Planet-driven density waves in protoplanetary discs: Numerical verification of non-linear evolution theory

Nicolas P. Cimerman and Roman R. Rafikov

1Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK
2Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA

ABSTRACT
Gravitational coupling between protoplanetary discs and planets embedded in them leads to the emergence of spiral density waves, which evolve into shocks as they propagate through the disc. We explore the performance of a semi-analytical framework for describing the non-linear evolution of the global planet-driven density waves, focusing on the low planet mass regime (below the so-called thermal mass). We show that this framework accurately captures the (quasi-)self-similar evolution of the wave properties expressed in terms of properly rescaled variables, provided that certain theoretical inputs are calibrated using numerical simulations (an approximate, first principles calculation of the wave evolution based on the inviscid Burgers equation is in qualitative agreement with simulations but overpredicts wave damping at the quantitative level). We provide fitting formulae for such inputs, in particular, the strength and global shape of the planet-driven shock accounting for non-linear effects. We use this non-linear framework to theoretically compute vortensity production in the disc by the global spiral shock and numerically verify the accuracy of this calculation. Our results can be used for interpreting observations of spiral features in discs, kinematic signatures of embedded planets in CO line emission (’kinks’), and for understanding the emergence of planet-driven vortices in protoplanetary discs.

Key words: accretion, accretion discs – hydrodynamics – shock waves – methods: analytical – methods: numerical – planets and satellites: formation.

1 INTRODUCTION
Gravitational coupling of young planets with the protoplanetary discs in which they form is known to give rise to global spiral density waves. These planet-induced waves may be responsible for the non-axisymmetric structures observed around multiple disc-hosting stars, for example spiral patterns seen in scattered light observations of MWC 758 (Grady et al. 2013; Benisty et al. 2015), SAO 206462 (Muto et al. 2012; Garufi et al. 2013), and other systems.

Density waves launched by planets carry angular momentum and energy, which can be deposited into the disc fluid giving rise to disc evolution (Goodman & Rafikov 2001; Rafikov 2002a, 2016) and gap formation (Lin & Papaloizou 1993; Rafikov 2002b), provided that the wave can be damped either linearly (Takeuchi, Miyama & Lin 1996; Miranda & Rafikov 2020a, b) or non-linearly (Goodman & Rafikov 2001; Rafikov 2002a). Non-linear wave dissipation naturally results from non-linear steepening of the wave profile and its eventual evolution into a shock, even for low wave amplitudes. Irreversible energy dissipation at the shock is the ultimate cause of the non-linear damping of a density wave (Goodman & Rafikov 2001; Rafikov 2016).

The details of propagation and evolution (damping) of weakly non-linear planet-driven waves have been studied in the local (homogeneous shearing sheet) approximation by Goodman & Rafikov (2001) (hereafter GR01). They showed that for low mass planets, \( M_p \lesssim M_{\text{th}} \), the problem of linear wave excitation by planetary gravity can be naturally separated from the subsequent wave propagation affected by non-linear effects, and explored both stages. Here \( M_{\text{th}} \) is the characteristic mass scale, the so-called thermal mass

\[
M_{\text{th}} = \frac{c_s^3}{\Omega G} = \left( \frac{H_p}{R_p} \right)^3 M_*,
\]

where \( c_s \) is the sound speed, \( \Omega \) is the orbital angular frequency, \( H_p \) is the disc scale height at the planetary distance \( R_p \), and \( M_* \) is the stellar mass such that planets with \( M_p \sim M_{\text{th}} \) launch density waves which are non-linear already at excitation. Subsequently Dong et al. (2011a), Dong, Rafikov & Stone (2011b) carried out high resolution hydrodynamical simulations of planet-launched density waves in the shearing sheet approximation. They verified the main results of the GR01 theory, in particular the prediction for the distance from the planet \( l_{\text{sh}} \) at which the wave shocks (see also Yu et al. 2010), and the evolution of the amplitude and width of the wave profile in the asymptotic ’N-wave’ regime. Additionally, they investigated the production of vortensity \( \Delta \xi \) by the planet-driven spiral shocks, showing it to be a steep function of planet mass (\( \Delta \xi \propto M_p^3 \)).

The local theory of GR01 has been subsequently extended to global discs in Rafikov (2002a) (hereafter RR02), fully accounting for radial gradients of the protoplanetary disc properties (e.g. \( c_s \) and gas surface density \( \Sigma \)) and curvature effects. This more general, global framework has in particular allowed Rafikov (2002b) to explore gap
opening by migrating planets accounting for the non-locality of non-linear wave damping, which was later verified numerically in Li et al. (2009) and Yu et al. (2010).

The goal of this work is to quantitatively verify the global theoretical framework of RR02 using hydrodynamical simulations. We study various aspects of excitation, propagation, and decay of the density waves driven by subthermal mass ($M_p \leq M_\text{th}$) planets. Also, following Dong et al. (2011b), we use the global theoretical framework of RR02 to compute vortensity generation by global planet-driven spiral shocks and verify these predictions numerically.

There are several key motivations for this study.

First, there has been limited amount of past work trying to verify non-linear propagation of global density waves, mainly documenting the evolution of the wave profile towards the asymptotic N-wave regime (Duffell & MacFadyen 2012) and examining the deviation of the spiral shape from the linear prediction (see Section 5.2; Zhu et al. 2015). Here we aim to provide a systematic study of the wave excitation and evolution, covering a broad range of the relevant parameters – disc aspect-ratio $h_p = H_p/R_p$, planet mass $M_p$, and disc surface density profile $\Sigma(R)$ – using a variety of diagnostics.

Secondly, theoretical frameworks of GR01 and RR02 reduce the full set of fluid equations to a single inviscid Burgers equation in the limit of a weakly non-linear density wave. So far the accuracy of this approximation (which has been recently employed in Bolatti et al. 2021) has not been studied, and we provide its systematic test in this work.

Thirdly, the global theory of RR02 explicitly assumes that a linear effect – excitation of a single armed spiral wave (Ogilvie & Lubow 2002) by the planetary potential – takes place only close to the planet, and that far from it only non-linear effects regulate wave evolution. However, recent studies (Bae & Zhu 2018; Miranda & Rafikov 2019a) have shown that linear effects (in the form of evolving interfer-

2.2 Problem setup

In this section, we first describe the physical setup for the problem under consideration (Section 2.1), and then provide relevant theoretical background (Sections 2.2 and 2.3).

2.1 Problem setup

We consider a planet of mass $M_p$ orbiting a central star of mass $M_*$. on a circular orbit with a semimajor axis $R_p$ that lies within a 2D gas disc. The 2D approximation is appropriate for thin discs, such that the disc aspect-ratio $h = H/R = c_s/(\Omega R) < 1$. Planetary gravity perturbs the motion of disc fluid, giving rise to a density wave that we explore in this work. We adopt polar coordinates $(R, \phi)$ to describe this problem.

The background disc state, unperturbed by the planet, has a power-law profile of the surface density

$$\Sigma_0(R) = \Sigma_p \left( \frac{R}{R_p} \right)^{-p},$$

where $p$ is a constant and $\Sigma_p = \Sigma_0(R_p)$. Throughout this work, we assume the mass of the disc to be small, $M_d \ll M_*$, such that its self-gravity can be neglected.

For all our models, we adopt a globally isothermal equation of state (EoS), $P = c_s^2 \Sigma$, where $P$ is the vertically integrated pressure, $\Sigma$ is the surface density, and $c_s$ is the spatially constant sound speed. We opted to use this barotropic EoS instead of an often used non-barotropic locally isothermal EoS, for which the sound speed follows a prescribed radial profile $c_s(R)$, for several reasons.

First, it has recently been shown by Miranda & Rafikov (2019b, 2020a, b), that the use of a locally isothermal EoS leads to a non-conservation of angular momentum flux (AMF) of the density waves freely propagating through the disc even in the absence of explicit dissipation. On the other hand, a barotropic EoS (in particular, the globally isothermal EoS) conserves wave AMF after excitation (see also Lin & Papaloizou 2011; Lin 2015), which greatly simplifies our analysis of the problem (see below). Secondly, the globally isothermal assumption eliminates baroclinic vorticity driving, see Section 8.3. Thirdly, the use of this EoS reduces the number of parameters characterizing the problem at hand.

The unperturbed background disc is in radial centrifugal balance, accounting for the radial pressure gradient: radial velocity $u_{R,0}(R) = 0$, azimuthal velocity $u_{\phi,0}(R) = R\Omega_0(R)$, where

$$\Omega_0^2(R) = \Omega_K^2(R) + \frac{c_s^4}{R^2 \Sigma_0(R)} \frac{d \Sigma_0(R)}{d R},$$

and $\Omega_K = \sqrt{G M_*/R^3}$ is the Keplerian orbital frequency.

Our fiducial disc model has an aspect ratio at the planet’s radius, $h_p = 0.05$ and a surface density slope $p = 3/2$. The latter results in an initial vortensity profile that is almost constant for slightly sub-Keplerian discs, see Section 6.

When presenting our results we adopt units where $G = M_* = \Omega_K(R_p) = \Sigma_0(R_p) = 1$.

2.2 Linear evolution of planet-driven density waves

The linear excitation of spiral density waves by massive perturbers has been extensively studied in the literature (e.g. Goldreich & Tremaine 1980; Ward 1997). Planetary gravity gives rise to the torque density per unit radius

$$\frac{dT(R)}{d R} = -R \int \Sigma(R, \phi) \frac{\partial \Phi_p(R, \phi)}{\partial \phi} d\phi,$$

(here $\Phi_p$ is the planetary potential) which imparts angular momentum flux on the density wave close to the planet, with the main contribution coming from a region $\lesssim 2H_p$ away from the planet (GR01; Dong et al. 2011a). Further away, the waves can be considered as freely propagating (GR01).

In the absence of dissipation and in barotropic discs, freely propagating waves preserve their AMF $F_J$ defined as

$$F_J(R) = R^2 \int \Sigma(R, \phi) u_{\phi}(R, \phi) \delta u_{\phi}(R, \phi) d\phi,$$

(with $\delta u_{\phi}(R, \phi) = u_{\phi}(R, \phi) - u_{\phi,0}(R)$), i.e. $\partial_R F_J = 0$. For planet-driven waves, the characteristic amplitude of the AMF (one-sided
Evolution of planet-driven density waves

Lindblad torque is (Goldreich & Tremaine 1980)

\[ F_{J,0} = \left( \frac{M_p}{M} \right)^2 \frac{M_p}{h_p^2} \sum_{p} \Omega_p \Omega_p^2 \sum_{p} \Omega_p \Omega_p \]

and we will use it as a reference value.

To zeroth order, the freely propagating planet-driven density wave appears as a narrow, single-armed spiral wrapped by differential rotation in the inner and outer disc, see Fig. 1(a) for illustration. A simple linear prediction for the azimuthal location of this spiral (RR02; Ogilvie & Lubow 2002), based on phase coherence arguments for linear WKBJ modes, reads

\[ \phi_{in}(R) = \phi_p + \text{sign} (R - R_p) \int_{R_p}^{R} \frac{\Omega(R') - \Omega_p}{c_0(R')} dR'. \]  

This relation gives an approximate shape of the curves traced out by the peak of the density perturbation in \( R - \phi \) coordinate.

However, it has been realized recently that a single-peak structure does not fully capture the shape of the spiral density wave in the inner disc (Dong et al. 2015; Fung & Dong 2015). Instead, far enough from the planet, the wave is comprised of several narrowly spaced (for low \( M_p \)) spiral arms, as indicated in Fig. 1(a). This redistribution of AMF was understood in Bae & Zhu (2018) and Miranda & Rafikov (2019a) as a manifestation of linear evolution of the density waves in differentially rotating discs, following from their weakly dispersive nature. It was shown that the interference of different azimuthal harmonics comprising the perturbation pattern evolves in the inner disc in such a way as to cause the appearance of secondary, tertiary, etc. arms after the wave has travelled far enough from the planet. During this process, the primary spiral arm steadily transfers some of its AMF to the secondary (and higher order) spiral(s), thereby decreasing in amplitude even in the absence of any dissipative effects (the full wave AMF is conserved in the absence of damping in barotropic discs). Arzamasskiy & Rafikov (2018) and Miranda & Rafikov (2019a) have also shown that this phenomenon is not unique to waves driven by a planet, but occurs for any passively propagating density wave. In this work, we will explore how this linear effect impacts non-linear wave evolution.

2.3 Non-linear evolution of global density waves

While the excitation of waves by planets can be described by linear theory, their propagation over large distances is subject to non-linear effects, even for small wave amplitudes. This is the process that ultimately makes waves steepen, form a shock, and dissipate, thus depositing their angular momentum to the disc.

Keeping the most important non-linear terms in their analysis, GR01 have shown that the evolution of weakly non-linear waves can be reduced to a 1D non-linear wave propagation problem described by the inviscid Burgers equation (see below). Their calculation was local as it adopted the shearing-sheet geometry and assumed a uniform background state of the disc. RR02 extended that analysis to the global case by allowing for an inhomogenous, radially structured disc and accounting for the cylindrical geometry, still reducing the problem of wave propagation to the same mathematical form.

The basis of RR02 analysis lies in the special rescaling of variables: radius \( R \) gets replaced with a time-like coordinate \( \tau \) defined as a function of \( R \) as

\[ \tau(R) = \text{sign}(R - R_p) \times \frac{3}{2} \frac{M_p}{M} \frac{R_p}{h_p} \int_{R_p}^{R} \frac{\Omega(R') - \Omega_p}{c_0(R')} dR', \]

where the auxiliary function \( g(R) \) is defined by

\[ g(R) = \frac{2^{1/4}}{R_p c_p \Sigma_p \Omega_p^{1/2}} \left( \frac{R \Sigma_p^2}{\Omega_p - \Omega_p} \right)^{1/2}, \]

while azimuthal angle \( \phi \) gets replaced by a space-like coordinate \( \eta \) defined as

\[ \eta(R, \phi) = \frac{3}{2} \frac{R_p}{c_p} \left[ \phi - \text{sign}(R - R_p) \int_{R_p}^{R} \frac{\Omega(R') - \Omega_p}{c_0(R')} dR' \right]. \]

Note that \( \tau(R_p) = 0 \) and also \( \eta(R, \phi) = 0 \) for \( \phi = \phi_{in}(R) \), i.e. at the location of the wake in linear theory, see equation (7). Finally, the density perturbation \( \Sigma - \Sigma_0 \) gets rescaled to a new independent variable \( \chi \) defined as

\[ \chi(R, \phi) = \chi(\tau, \eta) = \frac{\gamma + 1}{2} \frac{M_p}{M} \frac{\Sigma - \Sigma_0}{\Sigma_0} g(R). \]

In all these definitions \( \Sigma_0 \) and \( c_0 \) can in general be functions of \( R \) (unlike GR01).

With these new variables the free propagation of a weakly non-linear density wave is described by the inviscid Burgers equation (RR02)

\[ \partial_t \chi - \partial_\phi \chi = 0, \]

the same as in the local limit of GR01. If we neglect non-linearity, i.e. drop the second, quadratic in \( \chi \), term in this equation, then \( \chi = \chi(\eta) \) becomes independent of \( \tau \) and \( R \). Then equation (11) allows one to directly determine how the surface density perturbation \( \Sigma - \Sigma_0 = \Sigma_0/\chi(\eta)(\phi_{in}(R)) \) varies due to AMF conservation in the course of linear wave propagation in a differentially rotating, non-uniform disc (RR02).

Making use of the WKBJ approximation, one can obtain the following expression for the wave AMF in terms of \( \chi \) (GR01, RR02):

\[ F_{\chi,0}^{WKBJ}(\tau) = \frac{\sqrt{5} \Sigma_p R_p}{3 \Omega_p} \left( \frac{M_p}{M} \right)^2 \Phi(\tau), \]

Thus, the evolution of the AMF \( F_{\chi,0}^{WKBJ} \) is entirely dictated by the behaviour of the integral

\[ \Phi(\tau) = \int \chi(\eta, \tau) d\eta. \]

In particular, in the absence of non-linearity, when \( \chi \) is independent of \( \tau \), one finds that both \( \Phi \) and \( F_{\chi,0}^{WKBJ} \) are conserved in the course of wave propagation, as expected.

We now summarize some key results of GR01 and RR02 regarding propagation of weakly non-linear planet-driven waves.

(i) In the Burgers equation framework (8)–(12) wave evolution is self-similar. In other words, equation (12) does not contain any physical parameters of the problem \( (M_p, \Sigma_0(R), c_0(R), \Omega_p, \text{etc.}) \), all of which are absorbed into the definitions of variables \( \tau, \eta, \chi \), and initial conditions.

\[ (\text{ii}) \text{We discuss in Section 4 how well this approximation describes the full wave AMF } F_{\chi,0} \text{ given by the equation (5) in our simulations.} \]
which results in a distance from the planet, at which the wave starts to shock (i.e. where \( \eta = 0 \)).

The linear prediction (equation 21) is shown by the black dotted line. (b) Radial profile of the azimuthally averaged vortensity perturbation \( \delta \zeta \). (c) Radial profile of the azimuthally averaged surface density perturbation \( \delta \Sigma / \Sigma_1 \). (d) Map of the vortensity perturbation \( \delta \zeta \). The dashed horizontal lines show the predicted shock locations \( R = R_p \pm l_{sh} \).

This long-term simulation was run at the resolution \( N_R \times N_p = 3448 \times 7200 \).

(ii) For a weakly non-linear wave excited by a planet with \( M_p \ll M_0 \), the regions of linear wave excitation (within \( (1 - 2)H_p \) from the planet) and its subsequent non-linear propagation are spatially separated.

(iii) As a result of non-linear evolution, the wave inevitably shocks at some radial separation \( l_{sh} \) away from the planet. This shocking length is set by the value of \( \tau = \tau_{sh} \) at which the characteristics of the Burgers equation (12) first cross. Using the linear wake profile from the excitation region (determined by solving the linearized perturbation equations) as an initial condition for the Burgers equation, GR01 found

\[
\begin{align*}
\tau_{sh} & \simeq \tau_0 + 0.53, \\
\tau_0 & = 1.89 M_p / M_{th},
\end{align*}
\]

being the point at which the initial conditions are applied. The radial distance from the planet, at which the wave starts to shock (i.e. where \( \tau = \tau_{sh} \)) is

\[
l_{sh} \simeq 0.8 H_p \left( \frac{\gamma + 1}{12/5} \frac{M_p}{M_{th}} \right)^{-2/5}.
\]

One can see that for \( M_p \ll M_0 \), the shocking length \( l_{sh} \) indeed lies outside the linear wave excitation region, \( l_{sh} \gg H_p \).

(iv) After shocking, the wave profile asymptotically (for \( \tau \gg \tau_{sh} \)) evolves into an N-wave shape (Landau & Lifshitz 1959; Whitham 2011), which is a typical outcome of the wave evolution governed by the Burgers equation. In this regime, the azimuthal width of the wake grows as \( \Delta \eta \propto \tau^{1/2} \) while the wave AMF decays as \( \Phi \propto \tau^{-1/2} \), which results in \( F_j^\text{WKB} \propto \tau^{-1/2} \) (GR01).

Assuming a Keplerian rotation profile and taking the globally isothermal limit \( (c_0 = c_s = \text{const.}, \gamma \to 1) \), the variables \( \tau, \eta, \chi \) take the following form:

\[
\begin{align*}
\tau & = \text{sign}(R - R_p) \times \frac{3}{25/2} \frac{M_p}{M_{th}} R_p^{-5/2} \\
& \times \left[ \int_{|x|}^{R/R_p} \left( |x^{3/2} - 1|^{1/2} x^{p/2 - 11/4} \right) \, ds \right], \quad (18) \\
\eta & = \frac{3}{2 h_p} \left[ \phi - \text{sign}(R - R_p) \frac{1}{h_p} \left( 3 - 2 \sqrt{R_p / R - R_p} \right) \right], \quad (19) \\
\chi & = \frac{M_{th} \Sigma}{M_p} \left( \frac{\Sigma_0}{\Sigma_0} \right)^{1/2} \left( \frac{\sqrt{2 h_p}}{R/R_p} \left( R/R_p \right)^{-(p+1)/2} \right)^{1/2}, \quad (20)
\end{align*}
\]

\[
\phi_{\text{lin}} = \phi_p + \text{sign}(R - R_p) \frac{1}{h_p} \left( 3 - 2 \sqrt{R_p / R - R_p} \right). \quad (21)
\]

The behaviour of \( \tau \) for the two values of surface density slopes \( p \) that we consider in this work is shown in Fig. 2. In a uniform disc \( (p = 0) \) higher (lower) values of \( \tau \) are reached in the inner (outer) disc, as compared to the fiducial value \( p = 1.5 \), meaning that non-linearity is accelerated (slowed down) in a uniform disc. Also, as can be seen from equation (18), \( \tau \propto h_p^{3/2} \), so that increasing the disc scale height would lead to slower non-linear wave evolution over the same radial distance.\footnote{We use the assumption \( \Omega(R) = \Omega_{K}(R) \) in order to keep the analytical form of the coordinates simpler. We checked that accounting for sub-Keplerian rotation results in negligible changes for thin discs.}
Runge–Kutta scheme. Our results. The equations are integrated in time using a second-order Roe’s approximate Riemann solver. We have performed additional solved. Our fiducial setup computes the fluxes at cell-interfaces where \( \Phi_1 \).

4 ATHENA+ ++ is publicly available on GitHub.

5 As noted by Miranda & Rafikov (2019a), the indirect term is linear in \( R \) and thus becoming more important for \( R \gg R_0 \). It might be more significant when considering even larger domains of the disc.

\[ \Phi_1 = \Phi_1^{(e)} = -GM_p \frac{d^2 + (3/2)r_i^2}{(d + r_i^2)^{3/2}}. \]

where \( d = |r - r_p| \) is the distance from the planet and \( r_p = \epsilon HR = R_0 \) is a smoothing length. This form of the potential has been used by Dong et al. (2011a, b), who have shown it to give better agreement with linear shearing-sheet calculations than not employing any smoothing, than the typical Plummer-type potential used in literature, see Appendix A for more details. In cases where a different potential is used, we state it explicitly. We adopt \( \epsilon = 0.6 \) in this work, following Müller, Kley & Meru (2012).

In order to avoid spurious shocks, we introduce the planetary gravitational potential over a time-scale \( t_{\text{ramp}} \) that is typically set to 10 planet orbits, except for the highest resolution simulation in which we reduced it to 5 planet orbits (since it cannot be evolved for a very long time). Explicitly, the planet mass is introduced as

\[ M_p(t) = M_p \left\{ \begin{array}{ll} (1/2) \left[ 1 - \cos \left( \frac{\pi t}{t_{\text{ramp}}} \right) \right] & \text{if } t < t_{\text{ramp}}, \\ 1 & \text{else.} \end{array} \right. \]

We also employ an orbital advection scheme that is based on the FARGO algorithm (Masset 2000). Its implementation in Athena is described by Sorathia et al. (2012). It has now also been implemented in ATHENA++ and is available in the latest public release (version 21.0). In our models orbital advection is performed at intermediate integration time-steps to keep the scheme second-order accurate in time. In our version, orbital advection is done using the Keplerian background velocity \( u_\text{k}(R) = \sqrt{GM/R} \), which is only a function of \( R \) and constant in time. This allows for a larger time-step, leading to a net speed-up of more than one order of magnitude for our typical setup. We find that using this method leads to increased numerical diffusion as compared to models that do not employ this scheme. We have tested the method carefully, including tests with discs that do not host planets. As an additional check, we also performed several simulations using FARGO3D (Benítez-Llambay & Masset 2016) and found good agreement between the results obtained with two different codes, see Appendix B for details.

We vary the parameters of the problem in the following way: the planet mass is \( M_p/M_\odot \in [0.01, 0.1, 0.25, 0.5, 1] \), the disc aspect ratio at the planet location is \( h_\text{p} \in [0.05, 0.07, 0.1] \), and the background disc surface density slope is \( p \in [0, 1.5] \).

The simulation domain extends over a radial range \( 0.2 \leq R/R_\odot \leq 4.0 \) with logarithmic spacing and uniformly extends over the full azimuthal angle \( 0 \leq \phi \leq 2\pi \). We employ wave-damping zones (de Val-Borro et al. 2006) close to the radial boundaries in the zones \( 0.2 \leq R/R_\odot \leq 0.28 \) and \( 3.4 \leq R/R_\odot \leq 4 \) to avoid reflections. Only the radial velocity and density are damped towards their initial values, while the azimuthal velocity remains unchanged.

Probing the evolution of very small perturbations in the inviscid regime requires very high resolution to suppress numerical viscosity, which can lead to linear damping of the wave (e.g. Dong et al. 2011a, b). Regarding this issue, we perform an extensive resolution study to judge the convergence of our results. Resolutions ranged from the lowest \( N_R \times N_\phi = 3448 \times 7200 \) up to \( N_R \times N_\phi = 27584 \times 57600 \). This corresponds to 58 and 464 cells per disc scale-height at the planet radius.

We evolved the lowest resolution cases for about 1000 planetary orbits, while the highest resolution simulations were evolved for 20 orbits, due to their high computational cost. However, this is still long enough for the wave to reach a quasi-steady state, as the longest sound crossing time across the domain (for \( h_\text{p} = 0.05 \)) is less than 20 planet orbits.
During the simulations we keep track of the fluxes computed by the Riemann solver, flux divergences and source terms in the conservation equations, as well as azimuthal averages of all conserved quantities and their time-derivatives on a time resolution given by the actual time-stepping of the code. We calculate the vorticity as a line-integral over cell edges using reconstructed values of velocities:

$$\omega_c = \frac{1}{\Delta x^2} \int_c (V \times u) \, dA = \int_{\partial c} u \cdot ds$$  \hspace{1cm} (26)

and divide it by the cell-centred surface density to compute the vortensity in a cell. This method avoids calculating the velocity gradients in post-processing which can lead to diverging results close to discontinuities (shocks).

### 3.2 Linear calculations

To provide a benchmark for Athena++ to meet in the low planet mass limit ($M_p \ll M_{\text{th}}$), we use the numerical method of Miranda & Rafikov (2019a) to calculate the global structure of linear perturbations in globally isothermal discs. This calculation, relying on solving the linearized fluid equations in fully global setup, allows us to confirm the precision of the implemented methods in fully non-linear simulations and to demonstrate the transition into weakly non-linear behaviour. For consistency, we employ the same potential $\Phi_p^{\text{lin}}$ in these calculations (see also Appendix A).

### 3.3 Solutions of Burgers equation

Our verification of the weakly non-linear theory described in Section 2.3 relies on solving the inviscid Burgers equation (12) in order to directly obtain wave profiles and to compare their evolution with our simulations results. We numerically solve equation (12) adopting a finite-volume scheme with Engquist-Osher flux-splitting and a second-order Runge–Kutta integrator for stepping in $\tau$. As initial wave profiles, we use data from our non-linear simulations, close to the planet at $\tau = \tau_0$, where excitation of the wake is expected to be almost complete. This corresponds to $R/R_p \approx 0.1$ and $R/R_p \approx 1.068$ (in the inner and outer disc) for $h_0 = 0.05$.

### 4 EVOLUTION OF THE PLANET-DRIVEN WAVES AS A FUNCTION OF $M_p$

In this section, we compare different descriptions of the planet-driven wave density evolution for two different planet masses assuming the fiducial disc with a density slope $p = 3/2$ and aspect ratio at planet location $h_0 = 0.05$ (dependence on disc parameters is explored in Section 7). Unless specifically mentioned, e.g. when discussing long-term behaviour, results are shown after 20 planetary orbits, when the initial distribution of $\Sigma$ in the disc is not yet strongly perturbed.

#### 4.1 The low-mass case, $M_p = 0.01M_{\text{th}}$

In order to verify our implementation of the planetary potential and as an additional check of the orbital advection algorithm in Athena++, we compare the results of a simulation featuring a low-mass planet $M_p = 0.01M_{\text{th}}$ to the solutions of the linear problem obtained from the method described in Miranda & Rafikov (2019a), see Section 3.2. Since $\Delta \phi \approx 5.4\sigma_{\text{ph}}$ for this $M_p$, the regions of wave excitation (within $(1 - 2)\Delta \phi$ from the planet) and weakly non-linear propagation should be spatially well-separated. This particular run features the highest resolution we explored ($N_R \times N_{\text{ph}} = 27\,584 \times 57\,600$), which is needed to correctly capture the weak non-linearity of the low-amplitude wave. Lower resolutions show signs of purely numerical wave decay before the shock is formed (see also Section 6).

We start by displaying in Fig. 3 azimuthal (or $\eta$) slices of the dimensionless wave perturbation $\chi$ defined by equation (11), at different radii $R$. We compare results of linear calculations (top row), solutions of the Burgers equation (middle row), and the fully non-linear simulation results (bottom row). The left-hand (right-hand) panel show the inner (outer) wake. As has been pointed out by Bae & Zhu (2018) and Miranda & Rafikov (2019a), a secondary arm naturally appears in the linear calculation at $R \approx 0.5R_p$, without the need for non-linearity, see panel (a). There we also see that the peak of the primary arm decreases in amplitude (and slightly shifts horizontally leftwards), as a result of its AMF being transferred to the secondary arm. In the outer disc, the linear calculation (panel b) predicts a single peak that stays around $\eta = 0$ with fixed amplitude and no additional arms are formed. Some transfer of AMF occurs only between the leading and trailing troughs surrounding the peak. Due to the linear nature of this calculation, wave breaking does not occur.

Moving to solutions of the Burgers equation (panels c & d), we see that the inner and outer disc evolve very similarly, as expected from the form of the equation. Since this equation does not account for the linear wave evolution, its solutions do not capture the formation of the secondary arm in the inner disc and transfer of flux between the troughs in the outer disc. The non-linear nature of this approximation leads to steady wave steepening that eventually leads to the formation of a shock (a sharp jump in the wave profile). After the shock has formed, the resulting dissipation gradually reduces the wave amplitude while the wave profile broadens.

Comparing the top and bottom rows, we find excellent agreement in both the shape and amplitude of the excited wake close to the planet, confirming that our implementation in Athena++ reproduces the linear prediction for $M_p \ll M_{\text{th}}$. But as the distance from the planet increases, the wave profile in a simulation (panels e & f) gets additionally distorted by non-linear effects, leading to a horizontal shift of its peak to more positive (negative) values of $\eta$ as compared to the linear prediction in the outer (inner) disc. After a shock is formed, the wave starts to decay due to dissipation. Note that far from the planet ($R \gtrsim 2.4$) and ($R \lesssim 0.5$), the wakes in panels (e) & (f) do not display such steep jumps as the solutions of Burgers equation, most likely because of wave dispersion (and some numerical diffusion). This comparison clearly shows that the full calculation (bottom row) naturally combines linear and non-linear effects, with both playing important roles.

To further understand the different approximations used in this work, in Fig. 4 we show the rescaled angular momentum flux $F_{\text{J}}/F_{\text{J,0}}$ (see equation 6) computed in different ways, as a function of radius $R$. For the Athena++ simulations and the linear calculation, we compare results for $F_{\text{J}}^{\text{WK}}$ in the WK module (i.e. with $F_{\text{J}}$ obtained from $\Phi$, see equation 13) to $F_{\text{J}}$ obtained directly from the full solutions for density and velocity perturbations, see equation (5). In the case of Burgers equation that has been reduced to the variable $\chi$, we show only the former.\(^7\)

\(^7\)The divergent behaviour of $\Phi$ around $R = R_p$ is due to the fact that the atmosphere of the planet leads to a finite surface density perturbation there, while $g(R)$ in the definition (11) of $\chi$ diverges as $R \to R_p$. 

\(\text{MNRAS 508}, 2329–2349 (2021)\)
Evolution of planet-driven density waves

Figure 3. Evolution of the wave profile for a low mass planet, $M_p = 0.01 M_\oplus$. From top to bottom we show azimuthal slices (at several radii) of the planet-driven density perturbation obtained by (panels a–b) solving the linear perturbation equations, (panels c–d) solving Burgers equation and (panels e–f) full non-linear simulation with ATHENA++. The left-hand and right-hand rows show inner and outer disc, respectively. As a result of non-linear effects (i.e. as compared to the top row showing the linear results), the wave steepens, shocks and subsequently decays in amplitude (while getting stretched azimuthally). In the inner disc, one can also see the emergence and non-linear steepening of the secondary arm.

Figure 4. A detailed comparison of different wave angular momentum metrics for $M_p = 0.01 M_\oplus$. We compare the results for $F_J$ obtained using equation (5) (left-hand column in the legend) as well as calculated using the WKB approximation via equation (13) ($F_{JWKB}$, right-hand column in the legend). AMFs computed in the linear approximation (Section 3.2) are shown with grey dash-dotted and purple tripod curves; $F_{JWKB}$ obtained using the solutions to Burgers equation is shown with red dots; $F_J$ and $F_{JWKB}$ computed using ATHENA++ models are shown via the blue solid and orange points; the black dashed curve shows the integrated torque. All fluxes are rescaled by the characteristic amplitude $F_{J0}$. The dotted grey lines indicate the scaling expected for an N-wave, $F_J \propto \tau^{-1/2}$, with two different normalizations.

wound (and WKB approximation becomes accurate) only after propagating a certain distance away from the planet. Beyond $\simeq 0.3 R_p$ from the planet, both ways of computing AMF agree to within a few per cent. Note the asymmetry in the saturated $F_J$ levels between the inner and outer discs, that is ultimately responsible for planet migration.

The $F_{JWKB}$ computed using the solution of Burgers equation (red dots) shows an initially constant flux (which is below the linear prediction since the initial condition for the Burgers evolution was set close to the planet, where the linear $F_J$ has not yet fully accumulated), until the shock is formed $\simeq 0.3 R_p$ away from the planet, in good agreement with the predicted shocking distance $l_s \simeq 0.27 R_p$. Beyond that point, the Burgers AMF decays and asymptotically follows the N-wave scaling, $F_J \propto \Phi \propto \tau^{-1/2}$ (grey dotted curves with different normalization), although convergence to the asymptotic scaling is slow, as expected from GR01 and the fact that $\tau$ is small for low $M_p$, see equation (18).

Finally, for the $F_J$ derived from our simulations (blue solid curve) we find excellent agreement with the linear $F_J$ close to the planet, for $|R - R_p| \lesssim l_s$ – the location and amplitude of the peak $F_J$ match precisely. This means that $F_J$ is accurately conserved after excitation, indicating negligible numerical dissipation. To provide yet another check, we plot the integrated torque density from the simulation with the black dotted curve, which agrees perfectly with $F_J$ close to the planet (and with the linear $F_J$ far from it).

At a distance that agrees with the theoretical prediction $l_s$ and results from Burgers equation, the wave in our simulation shocks and starts to decrease. Its $F_J$ drops below the linear $F_J$ and the integrated torque density. At the same time, both in the inner and outer disc, $F_J$ decays notably slower than what Burgers equation predicts. The slow decay in the inner disc between $0.4 \lesssim R/R_p \lesssim 0.6$ can be easily...
4.2 An intermediate mass planet, $M_p = 0.25M_{\text{th}}$

We now examine the case of a more massive planet $M_p = 0.25M_{\text{th}}$, keeping the same (fiducial) disc parameters.

In Fig. 5, we again plot azimuthal wake profiles, this time omitting the linear calculation, which is no longer relevant for this $M_p$ because of the increased importance of non-linear effects. As compared to the low-$M_p$ case, we see that the shock forms closer to the planet and the wave evolves faster, e.g. the wake decays and broadens azimuthally more rapidly. This is because equation (18) yields larger values of the time-like coordinate $\tau$ for higher $M_p$, everything else being equal. As expected, solutions of the Burgers equation again differ from the full simulation results by not capturing the formation of the secondary spiral arm in the inner disc, see panels (a) and (c). Also, similar to the low-$M_p$ case shown in Section 4.1, Burgers framework (top row) predicts a stronger wave decay and a slower advance of the shock front compared to the simulation (bottom row) at intermediate distances from the planet ($R/R_p \approx 0.8$ and $1.3 \approx R/R_p$). Nevertheless, the evolution of the wake shape in an $M_p = 0.25M_{\text{th}}$ simulation bears more resemblance to the Burgers equation solutions than in the $M_p = 0.01M_{\text{th}}$ case, because of the increased importance of wave non-linearity.

These conclusions are corroborated by the behaviour of the angular momentum fluxes in $M_p = 0.25M_{\text{th}}$ case, see Fig. 6. In a simulation, shortly after reaching the maximum value that agrees with the

Figure 5. Same as middle and bottom rows of Fig. 3 but for an intermediate-mass planet, $M_p = 0.25M_{\text{th}}$. Note a faster non-linear evolution of the wave profile.

Figure 6. Same as Fig. 4 but for $M_p = 0.25M_{\text{th}}$. Results from the linear calculation are identical to those presented in Section 4.1 and are shown just for reference. The green curve and brown crosses show the results of an $M_p = 0.25M_{\text{th}}$ run with a spatially truncated potential (see equation 42), discussed in detail in Section 8.1. Note that AMFs show no extended plateau after wave excitation, as for this $M_p$ the shock forms at a distance $l_0 \approx 0.075R_p$ from the planet. The higher amplitude wave leads to a stronger shock, resulting in faster wave decay. See the text for details.
integrated torque density, $F_J$ begins to decay as a result of wave shocking at $\rho_b \simeq 0.075R_\oplus$ away from the planet. The $F_J$ damping is faster for this higher $M_\text{p}$, and the AMF decreases by more than a factor of 10 when the wave reaches the damping zones. In the inner disc, this decay is again delayed by the formation of the secondary arm, but overall, both in the inner and outer discs, $F_J$ is not too deviant from the $\tau^{-1/2}$ scaling expected from the Burgers framework, which is reasonably well followed by the Burgers $F_J^{\text{WKB}}$ (red dots). Even though $F_J$ in a simulation still decays with the distance slower than the Burger's $F_J^{\text{WKB}}$, the disagreement between the two is reduced for higher $M_\text{p}$.

5 CHARACTERISTICS OF THE PLANET-DRIVEN SPIRAL SHOCKS

In agreement with the local version of the non-linear evolution framework (18)–(12), Dong et al. (2011b) found that shocks forming in their simulations as a result of non-linear evolution of planet-driven density waves possess certain universal characteristics. We now examine whether the same is true in our global setup.

5.1 Evolution of the shock strength

The main characteristic of a shock wave is its strength $\epsilon \equiv \Delta \Sigma/\Sigma_0$, where $\Delta \Sigma$ is the surface density jump (in 2D) across the shock front. Measuring $\epsilon$ in isothermal simulations can be surprisingly difficult (especially for weak shocks), which has been noted before. This issue does not arise for non-isothermal shocks, for which the entropy jump can be used to effectively estimate the shock strength (e.g. Ziampras et al. 2020). This approximation could be reasonable for a density wave that has already evolved into an N-wave. However, it will fail for the waves that are just starting to develop shock features.

At the next level of sophistication, one can compare the minimum and maximum values of the density perturbation $\delta \Sigma_0/\Sigma_0$ inside some azimuthal interval $\delta \phi$ around the discontinuity. We found that the results of this method depend strongly on the choice of $\delta \phi$, which can lead to inconsistent results. For these reasons, we use an alternative method. We use a known fact that the energy and angular momentum losses of a density wave are connected (Lynden-Bell & Kalnajs 1972, GR01). Using this argument, one can show (Rafikov 2016) that the angular momentum dissipation rate (per unit radial distance) of a spiral shock with $m$-fold azimuthal symmetry and pattern frequency $\Omega_\phi$ is given by

$$\frac{\partial F_J}{\partial R}_{\text{diss}} = \text{sign}(\Omega_\phi - \Omega(R)) m R (\Sigma/\Sigma_0)^2 \Psi(e),$$

(27)

where for an isothermal EoS (Belyaev, Rafikov & Stone 2013; Rafikov 2016)

$$\Psi(e) = \frac{(2 + e) - 2(1 + e) \ln(1 + e)}{2(1 + e)}.$$  

(28)

In the case of a planet-driven shock $\partial R F_J_{\text{diss}}$ is in general not equal to the radial derivative of the AMF, $\partial R F_J$. This is because the latter is also affected by the torque deposition due to the planetary potential, $dT/dR$, so that

$$\frac{\partial F_J}{\partial R} = \frac{\partial F_J}{\partial R}_{\text{diss}} + \frac{dT}{dR}.$$  

(29)

We use this relation in our simulations to find the angular momentum deposition rate due to the shock $\partial R F_J_{\text{diss}}$ as the difference between the radial angular momentum flux divergence and the torque density (equation 4) directly computed from the simulation data (i.e. we use divergences of the Riemann fluxes and the source term monitored in ATHENA + on the run). At radii where there is only one shock (i.e. that of the primary arm), the knowledge of $\partial R F_J_{\text{diss}}$ then allows us to deduce the shock-strength $\epsilon$ by solving equations (27)–(28) for $\epsilon$, assuming $m = 1$ and using a Newton–Raphson root finder. We note that this method requires high resolution simulations to ensure that linear damping due to numerical viscosity does not significantly affect $\partial R F_J$, which would otherwise bias our estimate of $\epsilon$.

We illustrate the application of this procedure to our simulation results in Fig. 7, where we display the shock strength expressed in terms of $\Delta \chi$ – the jump of $\chi$ across the shock, which can be related to $\epsilon$ via equation (11). We show $\Delta \chi$ as a function of the coordinate $\tau$ for several planet masses (increasing from top to bottom) and fiducial disc parameters, for both the inner (blue) and outer (orange) disc. This figure clearly reveals a rather universal evolution of the shock strength: around $\tau = \tau_0$ defined by the equation (15) a shock appears, $\Delta \chi$ becomes non-zero and rapidly reaches its maximum value, after which it gradually falls off, showing a scaling close to that expected from Burgers equation, $\Delta \chi \propto \tau^{-1/2}$. For all masses, we find that the outer wake produces a stronger shock, consistent with higher wave amplitudes at excitation for the fiducial disc parameters. Comparing $\Delta \chi$ obtained by this method to the jumps in azimuthal profiles of $\chi$ in Figs 3, 5, we find good agreement.

The previously noted approximate universality of the shock strength behaviour motivates us to provide a fitting formula that would uniformly approximate $\Delta \chi(\tau)$ behaviour in Fig. 7 for different $M_p$. To this effect, we converged on a smoothly broken two-component power law for $\Delta \chi$ as a function of $\tau$ with $\tilde{\tau} \equiv \tau - \tau_0$ (see equation 16 for $\tau_0$):

$$\Delta \chi(\tau) = A \left( \frac{\tilde{\tau}}{\tilde{\tau}_b} \right)^{-\alpha_1} \left\{ \left[ 1 + \left( \frac{\tilde{\tau}}{\tilde{\tau}_b} \right)^{1/\Delta} \right] \right\} \left( \tilde{\tau}_1 - \tilde{\tau}_2 \right)^{\Delta},$$  

(30)

where $A$ is an amplitude and $\tilde{\tau}_b$ marks the breaking point between the two asymptotic slopes $\alpha_1$ and $\alpha_2$. The parameter $\Delta$ determines the width of this transition. Our derived set of the fit parameters is shown in Table 1. It best approximates the results of the intermediate mass $M_p = 0.25M_\text{th}$ simulation, see Fig. 7(c). Despite certain deviations of this fit from the data in the very low $M_p \ll M_\text{th}$ and $M_p \sim M_\text{th}$ regimes, it still covers a broad range of $M_p$ values reasonably well.

Note that for large values of $\tau$ in the inner disc, the shock strength inferred from the simulations displays a second peak. It is associated with the emergence of the secondary spiral arm and its evolution into a shock. Since the appearance of a secondary arm is a linear effect (Miranda & Rafikov 2019a), taking place at a fixed distance from the planet, the shift of the second peak towards higher $\tau$ for larger $M_\text{p}$ can be well understood using equation (18). Once the secondary shock has formed, our fit (30) no longer works in the inner disc.

In the outer disc, the rising $\Delta \chi$ that is seen for large $\tau$ at low planet masses (0.05–0.25$M_\text{th}$) is simply due to the fact that the wave reaches the damping zones, making our estimate of the shock strength invalid.
In linear theory the wave shape (defined as the $R - \phi$ curve along which $\eta = \text{const.}$) should be given by $\phi = \phi_{\text{lin}}(R)$, see equation (7). However, non-linear effects cause broadening of the wake, which steadily displaces wake maximum in the azimuthal direction compared to the linear prediction, see Fig. 5 for a clear illustration of this phenomenon (note that $\eta \propto \phi - \phi_{\text{lin}}$).

Weakly non-linear wave evolution theory (Landau & Lifshitz 1959, GR01) predicts that, in the $N$-wave regime, the broadening of the wake in the $\eta$ coordinate (as well as the $\eta$-displacement of its peak) behaves as $\Delta \eta \propto (\tau - \tau_0)^{1/2}$, where $\tau_0$ (equation 16) corresponds to the point where wave excitation is mostly complete (GR01). Then equation (19) suggests that the azimuthal deviation $\Delta \phi_{\text{sh}} = \phi_{\text{sh}} - \phi_{\text{lin}}$ of the shock position from the linear prediction $\phi_{\text{lin}}$ should scale as

$$\Delta \phi_{\text{sh}} \propto \text{sign}(R - R_p) \ h_p \ (\tau - \tau_0)^{1/2}.$$  

From this one would expect the position of the shock front corrected for the non-linear effects to be given by

$$\phi_{\text{sh}}^{\text{nonlin}}(R) = \phi_{\text{lin}}(R) + \text{sign}(R - R_p) \Delta \phi_0 \ h_p \ (\tau - \tau_0)^{1/2},$$  

where $\Delta \phi_0$ is the only fitting constant, which is fixed using our simulations as shown below. A similar approach has been used by Zhu et al. (2015), who studied spiral shocks for higher $M_p > M_{\text{sh}}$.

In our runs the shock front is found by searching for a maximum value of $\partial_\phi \ln \Sigma$ at a fixed $R$, a method which has been used before by Arzamasskiy & Rafikov (2018) to locate the wave front as the steepest part of its profile. In the inner disc, care has to be taken at radii where the secondary arm forms and eventually creates a shock. As we have seen in previous sections, both linear and non-linear effects can lead to a shift of the wave front, contributing to $\Delta \phi_{\text{sh}}$ in the inner disc. Recall that even for the global linear calculation, the front of the primary wake is not centred at $\eta = 0$ (see Fig. 3).

In Fig. 8, we display the global wave pattern in a simulation of the fiducial disc model with $M_p = 0.25M_{\text{sh}}$, and compare it with both linear (equation 7) and non-linear (equation 32) predictions. Around excitation, the linear wake prediction $\phi_{\text{lin}}$ (centred on $\eta = 0$) is already slightly offset from the wave front (see Figs 3a and b). As the wave propagates away from the planet, the disagreement in $\phi$ between the $\phi_{\text{lin}}$ and the actual position of the shock increases, mainly due to the aforementioned non-linear effects. Additionally, in the inner disc, the secondary arm forms behind the primary at $R \approx 0.66 \theta_p$, eventually passing through $\phi_{\text{lin}}$.

In Fig. 9, we plot the azimuthal deviation from the linear prediction $\Delta \phi_{\text{sh}}$ found in simulations as a function of $\tau$, for two different planet masses, $M_p = 0.05M_{\text{sh}}$ and $0.25M_{\text{sh}}$. One can see that $\Delta \phi_{\text{sh}}$ steadily increases with $\tau$ as the non-linear effects accumulate in the course of wave propagation. We fit these numerical results with a behaviour in the form (31), determining the values of the coefficients in front of the scaling as indicated in the legend of Fig. 9(a) (which uses $h_p = 0.05$). We find that in a fiducial disc the value of $\Delta \phi_{\text{sh}}$ fitted to the numerical data varies by only a few per cent as $M_p$ changes from 0.05$M_{\text{sh}}$ to 0.5$M_{\text{sh}}$. The dependence of $\Delta \phi_{\text{sh}}$ on $h_p$ or surface density slope is studied in Section 7, and shown to be negligible. This is expected as any dependence on these variables and $M_p$ is already absorbed in the equation (32) and the definition of $\tau$, see equation (18).

The differences between the inner and outer discs are rather small, even despite the emergence of the secondary arm in the inner disc. Thus, for simplicity, in the following we use

$$\Delta \phi_0 \approx 1,$$  

which works well both in the inner and outer disc.

### 5.2 Location of the shock front

Another important characteristic of a planet-driven density wave is its global shape in the $R - \phi$ coordinates. Understanding the factors determining the shape of a density wave is crucial for interpreting observations of the spiral arms in protoplanetary discs (e.g. Zhu et al. 2015; Bae & Zhu 2018). Also, as we will show in Section 6, the geometry (curvature) of the spiral shock has direct effect on the generation of vortensity. Here we outline theoretical expectations for the shape of the spiral shock and compare them to simulations.

### Table 1. Parameters of the fit to $\Delta \chi(\tau)$ (see equation 30).

|          | $A$   | $b_0$ | $\alpha_1$ | $\alpha_2$ | $\Delta$ |
|----------|-------|-------|------------|------------|----------|
| Inner disc | 2.07  | 0.300 | -10.84     | 0.505      | 0.623    |
| Outer disc| 3.11  | 0.181 | -8.63      | 0.525      | 0.766    |
Evolution of planet-driven density waves

Figure 8. Global pattern of the primary wave front, defined as the steepest part of the azimuthal density profile for the fiducial disc and planet with \( M_p = 0.25 \) \( M_{\text{th}} \). The blue and green dots mark locations of the primary and secondary arm from the simulation, respectively. The red dashed curve follows the theoretical linear prediction (equation 21) and the black dash-dotted curve is a theoretical fit given by the equations (32) and (33). The radial separation \( l_{\text{sh}} \) at which the shock forms is marked by vertical dashed lines. See Section 5.2 for details.

Figure 9. Deviation of the azimuthal shock location from linear theory prediction \( \Delta \phi = \phi_{\text{sh}} - \phi_{\text{lin}} \). Simulations results for fiducial disc parameters with \( M_p = 0.05 \) \( M_{\text{th}} \) (top) and \( M_p = 0.25 \) \( M_{\text{th}} \) (bottom) are shown. The blue and orange dots mark simulation results for the outer and inner disc, respectively. A fit for equation (31) is given by the red and green dashed lines, with a proportionality constant indicated in the legend.

6 VORTENSITY EVOLUTION DUE TO PLANET-DRIVEN SHOCKS

The vortensity (or potential vorticity) of a flow \( \zeta \) is given by the ratio of vorticity \( \omega = \nabla \times \mathbf{u} \), where \( \mathbf{u} \) is the fluid velocity, and density \( \rho \). For a 2D disc it reduces to the \( z \)-component of vorticity divided by the surface density \( \Sigma \):

\[
\zeta = \frac{\omega_z}{\Sigma}. \tag{34}
\]

For an axisymmetric background state, equation (34) reduces to

\[
\zeta = \frac{1}{R \Sigma} \frac{d}{dR} (R^2 \Omega) = \frac{\kappa^2}{2 \Sigma \Omega}, \tag{35}
\]

where \( \kappa \) is the epicyclic frequency in the disc (Lin & Papaloizou 2010). In a purely Keplerian disc, the initial vortensity is given by

\[
\zeta_{0, K}(R, p) = \frac{\Omega_p}{2 \Sigma_p} \left( \frac{R}{R_p} \right)^{p-3/2}, \tag{36}
\]

which is independent of \( R \) for \( p = 3/2 \). Corrections due to sub-Keplerian rotation are small for thin discs.

In barotropic 2D flows that are free of shocks, the vortensity is constant along streamlines and obeys

\[
\frac{D \zeta}{Dt} = \frac{\partial \zeta}{\partial t} + \mathbf{u} \cdot \nabla \zeta = 0. \tag{37}
\]

However, when the shocks are present \( \zeta \) is no longer conserved and experiences a discontinuous jump at the shock front.

\[^9\text{Note that } \zeta_{0}(R) \text{ is slightly different from } \zeta_{0, K} \text{ due to the modification of the } \Omega \text{ profile by the radial pressure gradient.}\]

\[^{10}\text{We comment on long-term evolution and corrections to this simplifying assumption in Section 6.4.}\]
6.1 Vortensity generation by the global shock

We now turn to understanding the production of vortensity at the front of a planet-driven shock. For a globally isothermal EoS, the jump of $\zeta$ that a fluid parcel experiences upon crossing a shock of strength $\Delta \Sigma/\Sigma$ is given by (Kivelson 1997; Lin & Papaloizou 2010; Dong et al. 2011b)

$$\Delta \zeta = \frac{c_s}{2 \Sigma \mathcal{M}^2} \left( \frac{\Delta \Sigma}{\Sigma} \right)^2 \frac{d}{dS} \left( \frac{\Delta \Sigma}{\Sigma} \right),$$

$$= \frac{c_s}{2 \Sigma (1 + \Delta \Sigma/\Sigma)^{5/2}} \frac{d}{dS} \left( \frac{\Delta \Sigma}{\Sigma} \right),$$

where $S$ is distance measured along the shock, increasing away from the planet and we used $\mathcal{M}^2 = 1 + \Delta \Sigma/\Sigma$ as appropriate for the isothermal EoS, where $M$ is the Mach number of the flow normal to the shock front. The sign of $\Delta \zeta$ is solely determined by the derivative of shock strength along the shock in this case, since all other factors are positive.

Extending the local calculation of Dong et al. (2011b), we use equation (39) to derive the following analytical expression for the vortensity jump at the shock that is valid globally (see Appendix C for details):

$$\Delta \zeta(R) = \frac{c_s}{2^{1/4} \Sigma \mathcal{M}^2} \left( \frac{M_p}{M_b} \right)^3 \frac{d}{dR} \left( \frac{B(R)}{2^{3/4} \mathcal{M}^2} |\partial \chi(\tau)| \right)^{-5/2} \times \left[ 1 + \frac{M_p}{M_b} \frac{B(R)}{2^{3/4} \mathcal{M}^2} \frac{\Delta \Sigma}{\Sigma} \right]^{5/2} \times C(R) \frac{d}{dR} \left[ \frac{B(R)}{2^{3/4} \mathcal{M}^2} |\partial \chi(\tau)| \right],$$

where $\tau = \tau(R, R_p/M_b)$. The scaling functions $B(R)$ and $C(R)$ are due to conversion between $(\Delta \Sigma/\Sigma_0)(R)$ and $\chi(\tau)$ and the geometry of the shock, respectively. We note that the term in square brackets is $M^{-3}$ and is typically of the order of unity for the shock strengths considered here.

While equation (40) shows similarity to the local expression for $\Delta \zeta$ in Dong et al. (2011b), we note that due to the non-linear shear and the more complex shock geometry appearing in the global cylindrical disc, this formula for $\Delta \zeta$ does not reduce to an expression in terms of $\tau$ alone. However, we can still evaluate this equation numerically for different disc parameters and planet masses and compare the results directly to fully non-linear simulations to verify the analytical prediction (40).

6.2 Semi-analytical model for vortensity production by the planet-driven shock

Results of Section 5 provide us with theoretical expectations for the global behaviour of the shock strength $\Delta \Sigma/\Sigma_0$ and the shock position as functions of disc parameters and planet mass. Combining them with the calculations in Section 6.1, we can formulate a semi-analytical prescription for vortensity generation by the planet-driven shock, based on the framework of RR02 (see Section 2.3).

Given a set of disc and planet parameters – $p, h_p, M_p/M_b –$ this prescription consists of the following sequence of steps:

1. Use the definition (18) to compute $\tau(R)$.
2. Use the fit equation (30) with parameters in Table 1 to predict shock strength $\Delta \Sigma_p$ as a function of $\tau(R)$.
3. Utilize an expression for the shock position $\phi_{sh}$, which sets $C(R)$, see equation (C5). We will examine two approximations for $\phi_{sh}$ in this work:

   (i) A simple linear prediction for shock position, equation (7), giving $\phi_{sh}(R) = \phi_{sh}(R_0)$.

   (ii) A more sophisticated prescription $\phi_{sh}(R) = \phi_{sh}^{\text{nonlin}}(R, \tau(R))$, equation (32), accounting for non-linear effects.

4. Plug the resultant expressions for $\phi_{sh}(R)$ and $\Delta \chi(R)$ into equation (40) to finally obtain the radial profile of the vortensity jump across the shock $\Delta \zeta(R)$.

This recipe for calculating $\Delta \zeta(R)$ is tested and validated in the rest of the paper.

6.3 Simulation results for vortensity generation in a fiducial disc model

We now turn to vortensity generation observed in our simulations and compare it to the semi-analytical model described above.

The rate of change of vortensity per unit time is directly related to the vortensity jump $\Delta \zeta$ experienced in a single shock crossing via the synodic period:

$$\frac{\partial \zeta(R)}{\partial t} = \Delta \zeta(R) \times \frac{|\Omega(R) - \Omega_p|}{2\pi},$$

where $\Omega_p$ appears since the shock is stationary in a frame that is corotating with the planet. We use this relation to obtain $\Delta \zeta$ from the time derivative of the azimuthally averaged vortensity, which is measured in our simulations. This time derivative is evaluated using snapshots separated by $\Delta t = 10$ planetary orbits (we found no dependence on the exact value of $\Delta t$, changing it within a few tens of orbits). For comparison with theory, we perform this calculation early on, after 20–30 planetary orbits, to ensure that $\Sigma$ is unperturbed by the wave damping. We comment on the effects of long-term perturbations to the disc in Section 6.4 (see also Fig. 12). In discs with a radial vortensity gradient, radial advection of vortensity can also add to variation of $\zeta$ at fixed $R$, see equation (37). However, we find that in our runs the advective contribution is negligible and shock-induced vortensity generation is by far the dominant source of $\zeta$.

In Fig. 10, we display the results for $\Delta \zeta(R)$ obtained from our fiducial disc model simulations (blue curves), with $M_p/M_b$ ranging from 0.01 to 1. Radial profiles of $\Delta \zeta$ are derived from our non-linear simulations after 20 planet orbits using equation (41) in which we approximate $\partial \zeta(R) = \partial \zeta(\chi(R))$. The vertical axis is rescaled by the third power of $M_p/M_b$, to absorb the most important scaling with $M_p$ in equation (40). Radial distance from the planet is rescaled by the shock distance $l_{sh}$, which is a natural length-scale of the problem in the local limit (GR01). These renormalizations lead to $\Delta \zeta(R)$ profiles that have very similar radial shape. However, slight variations are still visible as the planet mass varies, they are discussed below.

Before the shock forms, vortensity is mostly conserved, except for numerical artefacts that are amplified for the lowest planet masses. Also, in the coorbital region of the planet we observe the well-known numerical artefacts that are amplified for the lowest planet masses. As the density wave starts turning into a shock beyond $|R - R_p| = l_{sh}$ (marked by vertical dashed lines), a narrow positive peak in $\Delta \zeta$ forms between $1 \leq |R - R_p|/l_{sh} \leq 2$, followed by a wider but shallower negative trough at larger distances from the planet. The former is due to the rapid formation of the shock and the latter due to its subsequent decay. The overall shape of the $\Delta \zeta(R)$ curves is roughly the same in the inner and outer discs, although the outer $\Delta \zeta$ peak is higher than the inner one, illustrating the importance of global effects (in the local calculations of Dong et al. (2011b) the
valid and the code performs as expected. This agreement also shows that equation (40) is well fitted with max $\Delta \zeta$ in the outer disc as a function of planet mass for a fiducial disc model. We show results of simulations with several resolutions, which demonstrate that low planet masses require high resolution to show convergence. The dotted line shows the $\Delta \zeta \propto M_p^3$ scaling (see equation 40) familiar from the local study of Dong et al. (2011b).

In the highest mass case ($M_p = M_{th}$), the agreement with the simulation results is poor, since non-linear effects are strong and the shock forms while the wave is still accumulating angular momentum, in contrast to the assumptions of the GR01 and RR02, see Section 2.3. This spatial overlap of the excitation and propagation regions is also responsible for the shift of vortensity peaks away from $|R - R_p| = l_{th}$ at high $M_p$. This is because in the global case the rescaling of radial separation by $l_{th}$ is not identical to expressing the distance from the planet through the coordinate $\tau$.

Semi-analytical curves computed for the two different approximations for $\phi_{th}$ (see Section 6.2) differ from each other only for $M_p \lesssim 0.5 M_{th}$. This is expected, since at low masses non-linear effects are weak and $\phi_{th}^{\text{min}} \approx \phi_{th}$. We find that, for practical purposes, using the simple approximation $\phi_{th} = \phi_{th}^\text{min}$ for semi-analytical calculation of $\Delta \zeta(R)$ is at least not inferior to the more sophisticated assumption $\phi_{th} = \phi_{th}^{\text{min}}$.

Fig. 10 clearly illustrates that over a wide range of planet masses (0.05 − 0.5$M_{th}$), vortensity generation exhibits the self-similar behaviour expected in the framework of RR02, as long as the distance from the planet and $\Delta \chi$ amplitude are scaled according to theory. The vortensity amplitude scaling is additionally illustrated in Fig. 11, where we show the peak value of $\Delta \zeta$ in the outer disc as a function of planet mass in simulations using the fiducial disc model. One can see a clear power-law behaviour scaling that is well fitted with $\max \Delta \zeta \propto (M_p/M_{th})^3$ (dotted line in the figure). This scaling was found in Dong et al. (2011b) in the local (shearing-sheet) approximation, but clearly is valid in the global case as well. This is not surprising since $\Delta \zeta \propto M_p^3$ is the dominant dependence on $M_p$ in the global equation (40), with additional $M_p$-dependent terms playing a minor role.

Fig. 11 also illustrates the numerical convergence of our results over a wider range of resolutions, indicated by different symbols. One can see that convergence is very good, except at the lowest peaks are symmetric). Also, at small $M_p$ one can see a low amplitude secondary peak in the inner disc that recedes from the planet (in $|R - R_p|/l_{th}$ coordinate) as $M_p$ increases. This peak appears due to the shocking of the secondary arm.

Before detailed testing of the prescription outlined in Section 6.2, we checked the validity of the equation (40) by using it to compute $\Delta \zeta(R)$ with the main inputs obtained directly from our simulations – shock strength $\Delta \chi(R)$ from angular momentum flux decay, equation (27) and shock shape $\phi_{th}(R)$ – instead of using semi-analytical fits (30), (32), (33). This calculation is illustrated for $M_p = 0.25 M_{th}$ by magenta crosses in Fig. 10(d), providing an excellent match with simulation results and confirming that the shock is the only relevant source of vortensity. This agreement also shows that equation (40) is valid and the code performs as expected.

The green dotted and orange dashed curves in Fig. 10 illustrate the semi-analytical prescriptions for $\Delta \zeta(R)$ (described in Section 6.2), computed for $\phi_{th} = \phi_{th}^{\text{min}}$ and $\phi_{th} = \phi_{th}^{\text{nonlin}}$, respectively. By design, these prescriptions agree best with simulation results for $0.1 \leq M_p/M_{th} \leq 0.5$, since the parameters (see Table 1) of the fit (30) were determined for $M_p = 0.25 M_{th}$. The discrepancies at low $M_p$ arise most likely because of the increased importance of linear (validity of the fit for $\Delta \chi$) and numerical effects for small-amplitude density waves (see the discussion of Fig. 11 below).

We use a polynomial spline fit over a few cells in order to approximate the numerical parameters of the shock, which is needed to reduce the noise in computation of $\Delta \zeta$.

Figure 10. Radial profiles of the vortensity jump across the shock $\Delta \zeta(R)$. We compare simulation results (blue) with semi-analytical prediction described in Section 6.2, that use different inputs for the shock location $\phi_{th}$: $\phi_{th}^{\text{min}}$ (green dotted) and $\phi_{th}^{\text{nonlin}}$ (orange dashed). The dark grey area marks regions where vortensity changes, if present, are not due to the spiral shocks; vertical dashed lines mark $|R - R_p| = l_{th}$. Magenta crosses in panel (d) show $\Delta \zeta$ computed using equation (40) with inputs for shock position and strength obtained directly from simulations, showing excellent agreement with observed vortensity generation. For more details see Section 6.3.

Figure 11. Scaling of the peak value of the vortensity jump $\Delta \zeta$ in the outer disc as a function of planet mass for a fiducial disc model. We show results of simulations with several resolutions, which demonstrate that low planet masses require high resolution to show convergence. The dotted line shows the $\Delta \zeta \propto M_p^3$ scaling (see equation 40) familiar from the local study of Dong et al. (2011b).
First, suppression of the surface density near the planet reduces the strength of its gravitational coupling to the disc, resulting in weaker planetary torque and lowering the density wave amplitude around the planet’s orbit takes place, eventually resulting in gap opening. This will modify the planet-driven wave in several ways.

6.4 Effect of gap opening

Up to this point we always assumed the disc surface density near the planet to be unperturbed and vary only on scales of order $R_p$. However, as the angular momentum lost by the damping density wave is absorbed by the disc fluid, a radial redistribution of gas around the planet’s orbit takes place, eventually resulting in gap opening. This will modify the planet-driven wave in several ways.

First, suppression of the surface density near the planet reduces the strength of its gravitational coupling to the disc, resulting in weaker planetary torque and lowering the density wave amplitude at excitation (Petrovich & Rafikov 2012). Secondly, as the wave propagates across an inhomogeneous background at the gap edge, its non-linear steepening and decay get modified as described in RR02 equation (2). Different from the fiducial disc, this model has a global non-zero vortensity gradient, since its time-derivative (panel b), and the surface density profile (panel c). Results are shown for the fiducial disc model and $M_p = 0.25M_\text{th}$. For this point, since $\delta \Sigma$ is still relatively small, we observe only localized changes of $\delta \zeta$ profile: its peak broadens and goes down in amplitude, also showing a small kink around the maximum value (most likely due to vortex formation at late times). On longer timescales, when the gap gets deeper, we expect vortensity production at the shock to be modified more severely.

7 RESULTS: VARIATION OF THE DISC PARAMETERS

Next we examine how our results and the comparison with semi-analytical calculations change as we vary the underlying disc model. We will predominantly focus on the AMF behaviour and vortensity generation as our metrics for comparison since they provide very sensitive diagnostics of the non-linear wave evolution.

7.1 Variation of the surface density slope

First, we turn to a disc model that has the fiducial value of $h_0 = 0.05$ but a constant background surface density, i.e. $p = 0$, see equation (2). Different from the fiducial disc, this model has a global non-zero vortensity gradient, since its $\zeta_{\phi}(p = 0) \propto (R/R_p)^{-3/2}$, see equation (36). This may somewhat increase the role of the advective transport of vortensity. Given the generality of the framework outlined in Section 2.3, we expect the effects of a different slope of $\Sigma_\phi(R)$ on wave evolution to be fully absorbed in the definitions of the coordinates $\chi$ and $\tau$.

We begin by presenting in Fig. 13 the behaviour of AMF for a planet of mass $M_p = 0.01M_\text{th}$. Comparing it to Fig. 4, we see very (increased) surface density. This steady increase of $\delta \zeta$ eventually produces large radial gradients of vortensity near the planet that can lead to non-negligible radial advective transport of vortensity (which we normally find to be unimportant) during gap opening.

At early times, $\partial_t \langle \delta \zeta \rangle_{\phi}$ exhibits a very stable, time-independent profile that we use in Fig. 10. But once gap depth reaches $\delta \Sigma/R_0 \approx 0.15$, we find $\partial_t \langle \delta \zeta \rangle_{\phi}$ to deviate from the initial smooth profile. At this point, since $\delta \Sigma$ is still relatively small, we observe only localized changes of $\partial_t \langle \delta \zeta \rangle_{\phi}$ profile: its peak broadens and goes down in amplitude, also showing a small kink around the maximum value (most likely due to vortex formation at late times). On longer timescales, when the gap gets deeper, we expect vortensity production at the shock to be modified more severely.
Next we examine the vortensity generation. Our semi-analytical model (Section 6.2) provides a good match to simulation results. In Fig. 15, we show $\Delta \zeta$ as a function of $h_p$ (rows) for $M_f/M_{th} = 0.1$ (left-hand column) and $M_f/M_{th} = 0.25$ (right-hand column). For both masses, the general shape of $\Delta \zeta$ is maintained and $\Delta \zeta$ curves appear very similar when $R - R_p$ is rescaled by $l_{sh}$. Thus, our semi-analytical model (Section 6.2) provides a good match to simulation results.

8 DISCUSSION

The main goals of this work were to numerically verify the semi-analytical theory of planet-driven density wave propagation advanced in RR02, and to explore the impact of the emergence of secondary (and higher order) spiral arm on wave evolution in the inner disc. Using high-resolution ATHENA+ simulations we were able to confirm the main predictions of the semi-analytical framework of RR02 summarized in Section 2.3.

In particular, we find that the change of coordinates from $R, \phi, \Sigma$ to $\tau, \eta, \chi$ works very well when describing the planet-driven density wave evolution, as long as certain problem-specific inputs are properly calibrated using simulations. Examples of this include the scaling of $\phi_{sh}^{\text{nonlin}} - \phi_{sh}$ with $h_p$ and $\tau$ (see equation 32 in Section 5.2), self-similarity of the shock strength profile (see equation 30 in Section 5.1), semi-analytical calculation of the vortensity production in Section 6.2, and so on. In agreement with the findings of Dong et al. (2011b) and Duffell & MacFadyen (2012), we find very good agreement between the theoretically predicted shocking distance $l_{sh}$ (see equation 17) and our global simulation results, even though $l_{sh}$ was derived in GR01 in the local (shearing sheet) approximation.

Figure 14. Same as Fig. 10 but for constant surface density disc ($p = 0$). Note that, compared to the $p = 3/2$ disc (see Fig. 4), the peak of $\Delta \zeta$ is now higher in the inner disc. We omit simulation data in the corotation region (marked grey), where vortensity changes are due to advective mixing of material.

similar behaviour close to the planet: after wave excitation, its AMF reaches values comparable to the fiducial case and begins to decay at a similar distance. Thus, the shocking length $l_{sh}$ is still captured by the equation (17). Moreover, the $R - \phi$ shape of the wake is still accurately fit by the equations (32) and (33).

However, on the global scale, we see that wave damping is accelerated (slowed down) in the inner (outer) disc. As a consequence, (less) AMF reaches the damping boundary. This is true for the full simulation as well as for the solutions of the Burgers equation and can be directly understood from the effect of $p$ on $\tau$: according to Fig. 2, $\tau$ reaches higher (lower) values in the inner (outer) disc, as compared to the fiducial disc. This behaviour is observed across all planet masses. Qualitative differences between evolution described by the full set of hydro equations (ATHENA +) and by Burgers equation (see Section 4) still remain.

Next we examine the vortensity generation. Our semi-analytical framework should provide a good match for the $\zeta$ production at the shock when $p = 0$, since all variables in our framework naturally account for arbitrary $p$. In Fig. 14, we compare simulation results against the semi-analytical model (Section 6.2) for $p = 0$. The filled grey area marks the corotation region, where the horseshoe flow mixes regions of different initial vortensity advectively. We do not show simulation data in this region as $\zeta$ changes there are not due to shocks. First we note that, as opposed to the fiducial disc, the inner peak now dominates over the outer one in the linear regime ($M_p/M_{th} = 0.01$). Very importantly, this behaviour is correctly captured by our semi-analytical model (orange dashed curve). For the three higher planet masses ($0.1 \leq M_p/M_{th} \leq 0.5$), the agreement between our semi-analytical theory and simulations is very good.

7.2 Variation of the disc scale-height

Next we explore the effect of variation of the disc aspect ratio by studying simulations with $h_p = 0.07$ and 0.1. These are performed at the same resolution as the fiducial models that we deem converged, effectively increasing the number of grid cells per scale-height. Thus, simulations at larger $h_p$ are expected to be converged too, which we have confirmed.

Variation of $h_p$ is expected to have a direct effect on the amplitude of the wave perturbation at excitation. For a fixed planet mass, higher $h_p$ results in smaller amplitude of the perturbation (since $M_{th}$ is higher and $M_f/M_{th}$ is lower) and the density wave carries less angular momentum, see equation (6). But if we keep $M_f/M_{th}$ fixed, then the wave amplitude, expressed through $\chi$, remains the same. From equation (18), we also find $\tau \propto h_p^{-3/2}$, meaning that non-linear evolution, expressed via $\tau$-coordinate, is slowed down in hotter discs. In terms of radial distance, this dependence is approximately absorbed in the dependence of $l_{sh}$ on $h_p$, as we demonstrate below. And regarding the $R - \phi$ shape of the wake (Section 5.2), we find that the linear dependence on $h_p$ of the non-linear correction term $\phi_{sh}^{\text{nonlin}} - \phi_{sh}$ in equation (32) provides an excellent match to the simulation results.

In Fig. 15, we show $\Delta \zeta$ as a function of $h_p$ (rows) for $M_f/M_{th} = 0.1$ (left-hand column) and $M_f/M_{th} = 0.25$ (right-hand column). For both masses, the general shape of $\Delta \zeta$ is maintained and $\Delta \zeta$ curves appear very similar when $R - R_p$ is rescaled by $l_{sh}$. Thus, our semi-analytical model (Section 6.2) provides a good match to simulation results.

Note that as $h_p$ increases, peak values of $\Delta \zeta$ increase (decrease) in the outer (inner) disc. We traced this behaviour to the effect of $h_p$ on the initial (linear) wake profile: for higher $h_p$, we find the magnitude of minimum and maximum values of $\chi$ to increase (decrease) in the outer (inner) disc. But the main effect of changing $h_p$ is the variation of the spatial scale of the problem. Since the rescaled (horizontally, by $l_{sh} \propto h_p$) profiles of $\Delta \zeta$ in Fig. 15 remain largely unchanged, this implies that in physical space (in $R - R_p$) the vortensity profile is broader for higher $h_p$, see Fig. 16 for illustration.
The emergence of the secondary spiral arm in the inner disc certainly affects the performance of the RR02 theory, which did not take this subtle linear effect into account. In the inner disc, and especially at low $M_p$, that theory (as well as its local analogue, see GR01) misses the appearance of a second bump in the profile of $\Delta \chi$ at the shock as a function of $\tau$ (see Fig. 7), slow decay of the AMF $F_J$ in the inner disc (see Fig. 4), and secondary peaks of vortensity production (see Fig. 10). For these reasons, in the inner disc the semi-analytical framework of RR02 is valid only for $R \gtrsim (0.5 - 0.6)R_p$, before the secondary arm fully forms.

### 8.1 Validity of Burgers equation for describing evolution of planet-driven density waves

While, as described above, many predictions of the semi-analytical theory of RR02 work very well, once properly calibrated by simulations, some other aspects of theory, starting from first principles, show only qualitative agreement in many cases. In particular, when we attempt to reproduce wave evolution by directly solving the Burgers equation (rather than using simulations to calibrate semi-analytical prescriptions), we find quantitative differences with the simulations. They can be clearly seen in Fig. 3, which reveals a faster decay and slower azimuthal spreading of the density wave profile evolved using Burgers equation (12), as compared to the simulation results, for a low mass planet. Unsurprisingly, this leads to a considerably faster decay of the Burgers $F_{WKB}$ compared to both $F_J$ and $F_{WKB}$ in simulations, see Fig. 4.

This may seem surprising given that the Burgers equation (12) uses the simulated wake profile near the planet as a starting point for its subsequent evolution. This implies that Burgers equation is not capturing well some wave propagation physics that takes place in the actual simulation. We believe that, at least partly, this missing ingredient is the continued injection of the angular momentum into the density wave by the planetary potential happening in simulations far from the planet, outside of the nominal excitation region. Since the wave amplitude decays due to its dissipation at the shock, addition of even a small amount of angular momentum to the wave...
far from the planet can significantly slow down the decay of $F_1$ in simulations. By design, this injection is absent in the Burgers equation approach, resulting in the discrepant $F_{JWKB}$. An indirect support for this possibility can be seen in the less discrepant $F_{JWKB}$ for higher $M_p = 0.25M_\oplus$ planet, see Fig. 6: due to faster decay of the wave amplitude in the higher $M_p$ simulation, less angular momentum gets added to the wave far from the planet by its torque.

To test this hypothesis in a more direct way we performed an additional $M_p = 0.25M_\oplus$ simulation in which we artificially suppressed the planetary potential beyond some truncation radius $d$, thus preventing the continuing injection of angular momentum flux into the wave far from the planet. This is one way in which a freely propagating density wave can be realized (cf. Arzamasskiy & Rafikov 2018), bringing wave evolution closer to the regime described by the Burgers equation. In practice, we modify the potential (24) using a cut-off function, such that $\Phi_p(d) \rightarrow f(d) \times \Phi_p(d)$, where $d$ is the distance from the planet,

$$f(d) = \begin{cases} 1 & \text{for } d \leq d_t, \\ A \exp(-|d - d_t|/w) & \text{for } d > d_t, \end{cases}$$

$d_t = 0.11R_p \approx 2.2H_p$ is the truncation distance, $w = 0.01R_p$ is the characteristic scale for a potential decay, and $A$ is a normalization constant making the resulting potential continuous.

The AMF behaviour resulting in this run is shown in Fig. 6 via green solid curve ($F_J$) and brown crosses ($F_{JWKB}$), see the legend. One can see that around excitation, the original simulation and the one with truncated potential predict very similar AMF behaviour. But further out, the AMF for a truncated potential decays faster than for the non-truncated $F_J$, which can be seen by comparing blue and green curves. Both $F_J$ and $F_{JWKB}$ for a truncated potential stay closer to the Burgers $F_J$ (red points). This provides support to our hypothesis that it is the distant gravitational coupling of the density wave to the planetary potential, that causes the Burgers equation to overpredict the wave decay. On the other hand, the green curve does not fully converge to the red points, which suggests that some other factors may also be at play.

To summarize, while the Burgers equation (12) provides a useful framework for understanding the planet-driven wave evolution at the qualitative level, a more accurate approach (e.g. a theory calibrated by simulations) is needed for quantitative agreement with simulations. This can be important for certain applications, see Bollati et al. (2021) and Section 8.6.

8.2 Vortensity generation at the shock and its semi-analytical description

Our derivations in Section 6.2 and Appendix C provide a natural generalization of the local (i.e. shearing sheet geometry) calculation of the vortensity production in Dong et al. (2011b) to the case of a global (i.e. polar) disc geometry with radial gradients of $\Sigma$. Our calculation of the vortensity generation at the shock can also naturally account for the modification of the global shape of the spiral shock by the non-linear effects (Section 5.2), although we find the resulting corrections to be significant only for high $M_p$, see Section 6.3 and Fig. 10.

This calculation provides a basis for the semi-analytical framework for predicting the vortensity production at the shock, described in Section 6.2. This framework employs a single input – the jump of the dimensionless wave perturbation $\Delta \chi$ at the shock – that is calibrated using simulations. This calibration step yields an accurate description of $\Delta \chi(R)$, that compares very well against simulations, see Sections 6.3 and 7. This, in particular, implies that the radial transport of vortensity has negligible effect on its evolution in the vicinity of the planet, i.e. the variation of $\chi$ can be explained entirely through its production by the planet-driven shock.

The semi-analytical framework for vortensity production is one of the main results of our work and will be used in future calculations of the planet-induced vortensity evolution. Its accuracy allows us to make inferences about the vortensity evolution without running computationally expensive hydrodynamical simulations.

For example, we used this framework to determine the value of the surface density slope $p$ at which the height of the vortensity peaks in the inner and outer disc would be the same; this question is relevant for determining the side of the disc in which vortices triggered by the planetary perturbation might first appear (Cimerman & Rafikov, in preparation). Our semi-analytical approach allows a very efficient exploration of this problem. Given that we previously found the inner (outer) peak of $\Delta \chi$ to dominate for $p = 0 (p = 3/2)$, one should expect the transition (equal amplitude of the vortensity peaks) to take place for some $p$ in the interval (0, 3/2). However, it turns out that the value of $p$ at which this transition takes place also depends on the planet mass. We demonstrate this in Fig. 16 where we display several semi-analytical $\Delta \chi(R)$ profiles computed for $p = 1$, for three values of $M_p$ and as functions of different coordinates: $R/R_p, (R - R_p)/l_{sh}$ and $\tau(R)$. One can see that for this value of $p$ the inner peak dominates for low $M_p$ (blue curve), while for high $M_p$ the outer vortensity peak is higher (green curve). Our theory predicts that $\Delta \chi$ peaks should be almost equal in amplitude for $M_p = 0.05M_\oplus$. And indeed, in panel (a), we also show data from a simulation run for $M_p = 0.05M_\oplus$ and $p = 1$ (red dashed) that agrees very well with this theoretical prediction (orange solid).

This figure conveniently illustrates several other important points about the vortensity evolution: scaling of $\Delta \chi$ with $M_p$ (almost) self-similar appearance of $\Delta \chi$ when expressed in $\tau(R)$ and $(R - R_p)/l_{sh}$ coordinates, and the increase of radial scale of $\Delta \chi$ variation as $M_p$ is decreased, see panel (a).

8.3 Effect of equation of state and disc thermodynamics

Our calculations are restricted to globally isothermal discs. At the next level of sophistication, one might wonder what changes would be brought in by considering e.g. the locally isothermal EoS, in which $c_s$ is a function of $R$. We expect two modifications to be introduced by this EoS.

First, Lin & Papaloizou (2010) showed that in a locally isothermal disc the vortensity jump across a shock differs from equation (39) by an additional (baroclinic) term:

$$\Delta \xi = \frac{c_s}{2\Sigma} \frac{(\Delta \Sigma/\Sigma)^2}{(1 + \Delta \Sigma/\Sigma)^{5/2}} \frac{d}{d\Sigma} \left( \frac{\Delta \Sigma}{\Sigma} \right) + \frac{2\Delta \Sigma}{\Sigma^2(\Delta \Sigma/\Sigma + 1)^{3/2}} \frac{dc_s}{d\Sigma}.$$  

Lin & Papaloizou (2010) have also argued that this second (baroclinic) term should often be negligible, since $c_s$ varies on scales of order the disc size, whereas $\Delta \Sigma$ across the shock varies over shorter length-scales near the planet. Thus, we expect this term to be unimportant for high mass planets, for which $l_{sh} \ll R_p$. But for slowly decaying, lower amplitude waves driven by the low mass planets ($M_p \ll M_\oplus$), this term might contribute to vortensity generation at the shock12 more strongly.

12Additional baroclinic vortensity generation may occur during evolution of the flow, for example in the corotation region, where horseshoe orbits approach the planet (Paardekooper & Papaloizou 2008) or later on during the onset of the Rossby Wave Instability (Lovelace et al. 1999).
Secondly, the behaviour of $\Delta \Sigma$ is different in locally and globally isothermal discs. Part of the variation comes from the explicit difference in temperature profiles near the planet, but there is also a more subtle contribution. As shown in Miranda & Rafikov (2019b), angular momentum of the density waves propagating in locally isothermal discs is not conserved, unlike in globally isothermal ones. Instead, it scales with the local disc temperature, i.e. $\propto c_s^2$, as a result of additional coupling with the background flow (manifesting itself even in the linear regime). This translates into a different behaviour of $\Delta \Sigma (R)$ compared to the discs considered in this work.

We have explored evolution of the wave angular momentum flux in a locally isothermal disc using FARGO3D and ATHENA++. We found that, especially for sub-$M_{\text{th}}$ planets, wave damping is accelerated or slowed down substantially, beyond the expected effect of changing $\tau$, in agreement with the findings of Miranda & Rafikov (2019b). This changes the $\Delta \Sigma$ and $\Delta \chi (R)$ behaviour in a way that is not obviously generalizable within our current framework. The same is true in the even more general case of a disc with explicit heating/cooling, in which angular momentum flux of planet-driven density waves is known to not be conserved in general (Miranda & Rafikov 2020a, b; Zhang & Zhu 2020).

8.4 Limitations of this work

Our work necessarily employs a number of simplifications, in addition to the ones regarding disc thermodynamics (Section 8.3). One of them is our assumption that the planet-driven density waves decay only as a result of irreversible dissipation at the shock, into which they evolve due to non-linear effects. In real discs there are other, linear, mechanisms that can be responsible for wave damping. In particular, radiative damping has been demonstrated recently (Miranda & Rafikov 2020a) to play a strong role in wave dissipation in protoplanetary discs. Especially for lower mass planets, radiative losses can easily dominate wave dissipation compared to the non-linear damping at the shock (Miranda & Rafikov 2020b). While accounting for the radiative wave damping is possible in principle, as described in Miranda & Rafikov (2020a, b), in practice this procedure may be too cumbersome, leaving direct simulations with explicit heating/cooling as a better option.

We have also neglected the effects of viscosity in our study. While viscous damping of density waves is typically subdominant compared to either linear radiative or non-linear effects (Miranda & Rafikov 2020b), viscosity can smooth out the emerging density gradients and affect the vortensity evolution. However, recent studies (Flaherty et al. 2017, 2018, 2020; Rafikov 2017) have shown that effective viscosities in protoplanetary discs are likely low.

We also neglected the possibility of planet migration, which should result from the asymmetry of the torques exerted by the planet on the disc. Migration can produce additional asymmetry of the surface density and torque distribution around the planet (Rafikov 2002b). In this regard, our results should still be valid on time-scales over which the planet would not migrate significantly.

Finally, our 2D study neglects the possibility of vertical motions in the disc and other related 3D effects. However, previously Zhu et al. (2015) have shown that many aspects of the density wave propagation in 2D discs, including the non-linear effects, directly translate into fully 3D discs (see next).

8.5 Comparison to previous work

Some aspects of the non-linear density wave evolution have been studied numerically in the past. In particular, Duffell & MacFadyen (2012) looked at some properties of the global planet-driven spiral arms using high-resolution 2D simulations. We find good qualitative agreement with the behaviour they reported for $F_{\phi}$ derived the WKB approximation, see equations (13) and (14). Comparing our lowest mass case ($M_p/M_{\text{th}} = 0.01$) to theirs ($M_p/M_{\text{th}} = 0.0209$), we find very good agreement in the inner disc, including the effect of the secondary arm formation on $F_{\phi}$, which is now well understood. In the outer disc, our results show a slower than $F_{\phi} \propto \tau^{1/2}$ decay with $\tau$, whereas Duffell & MacFadyen (2012) find a behaviour close to this scaling. We speculate that his might be due to a different implementation of the wave damping at the outer boundary.

Using local (shearing sheet) simulations, Dong et al. (2011b) studied vortensity excitation at the planet-driven shock. Their results for $\Delta \zeta (R)$ are compatible with ours regarding the amplitude of the vortensity jump and its radial structure. Their predicted scaling of $\Delta \zeta \propto (M_p/M_{\text{th}})^{3}$ is also consistent with our findings. In our global calculations, we find that the amplitude of $\Delta \zeta$ can sometimes differ by a factor of $\approx 2$–3 between the inner and outer disc in the most extreme cases, which was not possible for Dong et al. (2011b) to observe due to the symmetry of the shearing sheet setup.

Both Dong et al. (2011b) and Duffell & MacFadyen (2012) found that the azimuthal width of the wave evolves as $\Delta \eta \propto \tau^{1/2}$, translating into a similarly behaving offset of the peak position of the non-linear wave from the linear prediction (7), in agreement with our findings, see Section 5.2. Moreover, in their 3D simulations of more massive planets embedded in discs Zhu et al. (2015) also found good agreement of the wave offset with this scaling.

8.6 Applications of our results

Semi-analytical framework for characterizing global planet-driven shocks (Rafikov 2002a), further developed and tested in this study, can be applied to improve understanding of various aspects of disc–planet interaction. For example, our results on the shock strength (Section 5.1) can be used to compute the contribution of the planet-driven shock heating (Rafikov 2016) to thermal balance of the disc; they can also be used for computing the associated mass accretion rate $M(R)$ through the disc. Our semi-analytical fit for the offset between the non-linear shock location and $\theta_{\text{ins}}$, see equation (32) in Section 5.2, can be employed for inferring planet masses from the shapes of spirals observed in protoplanetary discs.

Recently Bollati et al. (2021) applied the non-linear framework of GR01 and RR02 to model kinematic signatures (‘kinks’) of planets embedded in the disc, that have been recently observed by ALMA (Pinte et al. 2018, 2019). They used a procedure identical to the one used in making Figs 3(c) and (d) – numerical solution of the Burgers equation (12) with initial conditions from linear theory – to compute the velocity perturbation in the disc due to a planet-driven density wave and to obtain the CO emission channel maps. Although we did not compute a kinematic signature in this work, our results (Section 8.1) suggest that an approach based on solving Burgers equation may easily overestimate wave decay and underestimate the resultant velocity perturbation and the amplitude of a kink.

While the weakly non-linear theory of RR02 strictly applies only to the low planet mass regime, some of its ingredients may be useful also for describing the high-mass ($M_p \gtrsim M_{\text{th}}$) regime, relevant for the existing protoplanetary disc observations in CO emission and scattered light. In particular, we expect that, when expressed in terms of the generalized coordinates (8)–(11), the wave properties would still show a universal behaviour even for high $M_p$, which could be calibrated using simulations.
Our results on vortensity generation (Section 6) can be used for understanding the appearance of vortices in planetary vicinity (Koller, Li & Lin 2003; Li et al. 2005; de Val-Borro et al. 2007). They can also be employed to (semi-analytically) generate axisymmetric surface density profiles for the planet-induced gaps as a function of time, disc parameters and planet mass, following the recipe of Lin & Papaloizou (2010). These applications will be explored in a forthcoming work (Cimerman & Rafikov, in preparation). Finally, our results can be useful for benchmarking the codes used for studying disc–planet interaction, especially through diagnostics such as the angular momentum flux evolution and vortensity production at the shock.

9 SUMMARY

We have studied the non-linear evolution of density waves excited by planets embedded in inviscid, isothermal 2D cylindrical discs with the goal of verifying the weakly non-linear theory of global density waves developed in Rafikov (2002a). Using linear calculations, weakly non-linear theory and full hydrodynamical simulations with ATHENA++ we explored models for a variety of disc parameters, such as disc scale-heights and surface density slopes, and planet masses spanning two orders of magnitude such as disc scale-heights and surface density slopes, and planet evolves (equation 30 and Table 1). Effects, and the strength of the shock into which the wave ultimately of the density wave (equations 32–33), accounting for the non-linear of the key inputs of the semi-analytical framework: the global shape masses. A better approach would be to rely on numerical calibration provide a quantitatively accurate description of the wave evolution, using Burgers equation (12), while qualitatively correct, does not provide a quantitative accurate description of the wave evolution, in particular, overpredicting wave decay, especially for low planet masses. A better approach would be to rely on numerical calibration of the key inputs of the semi-analytical framework: the global shape of the density wave (equations 32–33), accounting for the non-linear effects, and the strength of the shock into which the wave ultimately evolves (equation 30 and Table 1).

(i) The semi-analytical framework of Rafikov (2002a) using rescaled variables $r, \eta, \chi$ (equations 8–11) provides an accurate description of the non-linear evolution of the global density waves driven by the subthermal mass planets, provided that it is properly calibrated using numerical simulations. In particular, many characteristics of the wave exhibit a behaviour close to self-similar when expressed in terms of these variables.

(ii) We showed that calculation of the density wave properties using Burgers equation (12), while qualitatively correct, does not provide a quantitative accurate description of the wave evolution, in particular, overpredicting wave decay, especially for low planet masses. A better approach would be to rely on numerical calibration of the key inputs of the semi-analytical framework: the global shape of the density wave (equations 32–33), accounting for the non-linear effects, and the strength of the shock into which the wave ultimately evolves (equation 30 and Table 1).

(iii) We derive analytical expressions for the vortensity jump across the global planet-driven shock and verify them using numerical simulations (Section 6). We confirm that vortensity generation at the shock scales as a high power of the planet mass, $\Delta \zeta \propto (M_p/M_\oplus){\Delta}$. The data underlying this article will be shared on reasonable request to the corresponding author.

(iv) Applicability of the weakly non-linear theory (Rafikov 2002a) in the disc interior to the planetary orbit is limited by the emergence of secondary spiral arms, which is a linear effect (Bae & Zhu 2018; Miranda & Rafikov 2019a).

(v) Our results have implications for understanding the appearance of planet-driven spiral arms in proplanetary discary, kinematic signatures of embedded planets (‘kinks’), and other phenomena.

Future work will apply the semi-analytical approach explored in this study to other problems in the area of disc–planet interaction, in particular, the formation of vortices at the edges of planetary gaps driven by the Rossby Wave Instability.

ACKNOWLEDGEMENTS

Software: NUMPY (Harris et al. 2020), SCIPY (Virtanen et al. 2020), IPYTHON (Perez & Granger 2007), MATPLOTLIB (Hunter 2007), ATHENA++ (Stone et al. 2020), FARGO3D (Benítez-Llambay & Masset 2016) and ASTROPY (Astropy Collaboration 2013, 2018).

We thank the anonymous referee for helpful suggestions, especially in clarifying our explanation of multiple spiral arm formation in the linear case. We are indebted to Ryan Miranda for many useful exchanges and for sharing his method and code to calculate global linear modes, which greatly enriched this work. We would also like to thank all developers of ATHENA++ for sharing their code with us and in particular Tomohiro Ono, James Stone, and Matt Coleman for helpful discussions. NPC is funded by an Isaac Newton Studentship and a Science and Technology Facilities Council (STFC) studentship. RRR acknowledges financial support through the NASA grant 15-XRP15-2-0139, John N. Bahcall Fellowship, and STFC grant ST/T00049X/1. A large part of the lower resolution simulations were performed on FAWCETT the HPC cluster at DAMTP, University of Cambridge. Special thanks to Kacper Kornet and Deryck Thake for all the support in using local resources. Part of this work was performed using resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service (www.csd3.cam.ac.uk), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/P020259/1), and DiRAC funding from the Science and Technology Facilities Council (www.dirac.ac.uk).

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Arzamasskiy L., Rafikov R. R., 2018, ApJ, 854, 84

Astropy Collaboration, 2013, A&A, 558, A33

Astropy Collaboration, 2018, AJ, 156, 123

Bae J., Zhu Z., 2018, ApJ, 859, 118

Belyaev M. A., Rafikov R. R., Stone J. M., 2013, ApJ, 770, 67

Benisty M. et al., 2015, A&A, 578, L6

Benítez-Llambay P., Masset F. S., 2016, ApJS, 223, 11

Bollati F., Lodato G., Price D. J., Pinte C., 2021, MNRAS, 504, 5444

de Val-Borro M. et al., 2006, MNRAS, 370, 529

de Val-Borro M., Artyomowicz P., D’Angelo G., Peplinski A., 2007, A&A, 471, 1043

Dong R., Rafikov R. R., Stone J. M., Petrovich C., 2011a, ApJ, 741, 56

Dong R., Rafikov R. R., Stone J. M., 2011b, ApJ, 741, 57

Dong R., Zhu Z., Rafikov R. R., Stone J. M., 2015, ApJ, 809, L5

Duffell P. C., MacFadyen A. I., 2012, ApJ, 755, 7

Flaherty K. M. et al., 2017, ApJ, 843, 150

Flaherty K. M., Hughes A. M., Teague R., Simon J. B., Andrews S. M., Liner D. J., 2018, ApJ, 856, 117

Flaherty K. et al., 2020, ApJ, 895, 109

Fung J., Dong R., 2015, ApJ, 815, L21

Garuti A. et al., 2013, A&A, 560, A105

Goldreich P., Tremaine S., 1980, ApJ, 241, 425

Goodman J., Rafikov R. R., 2001, ApJ, 552, 793 (GR01)

Grady C. A. et al., 2013, ApJ, 762, 48

Harris C. et al., 2020, Nature, 585, 357–362

Hunter J. D., 2007, Comput. Sci. Eng., 9, 90

Kevlahan N. K. R., 1997, J. Fluid Mech., 341, 371

Koller J., Li H., Lin D. N. C., 2003, ApJ, 596, L91

Landau L. D., Lifshitz E. M., 1959, Fluid Mechanics. Pergamon Press, Oxford

Li H., Li S., Koller J., Wendroff B. B., Liska R., Orban C. M., Liang E. P. T., Lin D. N. C., 2005, ApJ, 624, 1063

Li H., Lubow S. H., Li S., Lin D. N. C., 2009, ApJ, 690, L52

Lin D. N. C., 2005, ApJ, 624, 1063

Li H., Lubow S. H., Li S., Lin D. N. C., 2009, ApJ, 690, L52

Downloaded from https://academic.oup.com/mnras/article/508/2/2329/6371117 by guest on 15 October 2021
Lin D. N. C., Papaloizou J. C. B., 1993, in Levy E. H., Lunine J. I., eds, Protostars and Planets III. University of Arizona Press, Tucson, p. 749
Lin M.-K., 2015, MNRAS, 448, 3806
Lin M.-K., Papaloizou J. C. B., 2010, MNRAS, 405, 1473
Lin M.-K., Papaloizou J. C. B., 2011, MNRAS, 415, 1445
Lovalcice R. V. E., Li H., Colgate S. A., Nelson A. F., 1999, ApJ, 513, 805
Lynden-Bell D., Kalnajs A. J., 1972, MNRAS, 157, 1
Masset F., 2000, A&AS, 141, 165
Miranda R., Rafikov R. R., 2019a, ApJ, 875, 37
Miranda R., Rafikov R. R., 2019b, ApJ, 878, L9
Miranda R., Rafikov R. R., 2020a, ApJ, 892, 65
Miranda R., Rafikov R. R., 2020b, ApJ, 904, 121
Müller T. W. A., Kley W., Meru F., 2012, A&A, 541, A123
Muto T. et al., 2012, ApJ, 748, L22
Ogilvie G. I., Lubow S. H., 2002, MNRAS, 330, 950
Paardekooper S. J., Papaloizou J. C. B., 2008, A&A, 485, 877
Paardekooper S. J., Papaloizou J. C. B., 2009, MNRAS, 394, 2283
Perez F., Granger B. E., 2007, Comput. Sci. Eng., 9, 21
Petrovich C., Rafikov R. R., 2012, ApJ, 758, 33
Pinte C. et al., 2018, ApJ, 860, L13
Pinte C. et al., 2019, Nat. Astron., 3, 1109
Rafikov R. R., 2002a, ApJ, 569, 997 (RR02)
Rafikov R. R., 2002b, ApJ, 572, 566
Rafikov R. R., 2016, ApJ, 831, 122
Rafikov R. R., 2017, ApJ, 837, 163
Sorathia K. A., Reynolds C. S., Stone J. M., Beckwith K., 2012, ApJ, 749, 189
Stone J. M., Tomida K., White C. J., Felker K. G., 2020, ApJS, 249, 4
Takeuchi T., Miyama S. M., Lin D. N. C., 1996, ApJ, 460, 832
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Ward W. R., 1997, Icarus, 126, 261
Whitham G., 2011, Linear and Nonlinear Waves. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley, Hoboken, NJ
Yu C., Li H., Li S., Lubow S. H., Lin D. N. C., 2010, ApJ, 712, 198
Zhang S., Zhu Z., 2020, MNRAS, 493, 2287
Zhu Z., Dong R., Stone J. M., Rafikov R. R., 2015, ApJ, 813, 88
Ziampras A., Ataiee S., Kley W., Dullemond C. P., Baruteau C., 2020, A&A, 633, A29

APPENDIX A: EFFECT OF THE FORM OF SMOOTHED POTENTIAL

We tested how our results would change if we replace the $\Phi_4^{(s)}$ potential (equation 24) with a second-order smoothed potential of the form

$$\Phi_4^{(s)} = -\frac{GM_p}{(d^2 + r_h^2)^{3/2}}. \quad (A1)$$

which is commonly employed in modelling planet–disc interaction.

The effect on the wave can be seen by examining the wave profile close to the planet. In Fig. A1, we show azimuthal wave profiles for the fiducial disc and $M_p = 0.01M_\odot$ at $R = 1.068R_p$ and $R = 0.935R_p$. The same smoothing parameter of the gravitational potential, $\epsilon = 0.6$ (typical in the literature; Müller et al., 2012), is used for both potentials.

We find that $\Phi_4^{(s)}$ results in a wave amplitude that is about 25 percent smaller than when using $\Phi_4^{(4)}$. This is consistent with the observations of Dong et al. (2011b), who found that the second-order potential creates a shock further from the planet, that is weaker, than $\Phi_4^{(s)}$. Indeed, as non-linear effects directly depend on the wave amplitude it is clear that the change of the wave amplitude due to a different potential must affect the time it takes for characteristic to first cross and a shock to be formed. These differences are expected to reduce as the smoothing parameter $\epsilon$ is decreased.

APPENDIX B: COMPARISON WITH ANOTHER HYDRO CODE

To test the sensitivity of our results to numerical scheme we also run FARGO3D simulations in order to make a direct comparison with ATHENA++. We performed simulations at the highest resolution we could achieve on one GPU, $N_\theta \times N_\phi = 6896 \times 14400$, covering the same domain using the same grid structure, damping zones, and boundary conditions as the ATHENA++ models. For this comparison, we set $\Phi_4 = \Phi_4^{(s)}$ (equation 24), used the fiducial disc parameters ($\eta = 1.5$; $h_0 = 0.05$), and show results for $M_p = 0.25M_\odot$. The behaviour of AMF obtained by both codes is displayed in Fig. B1. In all regions of the disc we find very good agreement between the two codes. The small disagreement at a few per cent level might be due to the lower resolution of FARGO3D runs.

Figure A1. Initial wake profiles at $|R - R_p| = 1.068R_p$, 0.935R_p computed using the second and fourth order approximations for the planetary potential, given by the equations (A1) and (24), respectively. Fiducial disc parameters and $M_p = 0.01M_\odot$ are used.

Figure B1. Comparison of the wave angular momentum fluxes – curves for $F_\phi$ dots for $\Phi_4^{(s)}$, as shown in the legend – computed with ATHENA++ and FARGO3D. Calculation is performed for the same parameters as in Fig. 6.
**APPENDIX C: DERIVATION OF THE VORTENSITY JUMP**

We use equations (9), (11) to express the density jump across the shock in a Keplerian disc in terms of the jump of $\chi$ as

$$\frac{\Delta \Sigma}{\Sigma_0} = \frac{M_p}{M_\infty} \left( \sqrt{2} h_p \frac{\left( \frac{R}{R_p} \right)^{p+1} - 3/2 - 1}{\left( \frac{R}{R_p} \right)^{3/2} - 1} \right)^{1/2} \Delta \chi. \quad (C1)$$

Given the location of the shock $\phi_0(R)$, we can write the derivative along the shock as

$$\frac{d}{dS} = \text{sign}(R - R_p) \frac{dR}{\sqrt{dR^2 + R^2 d\phi^2}} dR = \frac{\text{sign}(R - R_p)}{\sqrt{1 + R^2 \left( \frac{d\phi}{dR} \right)^2}} \frac{d}{dR}. \quad (C2)$$

Substituting these results into equation (39), we obtain

$$\Delta \chi = \frac{c_s}{2^{7/4} \Sigma_0 h_p^{1/2}} \left( \frac{M_p}{M_\infty} \right)^3 \left( \frac{R}{R_p} \right)^{-3/2} \left( \frac{R}{R_p} \right)^{-p+1} (\Delta \chi)^2$$

$$\times \left[ 1 + \frac{M_p}{M_\infty} \left( \frac{R}{R_p} \right)^{-3/2} - 1 \right]^{1/2} \Delta \chi$$

$$\times \frac{\text{sign}(R - R_p)}{\sqrt{1 + R^2 \left( \frac{d\phi}{dR} \right)^2}} \frac{d}{dR} \left[ \left( \frac{R}{R_p} \right)^{-3/2} - 1 \right]^{1/2} \Delta \chi. \quad (C3)$$

The factor in the middle line of the above expression is $M^{-5}$, where $M$ is the normal Mach number of the shock. For $M_p \ll M_\infty$, it is typically of the order of unity. Defining

$$B(R) \equiv \left[ \left( \frac{R}{R_p} \right)^{p+1} - 1 \right]^{1/2}, \quad (C4)$$

$$C(R) \equiv \text{sign}(R - R_p) \left[ 1 + R^2 \left( \frac{d\phi_0}{dR} \right)^2 \right]^{1/2}, \quad (C5)$$

Equation (C3) can be written more compactly as equation (40).

Information about the shape of the shock enters equation (40) through the factor $C(R)$. When using the linear prediction for the location of the primary arm, equation (21), one finds

$$C(R) = \frac{\text{sign}(R - R_p)}{\sqrt{1 + h_p^2 \left( \frac{R}{R_p} \right)^2 \left[ \left( \frac{R}{R_p} \right)^{-3/2} - 1 \right]^2}}. \quad (C6)$$

Alternatively, when accounting for the non-linear correction to the shock position, equation (32), one gets

$$C(R) = \frac{\text{sign}(R - R_p)}{\sqrt{1 + h_p^2 \left( \frac{R}{R_p} \right)^2 \left[ \left( \frac{R}{R_p} \right)^{-3/2} - 1 \right]^2 + \left( \frac{R}{R_p} \right)^{3/2} - 1 \right]^{1/2} \Delta \phi_0 R_p}}. \quad (C7)$$

The non-linear correction acts to decrease the curvature of the shock and reduces the denominator in equation (C7), effectively increasing the derivative along the shock $d/dS$.

This paper has been typeset from a \TeX\,\@X\,\@X file prepared by the author.