, but in-plane cooling is not critical in this region anyway.
The implementation of this prescription for in-plane cooling in numerical simulations makes use of the quasi-linearity of planet-driven density waves, since in our code cooling is applied to each Fourier harmonic of the perturbation separately. For massive planets with \( M_p \gtrsim M_{\text{bh}} \), the WKBJ approximation, and hence our implementation of cooling in simulations, may begin to break down. However, for massive planets (\( M_p \gtrsim M_{\text{bh}} \)), nonlinear effects also play a more important role in the evolution of density waves than cooling does, rendering irrelevant any inaccuracy of the cooling prescription.

Since our hydrodynamical simulations do not explicitly treat the dust component, we simulate dust emission maps using a 1D treatment in post-processing, neglecting non-axisymmetric effects. We use an artificial smoothing applied to the dust surface density distribution to approximate possible non-axisymmetric effects (see Appendix D.2). Our calculations use dust particles with a single size (approximately representing a maximum dust size), rather than a distribution of sizes as expected in real protoplanetary disks. The resulting dust distributions are used to create raw (i.e., not convolved with an interferometric beam) maps of the submillimeter emission. Therefore, the resulting emission maps primarily serve to explore the observational impact of cooling and varying the disk parameters.

Finally, our simulations are 2D. In 3D, the treatment of cooling becomes considerably more complicated. For example, it is necessary to properly account for the vertical structure of the disk. Additionally, density wave propagation and dissipation may be modified in 3D (Lubow & Ogilvie 1998; Ogilvie & Lubow 1999; Lee & Gu 2015).

8. Summary

We have carried out a suite of hydrodynamical simulations of gap opening by planets, including a detailed treatment of the radiative physics. In addition to cooling from the disk surface, we have also considered the transport of radiation induced by density waves along the midplane of the disk. We have examined the global structure of gaps and rings produced by planets, using the gas surface density profiles and simulated dust continuum emission maps. We summarize our key results below.

1. In-plane cooling, rather than cooling from the disk surface, plays a dominant role in the evolution of planet-driven density waves. Neglecting in-plane cooling leads to a substantial overestimation of the cooling timescale, typically by more than an order of magnitude.

2. We propose and validate a simple estimate for the (dimensionless) cooling timescale—\( \beta_{\text{n*}} \) defined in Section 3.4.1—that can be used to assess the effect of radiation transport along the disk midplane on density wave propagation.

3. For the effective viscosity expected in protoplanetary disks (\( \alpha \lesssim 10^{-3} \)), viscous damping plays an insignificant role in dissipating planet-driven density waves. It can become competitive with radiative damping or nonlinear wave dissipation only if \( \alpha \gtrsim 10^{-2} \).

4. The number of gaps/rings produced by a planet, as well as their widths, separations, and relative contrasts, depends on the disk temperature, surface density, and opacity, all of which govern the efficiency of cooling for planet-driven density waves.

5. The locally isothermal assumption represents a poor approximation for describing gap opening by low-mass planets at orbital distances of \( \gtrsim 50 \) au. For planets at larger separations (\( \gtrsim 50 \) au) or more massive planets (\( \gtrsim M_{\text{bh}} \)), its performance may improve, but this depends sensitively on the details of the adopted disk model.

Our results on the role of radiative effects in wave dynamics and disk evolution can be naturally applied to other astrophysical systems featuring density waves: disks in stellar binaries (Kley et al. 2008; Ju et al. 2016), circumbinary disks (MacFadyen & Milosavljević 2008; Miranda et al. 2017), boundary layers of accretion disks (Belyaev et al. 2012, 2013), and so on.

Financial support for this work was provided by NASA via grants 14-ATP14-0059 and 15-XRP15-2-0139. We thank the referee, Ruobing Dong, for a careful reading of the paper and for helpful comments and suggestions, which have helped us to improve the paper.

Appendix A

Derivation of the Energy Equation

The evolution equation for the internal energy density \( e = \epsilon_f T \) per unit mass) in 3D is

\[
\frac{d}{dt}(\rho e) + \rho \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla \cdot (F + F_0),
\]

where \( \rho \) is the gas density (so that \( \rho e \) is the internal energy per unit volume), \( \mathbf{u} \) is the 3D gas velocity, \( F \) is the radiative flux associated with the gas, and \( F_0 \) is the radiative flux of the central star. The right-hand side represents the energy source term.

In an axisymmetric \((\partial/\partial \phi \to 0)\) and steady \((\partial/\partial t \to 0)\) state, when the left-hand side of Equation (A1) vanishes, thermal equilibrium requires \( F_0 = -F_0 \), where \( F_0 \) is the disk radiative flux in equilibrium. However, the passage of a density wave through the disk disturbs this equilibrium, inducing a perturbation \( \delta F = F - F_0 \) to the radiative flux; this turns the right-hand side of Equation (A1) into \(-\nabla \cdot \delta F\).

Integrating Equation (A1) vertically and making the conventional conversion from 3D to vertically integrated, 2D variables, one obtains

\[
\frac{d}{dt}(\Sigma e_2) + \gamma \Sigma e_2 \nabla \cdot \mathbf{u}_2 = -\int_{-\infty}^{\infty} (\nabla \cdot \delta F) dz,
\]

where \( e_2 \) and \( \mathbf{u}_2 \) are the appropriately vertically averaged values of internal energy and 2D velocity in the disk plane. Next we write \( \delta F = \delta F_\parallel + \delta F_\perp \hat{z} \), where \( \delta F_\parallel \) is the component of \( \delta F \) in the disk plane, which allows us to integrate the \( \partial \delta F_\parallel / \partial t \) term in the right-hand side of (A2). Also, using the 2D continuity equation to eliminate \( \nabla \cdot \mathbf{u}_2 \) and introducing the 2D pressure \( P = (\gamma - 1) \Sigma e_2 \), we arrive at

\[
\Sigma \left[ \frac{d e_2}{dt} + \gamma \frac{d}{dt} \left( \frac{1}{2} \Sigma \right) \right] = -\nabla_2 \cdot \delta F_2 - 2\delta F_\perp(\infty),
\]

where \( \delta F_2 = \int \delta F_\parallel dz \) is the vertically integrated perturbation of the radiative flux along the disk plane and \( \delta F_\perp(\infty) \) is the deviation of the radiative flux normal to the disk from its unperturbed value \( F_\perp(\infty) = \sigma T_\text{eff}^4 \), both measured high above
the disk plane (above its photosphere). The first term on the right-hand side represents energy transport along the disk plane, while the second term accounts for energy losses from the disk surface.

To relate $\delta F_2(\infty)$ to various parameters of the disk and their perturbations is, in general, a nontrivial exercise requiring a full consideration of the time-dependent transport of radiation across the vertical extent of the disk. To avoid these complications, we employ a simple but intuitive approximation, in which we relate $\delta F_2(\infty)$ to $F_2(\infty)$ and the energy perturbation $\delta e_2 = e_2 - e_{2,0}$ (relative to its unperturbed value $e_{2,0}$) according to

$$
\delta F_2(\infty) = \frac{\delta e_2}{e_{2,0}} F_2(\infty) = \frac{\sigma T_4^4}{e_{2,0}} \delta e_2.
$$

(A4)

The explicit form of $\delta F_2$ is specified in Appendix B.

Plugging the expression (A4) into Equation (A3), dividing it by $\Sigma$, and dropping the subscript “2” from all variables, we finally arrive at the 2D form of the energy equation (6) for a perturbed disk used in this work.

Note that our transition from the 3D energy equation to its 2D version is accurate up to insignificant factors of order unity (which depend on the detailed vertical structure of the disk) and is roughly as rigorous as the definition of the conventional thin-disk variables such as the 2D pressure.

Appendix B
Vertically Integrated Flux Perturbation $\delta F$

In the diffusion limit ($l_e \ll l_{rad}$) the radiative flux in 3D is (Rybicki & Lightman 1986)

$$
F = \frac{-16 \sigma T^3}{3 \kappa T} \nabla T = -\eta \rho \nabla e,
$$

(B1)

where $\eta$ is defined by Equation (18). Note that $\kappa$ is the Rosseland mean opacity, but for simplicity we use the same opacity in both the diffusive and streaming limits. To get the vertically integrated flux we assume (as we do everywhere else in this paper) that $\eta$ is a constant (i.e., does not depend on $z$) with a value set by $\rho = \rho_{min}$. Integrating (B1) over $z$, we get $F_2 = -\eta \Sigma \overline{\nabla e}$ (again, appropriate averaging of the thermal energy $e$ is assumed here).

Taking a perturbation to this equation and dropping the subscript “2,” we arrive at Equation (17).

In the streaming limit, we recall that, most generally, the radiative flux can be expressed through the specific intensity $I_r$, as (Rybicki & Lightman 1986)

$$
F = \int L_r(n)n \, d\Omega \, dv = 4\pi \bar{f} \int H_r \, dv, \quad (B2)
$$

$$
H_r = \frac{1}{2} \int_{-1}^{1} I_r \mu \, d\mu, \quad (B3)
$$

where $\bar{f} = \nabla T/|\nabla T|$ is the direction of radiation propagation and $\mu$ is the cosine of the angle that the unit vector $n$ makes with $\bar{f}$ ($\Omega$ is the solid angle).

In the streaming limit, we can neglect flux absorption and consider only its variation due to disk emissivity. Then, using the Kirchhoff law, the conventional expression for the first moment of the radiation transport equation becomes (Rybicki & Lightman 1986)

$$
\frac{dH_r}{ds} \approx \kappa \rho B_\nu, \quad (B4)
$$

where $s$ is the distance along $\bar{f}$ and $B_\nu$ is the Planck function.

Taking the divergence of Equation (B2), using Equation (B4), and integrating $B_\nu$ over $\nu$ (we take $\kappa$ to be frequency-independent), we get

$$
\nabla \cdot F \approx 4 \kappa \rho T^4. \quad (B5)
$$

Taking a perturbation to this equation, we integrate it over $z$, finding $\nabla \cdot \delta F_2 = 16 \kappa \Sigma \bar{T}^4 \delta T$. Dropping the subscript “2,” we finally arrive at Equation (23).

Appendix C
Implementation of Cooling in Simulations

In our formulation, the cooling timescale (Equation (33)) has explicit dependence on the azimuthal wavenumber $m$ of the perturbation. For this reason, in our simulations, which do not compute explicit radiation transfer, cooling is implemented using a spectral approach, in which the different Fourier harmonics of the specific internal energy perturbation are evolved independently. The cooling step takes place after the main hydrodynamical update step in FARGO3D. Given the specific internal energy $e(r, \phi, t)$, as a function of the polar coordinates $r$ and $\phi$ and at time $t$, we obtain the updated energy due to cooling at time $t + \delta t$, denoted by $e(r, \phi, t + \delta t)$, using the following procedure.

1. At each radius $r$, the specific internal energy at $t = 0$, denoted by $e_0(r)$, is subtracted from $e(r, \phi, t)$ in order to obtain the specific internal energy perturbation $\delta e(r, \phi, t)$. The perturbation is then decomposed into Fourier harmonics $\delta e_m(r, t)$ by means of a fast Fourier transform (FFT).

2. The evolution of the harmonics is described by Equation (31), with the dimensionless cooling timescale $\beta_m$ given by Equations (33) and (34) (see also Equations (4) and (13)–(15)). To compute $\beta_m$, we require the local temperature and surface density as inputs. For these we use the initial temperature profile—assuming the fixed equilibrium temperature distribution set by stellar irradiation—and the numerical azimuthally averaged surface density, $\langle \Sigma(\rho, \phi, t) \rangle$.

3. Each harmonic of the perturbation, up to the maximum azimuthal number $m_{max}$ (and including $m = 0$), is then updated using the exact solution to Equation (31),

$$
\delta e_m(r, t + \delta t) = \exp \left[ -\frac{\Omega(r) \delta t}{\beta_m} \right] \delta e_m(r, t). \quad (C1)
$$

For $m > m_{max}$, we set $\delta e_m(r, t + \delta t) = 0$.

4. Subsequently, the updated energy perturbation $\delta e(r, \phi, t + \delta t)$ is constructed from the updated Fourier harmonics $\delta e_m(r, t + \delta t)$ using the inverse FFT. Finally, the initial profile $e_0(r)$ is added back to $\delta e(r, \phi, t + \delta t)$ to obtain the final updated energy $e(r, \phi, t + \delta t)$.

As described in the third step above, the $\delta e_m$ harmonics are updated according to Equation (C1) for $m < m_{max}$ but are set to zero for $m > m_{max}$. There are three reasons for this simplification. First, for very large values of $m$ ($m \gg m_{max}$), the amplitude of $\delta e_m$ (or the Fourier harmonic of any other fluid
variable) is very small, and a negligible amount of angular momentum is carried by the harmonic. Second, the cooling timescale for these harmonics is typically very short, especially in the optically thick inner disk, where $\beta_m \propto m^{-2}$. Hence, if one of these harmonics acquires an appreciable amplitude, it should be erased very quickly anyway. Finally, for the maximum $m$, which can be potentially captured by the simulation, $N_c f/2$ (i.e., the Nyquist frequency) is very high (1024 in our simulations using the fiducial resolution), and updating every harmonic is computationally expensive. In our simulations we choose $m_{\text{max}} = 8h_p^{-1}$ and find that using a larger value has a negligible effect on the results.

Using this value of $m_{\text{max}}$, our simulations with cooling incur a moderate slowdown, about 40%–50%, relative to the corresponding locally isothermal simulations. In Appendix C.1, we present a test of our implementation of cooling. We find that the numerical implementation reproduces the results of the linear calculations very closely.

In some of the simulations with cooling, there is a numerical artifact that arises at the boundary of the inner damping zone. It is a result of a mismatch between the short cooling timescale ($\beta \sim 1$) inside the damping zone (associated with the artificially imposed wave damping) and the long damping timescale due to radiative cooling just outside of it. This only occurs when the cooling timescale in the inner disk is long, e.g., $\beta_{\text{eff}} \gtrsim 10$, which is the case only for small planetary orbital radii (this is never the case for $r_p \gtrsim 50$ au). The mismatch leads to an artificial density enhancement just outside the damping zone. When this feature is present, the region inside of $r = (0.12-0.15)r_p$, where the disk evolution is affected by the artifact, is removed when the numerical results are presented.

C.1. Verification of Numerical Cooling Treatment

Now we test our implementation of cooling in numerical simulations by comparing in Figure 15 the profiles of the Fourier harmonics of the AMF for the waves produced by an $M_p = 0.01M_\odot$ planet in a simulation with cooling implemented as described above to the predictions from linear theory. Here we use the fiducial disk model, with the planet located at $r_p = 50$ au. Overall, one can see excellent agreement between the numerical results and linear theory. Most importantly, the damping rates of the different harmonics (i.e., the rates at which the different $F_j^m$ components fall off with distance from the planet) are accurately reproduced in the numerical simulation. We also find similarly good agreement for different planetary orbital radii and disk models. This validates our numerical treatment of cooling.

We also experiment with several alternative implementations of cooling, in which the total internal energy perturbation (rather than its Fourier harmonics) is relaxed with a single cooling timescale, which is intended to produce approximately the correct behavior for all of the harmonics of a perturbation. This would bypass the need to decompose the internal energy perturbation $\delta e$ into Fourier harmonics, which is computationally convenient. Possible choices for this timescale include $\beta_{\text{eff}}$ (Equation (40)), as well as other similarly constructed quantities, e.g., the averages of $\beta_m$ weighted by either $|\delta e_m|$ or $|\delta e_m|^2$. Such “single effective $\beta$” methods generally do not adequately reproduce the linear behavior of the $F_j^m$ components. A common failure of these methods is an overestimation of the damping rate for harmonics with small values of $m$ (e.g., $\lesssim 3$). For these harmonics the cooling timescale $\beta_m$ (Equation (33)), set primarily by in-plane cooling, is usually much longer than that for higher-order harmonics (in the diffusive regime; see Figure 2). Therefore, using an effective cooling timescale that is weighted toward higher azimuthal numbers (which tend to have larger amplitudes) generally produces incorrect behavior for low-$m$ harmonics. In our simulations we therefore use the spectral method described in Appendix C.

Dust Emission Maps

D.1. Dust Dynamics

Dust evolution is treated by post-processing our hydrodynamical simulations using an approximate 1D method similar to the approach used in Miranda & Rafikov (2019b). However, in this work we treat dust as Lagrangian particles rather than as a pressureless fluid. In this approach, it is not necessary to mitigate the large density contrasts that arise at the outer boundary in a fluid-like treatment.

The positions of $2 \times 10^5$ dust particles are initialized at $t = 0$ following the distribution $\Sigma_0(r) = \eta \Sigma(r)$, where $\eta = 0.01$ is the initial dust-to-gas ratio. At each timestep, their positions are updated using the dust radial velocity (Takeuchi & Lin 2002)

$$u_{\text{r},d}(r) = \frac{1}{1 + \text{St}^2} \left( \bar{u}_r + \frac{\text{St}}{\langle \Sigma \rangle_0 \Omega_k} \frac{d \langle P \rangle_0}{dr} \right). \quad \text{(D1)}$$

Here St is the Stokes number of the particle, $\langle \Sigma \rangle_0$ and $\langle P \rangle_0$ are the azimuthally averaged gas surface density and pressure, and $\bar{u}_r = \langle u_r \rangle_0 / \langle \Sigma \rangle_0$ is the effective 1D gas radial velocity. The fluid variables in Equation (D1) are extracted from the hydrodynamical simulation at a frequency of 20$/\Omega_p$ ($\eta_p = 2\pi/\Omega_p$) and are linearly interpolated from the radial hydrodynamical grid to the particle position. Particles are dropped if they leave the grid, and so the total dust mass can decrease with time. However, typically only a few percent of the dust mass is lost.
The Stokes number of a particle is

\[ \text{St} = \frac{\pi \rho_d s_d}{2 \Sigma} = 0.016 \rho_{d,1} s_{d,1} \Sigma_{10}^{-1}, \tag{D2} \]

where \( \rho_{d,1} = \rho_d/(1 \text{ g cm}^{-3}) \) and \( s_{d,1} = s_d/(1 \text{ mm}) \) are the scaled bulk density and particle size. For most of our calculations, we choose a particle size of \( s_d = 1 \text{ mm} \) for the dust particles (the effects of varying the particle size are explored in Section 5.4). Since we consider a fixed particle size and the gas surface density decreases with \( r \), initially \( \text{St} \) increases with \( r \), so that the particles are more mobile, i.e., drift faster, at larger radii. In the fiducial disk model, the initial Stokes number at \( r_p \) denoted by \( \text{St}_0 \), ranges from 0.008 to 0.08 for 1 mm particles.

In some of our simulations (several cases with \( r_p = 100 \text{ au} \) and different disk models), the dust is so mobile (with \( \text{St}_0 \lesssim 0.1 \)) that it all drifts inward and is lost to the central star (i.e., leaves the inner edge of the grid) before appreciable pressure bumps in the gas can be built up to trap it. In these cases we hold the dust particles fixed for the first several hundred orbits of the simulation, allowing pressure bumps to be built up, before they are released to drift freely, ensuring that all the dust is not lost through the inner boundary. In these cases, the final results are not sensitive to the exact time at which the dust particles are released, as long as it is \( \gtrsim 200 \text{ orbits} \).

### D.2. Calibration of 1D Dust Surface Density

In our 1D treatment of dust dynamics, the particles accumulated in a gas pressure maximum tend to collapse into unphysically narrow radial structures. This is the result of neglecting different effects in our 1D treatment: explicit diffusion and the stirring of dust by non-axisymmetric density waves. Both of these effects would serve to smear out the radial dust distribution. Without accounting for them, the dust distribution near a pressure maximum resembles a delta-function. The width of the peak is therefore underestimated, and the concentration is overestimated.

In order to approximate the neglected stirring effects, we artificially smooth the dust density distribution produced in our calculations. The degree of smoothing is calibrated against the dust surface density profiles presented in Zhang & Zhu (2020), which came from hydrodynamical simulations with a self-consistent 2D treatment of dust dynamics.

To perform this calibration we carry out a dedicated set of numerical simulations following the “B17” setup described in Zhang & Zhu (2020). The parameters of these runs are the aspect ratio \( h_p = 0.07 \), temperature power-law index \( q = 0 \) (i.e., the temperature profile is globally isothermal), surface density power-law index \( p = 1 \), and viscosity parameter \( \alpha = 5 \times 10^{-5} \). The smoothing length for the planetary potential is \( 0.1 H_p \) (substantially smaller than the value used in our other runs, \( 0.6 H_p \)). The radial extent of the disk is \( [0.2 r_p, 2 r_p] \), the boundaries of the inner and outer wave-damping zones are \( 0.24 r_p \) and \( 1.6 r_p \), and the size of the grid is \( N_x \times N_y = 752 \times 2048 \) (approximately half that of Zhang & Zhu 2020).

We run these simulations with planet masses \( M_p = 0.1 M_{\oplus}, \ 0.3 M_{\oplus}, \ \text{and} \ 1 M_{\oplus} \), each for 500 orbits, and then perform 1D dust evolution calculations. The final dust surface density \( \Sigma_k(r) \) is convolved with a Gaussian profile in \( \ln(r) \), so that the width of the Gaussian relative to the radius \( r \) is constant. The convolution is parameterized by the FWHM of the Gaussian (FWHM \( \approx 2.36 \sigma \)). We choose FWHM/r = 0.025, 0.05, and 0.075.

The profiles of the gas surface density and convolved dust surface density for the three different dust sizes are shown in Figure 16. Here the dust size is parameterized by \( \text{St}_0 \), the Stokes number at \( r = r_p \) and \( t = 0 \). Figure 16 can be compared with Figure 8 of Zhang & Zhu (2020); specifically, their case with \( Q = 100 \), which is closest to our simulations, in which the disk self-gravity is not included. Our gas surface density profiles reproduce the results of Zhang & Zhu (2020) very closely.

We find that the dust profiles convolved with FWHM/r = 0.025 and FWHM/r = 0.05 produce peaks that are too narrow and whose amplitudes are too large. But convolution with FWHM/r = 0.075 reproduces the correct widths and amplitudes of the peaks quite well. The agreement is not as good for \( M_p = 0.1 M_{\oplus} \), however, only a small subset of the simulations in this paper use such a small planet mass.

Therefore, we find that smoothing our 1D distributions with FWHM/r = 0.075 provides an adequate approximation of 2D effects neglected in our 1D calculations. We use this value of the Gaussian FWHM for producing all of our dust maps.

### D.3. Submillimeter Emission

Continuum emission maps are computed from the dust surface distributions using a procedure that closely follows that of Zhang et al. (2018). However, we simplify the process by considering only a single particle size, rather than a size distribution.

We choose an observing frequency of \( \nu_{\text{obs}} = 240 \text{ GHz} \) or a wavelength of \( \lambda_{\text{obs}} = 1.25 \text{ mm} \), as in the DSHARP survey. We make use of the DSHARP opacity, tabulated as a function of particle size and wavelength, from Birnstiel et al. (2018). For 1 mm particles and a wavelength of 1.25 mm, the absorption opacity is \( \kappa_\nu = 2.2 \text{ cm}^2 \text{ g}^{-1} \) (per unit dust mass). Note that \( \kappa_\nu \) is the opacity that controls the emission at frequency \( \nu \), and should not be confused with \( \kappa_d \), the frequency-averaged opacity of smaller dust grains that mediates disk cooling (see Equation (3)).

The optical depth is \( \tau_\nu(r) = \kappa_\nu \Sigma_d(r)/2 \). For the fiducial disk model, the optical depth at \( t = 0 \) is \( \tau_\nu = 0.11 r_p^{-1} \). Using the optical depth \( \tau_\nu(r) \) of the dust distribution at 500 orbits, we compute the brightness temperature of the dust emission,

\[ T_b(r) = [1 - e^{-\tau_\nu(r)}] T(r). \tag{D3} \]

To bring our emission maps to the same scale, the brightness temperature is normalized by the disk temperature at \( r_p \),

\[ T(r_p) = 43 M_p, h_{50 \text{ au}}, 0.1 r_p^{2.5} \text{ K}, \tag{D4} \]

where \( r_{p,50} = r_p/(50 \text{ au}) \) and \( h_{50 \text{ au}} = h_{50 \text{ au}}/0.1 \).

In the Rayleigh–Jeans limit, the intensity of the emission is proportional to the brightness temperature, \( I_\nu(r) \propto T_b(r) \). The peak blackbody emission wavelength for \( T(r) \) is \( \lambda_{\text{peak}}(r) = 0.067 h_{50 \text{ au}}^{-2} r_p^{-0.5} \text{ mm} \). For the fiducial disk model, we therefore have \( \lambda_{\text{peak}} \ll \lambda_{\text{obs}} \) even at large radii (hundreds of astronomical units), and so the Rayleigh–Jeans limit is always applicable. Only for the cold disk model does the Rayleigh–Jeans limit become only marginally applicable far in the outer disk.
Figure 16. Gas surface density perturbation (relative to the initial profile; top row) and dust surface density perturbation for different-sized dust particles (parameterized by the initial Stokes number at $r_p$) for different planet masses $M_p$ (in terms of the thermal mass). The dust surface density is convolved in $\ln(r)$ with a Gaussian profile with different widths, described by the FWHM (different-colored curves), in order to approximate 2D effects not captured in our 1D calculation of the dust evolution.
Rybicki, G. B., & Lightman, A. P. 1986, Radiative Processes in Astrophysics (Weinheim: Wiley-VCH)
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 500, 33
Takeuchi, T., & Lin, D. N. C. 2002, ApJ, 581, 1344
Takeuchi, T., Miyama, S. M., & Lin, D. N. C. 1996, ApJ, 460, 832
Tsukamoto, Y., Takahashi, S. Z., Machida, M. N., & Inutsuka, S. 2015, MNRAS, 446, 1175

Zhang, S., & Zhu, Z. 2020, MNRAS, 493, 2287
Zhang, S., Zhu, Z., Huang, J., et al. 2018, ApJL, 869, L47
Zhu, Z., Dong, R., Stone, J. M., & Rafikov, R. R. 2015, ApJ, 813, 88
Zhu, Z., Stone, J. M., & Rafikov, R. R. 2013, ApJ, 768, 143
Ziampras, A., Ataiee, S., Kley, W., Dullemond, C. P., & Baruteau, C. 2020a, A&A, 633, A29
Ziampras, A., Kley, W., & Dullemond, C. P. 2020b, A&A, 637, A50