

The effect of self-gravity on vortex instabilities in disc–planet interactions

Min-Kai Lin[★] and John C. B. Papaloizou[★]

Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA

Accepted 2011 March 25. Received 2011 March 24; in original form 2010 December 1

ABSTRACT

We study the effect of disc self-gravity on instabilities associated with gaps opened by a giant Saturn mass planet in a protoplanetary disc that lead to the formation of vortices. We also study the non-linear evolution of the vortices when this kind of instability occurs in a self-gravitating disc as well as the potential effect on type III planetary migration due to angular momentum exchange via co-orbital flows.

It is shown analytically and is confirmed through linear calculations that vortex-forming modes with low azimuthal mode number, m , are stabilized by the effect of self-gravity if the background structure is assumed fixed. However, the disc's self-gravity also affects the background gap surface density profile in a way that destabilizes modes with high m . Linear calculations show that the combined effect of self-gravity through its effect on the background structure and its direct effect on the linear modes shifts the most rapidly growing vortex mode to higher m .

Hydrodynamic simulations of gaps opened by a Saturn mass planet show more vortices develop with increasing disc mass and therefore importance of self-gravity. For sufficiently large disc mass the vortex instability is suppressed, consistent with analytical expectations. In this case a new global instability develops instead.

In the non-linear regime, we found that vortex merging is in general increasingly delayed as the disc mass increases and in some cases multiple vortices persist until the end of simulations. For massive discs in which the vortices merge, the post-merger vortex is localized in azimuth and has similar structure to a Kida-like vortex. This is unlike the case without self-gravity where vortices merge to form a larger vortex extended in azimuth.

In order to study the properties of the vortex systems without the influence of the planet, we also performed a series of supplementary simulations of co-orbital Kida-like vortices. We found that self-gravity enables Kida-like vortices to execute horseshoe turns upon encountering each other. As a result, vortex merging is avoided on time-scales where it would occur without self-gravity. Thus we suggest that mutual repulsion of self-gravitating vortices in a rotating flow is responsible for the delayed vortex merging seen in the disc–planet simulations.

The effect of self-gravity on vortex-induced migration in low-viscosity discs is briefly discussed. We found that when self-gravity is included and the disc mass is in the range where vortex-forming instabilities occur, the vortex-induced type III migration of Lin & Papaloizou is delayed. There are also expected to be longer periods of slow migration between the short bursts of rapid migration compared to what occurs in a simulation without self-gravity. However, the extent of induced rapid migration is unchanged and involves flow of vortex material across the gap, independent of whether or not self-gravity is included.

Key words: planets and satellites: formation – planet–disc interactions – protoplanetary discs.

1 INTRODUCTION

Since the first detection of an extrasolar giant planet orbiting a main-sequence star in a 4-d orbit by Mayor & Queloz (1995), there have been many observations of hot Jupiters (Marcy et al. 2005; Udry & Santos 2007). One possible explanation for their close proximity

[★]E-mail: mkl23@cam.ac.uk (MKL); J.C.B.Papaloizou@damtp.cam.ac.uk (JCBP)

to their host stars is orbital migration due to interaction with the gaseous protoplanetary disc, from their site of formation to their current observed location (Lin & Papaloizou 1986).

Planetary migration was first studied by Goldreich & Tremaine (1980), and since then many analytical and numerical studies of disc–planet interactions have been undertaken (see Papaloizou et al. 2007, for a review). However, most of these adopt a low-mass disc model for which it may be assumed that self-gravity may be neglected. This is even the case when rapid type III migration that requires a massive disc is considered. It is the purpose of this paper to investigate the effect of the disc self-gravity on such disc–planet interactions and in particular to consider its effect on the stability of the disc to vortex formation, a phenomenon that has been found to occur in discs with very low viscosity.

Astrophysical discs may be dynamically unstable for a variety of reasons. One example is instability associated with steep surface density gradients or vortensity¹ extrema (Papaloizou & Pringle 1985, 1987; Papaloizou & Lin 1989; Papaloizou & Savonije 1991; Lovelace et al. 1999; Li et al. 2000) that can lead to vortex formation in non-linear regime (Li et al. 2001). The disc vortensity profile necessary for such instabilities to occur may be induced by disc–planet interaction. For sufficiently large planetary masses and/or sufficiently low disc viscosity, the tidal perturbation on the gaseous disc leads to a dip or gap in the surface density (Lin & Papaloizou 1986). There have been a series of studies of the vortex instability associated with planetary gaps (Koller, Li & Lin 2003; Li et al. 2005; de Val-Borro et al. 2007) and its consequences for planetary migration (Ou et al. 2007; Li et al. 2009; Lin & Papaloizou 2010; Yu et al. 2010). However, the effect of disc self-gravity is still unclear.

In the context of disc–planet interactions, disc self-gravity is often neglected. It has thus far only been studied in a limited number of works (Nelson & Benz 2003a,b; Baruteau & Masset 2008; Zhang et al. 2008). Li et al. (2009) included self-gravity in their study of type I migration in low-viscosity discs, where the vortex instability develops for α viscosity parameters $\leq 10^{-5}$, but the specific effect of self-gravity on the instability was not discussed analytically or explored numerically. In this paper we examine the effect of self-gravity on the vortex instability associated with edges of gaps opened by a Saturn mass planet and its effect on the subsequent non-linear evolution of vortices together with a first calculation of the subsequent effects on orbital migration.

This paper is organized as follows. In Section 2 we present the governing equations and models for our disc–planet systems. We discuss the linearized stability problem in Section 3. We show that vortex-forming modes with long azimuthal wavelength associated with vortensity minima at gap edges, induced by an embedded planet, are stabilized by self-gravity when the background model is held fixed. In Section 4 we perform linear mode calculations for disc models of varying mass. We confirm the analytic discussion but show in addition that changes to the gap structure that occur with increasing disc mass are destabilizing, an effect that is stronger for modes with short wavelength in the azimuthal direction. All of these effects cause the most unstable modes to shift to shorter azimuthal wavelength as the disc mass increases. However, in practice, these modes are eventually dominated by long azimuthal wavelength, global edge modes, which are not associated with localized vortex formation and which are studied in Lin & Papaloizou (2011). Edge modes dominate our models when the disc-to-star mass ratio is $\gtrsim 0.047$, corresponding to a Keplerian Toomre stability parameter

$\lesssim 2$ at the outer disc boundary (see Section 2.1 and Appendix A). As this paper is concerned with vortex formation, we focus on models with disc mass below this threshold.

We go on to present non-linear hydrodynamic simulations in Section 5. In accord with expectations from linear theory, the instability produces more vortices of smaller scale as the disc mass increases. The vortices are found to survive for much longer against merging than in the non-self-gravitating (NSG) limit and themselves become SG with strong over surface densities. This is found to enable pairs of vortices to undergo co-orbital dynamics with horseshoe trajectories which aid to inhibit merging. In Section 6 we perform simulations of interacting Kida-like vortices without an embedded planet in order to clarify results obtained for vortices at gap edges. These show that the co-orbital dynamics occurs independently of the embedded planet. In Section 7 we give a preliminary investigation of the effects of vortex formation in low-viscosity discs on orbital migration when self-gravity is significant but not strong enough to prevent effective vortex formation. As in the NSG case, episodes of rapid migration occur as vortices are scattered by the planet (Lin & Papaloizou 2010), but longer time intervals between them are expected on account of stabilization by self-gravity and slowed vortex merging. Finally we summarize and conclude in Section 8.

2 DISC MODEL AND BASIC EQUATIONS

We consider a gaseous disc of mass M_d orbiting a central star of mass M_* . We assume the disc to be razor thin and thus work in the two-dimensional flat disc approximation. We adopt cylindrical polar coordinates (r, φ) centred on the star and defined in the plane of the disc. The reference frame is non-rotating.

The governing hydrodynamic equations are the continuity and momentum equations

$$\frac{D\Sigma}{Dt} = -\Sigma \nabla \cdot \mathbf{u}, \quad (1)$$

$$\frac{D\mathbf{u}}{Dt} = -\frac{1}{\Sigma} \nabla p - \nabla \Phi + \frac{\mathbf{f}}{\Sigma}, \quad (2)$$

where Σ is the surface density, \mathbf{u} is the velocity field and p is the vertically integrated pressure. Numerical calculations adopt a locally isothermal equation of state, $p = c_s^2 \Sigma$ where $c_s = H\Omega_k$ is the local sound speed and $\Omega_k = \sqrt{GM_*/r^3} \equiv 2\pi/P$. Here $H = hr$ is the putative disc semithickness. We fix the constant $h = 0.05$. The viscous force per unit area is \mathbf{f} being characterized by a uniform kinematic viscosity ν . The gravitational potential Φ includes the indirect potential Φ_i , the potential due to the central star, $\Phi_* = -GM_*/r$, the potential due to the disc Φ_d when self-gravity is included and the potential due to an embedded planet, Φ_p , for disc–planet systems. Details of \mathbf{f} are given in Masset (2002). The potentials due to the planet and the disc are

$$\Phi_p = -\frac{GM_p}{\sqrt{r^2 + r_p^2 - 2rr_p \cos(\varphi - \varphi_p) + \epsilon_p^2}} \quad (3)$$

$$\Phi_d = -\int_{\mathcal{D}} \frac{G\Sigma(r', \varphi')}{\sqrt{r^2 + r'^2 - 2rr' \cos(\varphi - \varphi') + \epsilon_g^2}} r' dr' d\varphi', \quad (4)$$

where the integral is taken over the domain of the disc \mathcal{D} . The planet has mass M_p , position (r_p, φ_p) and its potential is softened by use of the softening length $\epsilon_p = 0.6H(r_p)$. Similarly the disc potential is softened by use of the softening length $\epsilon_g = 0.3H(r')$. The indirect potential takes account of the forces on the central star, which is at

¹ The term vortensity is used for the ratio of vorticity to surface density.

the origin of a non-inertial frame, due to the disc and planet. It is given by

$$\Phi_1 = r \int_{\mathcal{D}} \frac{G\Sigma(r', \varphi')}{r'^2} \cos(\varphi - \varphi') r' dr' d\varphi' + \frac{GM_p}{r_p^2} r \cos(\varphi - \varphi_p). \quad (5)$$

We adopt units such that $G = M_* = 1$. The Keplerian orbital period at $r = 1$ is then $P(1) = 2\pi$. As in our previous work, for disc–planet simulations we use a uniform viscosity $\nu = 10^{-6}$ in code units. This is an order of magnitude smaller than the typically adopted value of $\nu = 10^{-5}$ that suppresses the vortex instability (de Val-Borro et al. 2007; Lin & Papaloizou 2010).

2.1 Disc–planet model

Most of our discussion will be applied to disc–planet systems, so we describe these models here. Supplementary simulations in Section 6 employ a different set-up from that described below.

The disc occupies $r = [r_i, r_o] = [1, 10]$. The initial surface density profile is modified from Armitage & Hansen (1999):

$$\Sigma(r) = \Sigma_0 r^{-3/2} \left(1 - \sqrt{\frac{r_1}{r + H_1}} \right), \quad (6)$$

where $H_1 = H(r_1)$ is introduced to prevent a singular pressure force at the inner boundary because $\Sigma \rightarrow 0$ there. Σ_0 is chosen via the parameter $Q_o \equiv Q_{\text{Kep}}(r_o)$, where

$$Q_{\text{Kep}}(r) = \frac{c_s \kappa}{\pi G \Sigma} = \frac{h M_*}{\pi r^2 \Sigma(r)} \quad (7)$$

is the Toomre Q parameter for a thin Keplerian disc with a locally isothermal equation of state and $\kappa^2 = 2\Omega r^{-1} d(r^2 \Omega)/dr$ is the square of the epicycle frequency, with Ω being the disc angular velocity. We use Q_o to label disc models and also define $Q_p \equiv Q_{\text{Kep}}(r_p)$. We note that specifying Q_o also determines the disc-to-star mass ratio, M_d/M_* . The conversion between Q_o , Q_p and M_d/M_* for the models used in disc–planet interactions is given by Table A1 in Appendix A.

The disc is initialized with the azimuthal velocity required by hydrostatic equilibrium. The initial radial velocity is $u_r = 3\nu/r$, corresponding to the initial radial velocity profile of a Keplerian disc with uniform kinematic viscosity and surface density $\propto r^{-3/2}$. Note that $|u_r| \ll |u_\phi|$.

In most of this work we focus on stability of the gap induced by an embedded planet, accordingly we fix the planet on a circular orbit of radius $r_p = 5$. We quote time in units $P_0 = 2\pi/\Omega_k(r_p)$. The planet is introduced at $t = 25P_0$ with azimuthal velocity that takes account the contribution from the gravitational force due to the disc. The planet potential is ramped up over a time period of $10P_0$.

3 LINEAR MODES IN DISCS WITH STRUCTURED SURFACE DENSITY

In this section and the next we study linear disturbances associated with internal surface density depressions in a disc in which self-gravity is not neglected. We will consider dips/gaps self-consistently opened by a giant planet. To simplify the discussion, we ignore viscosity, indirect potentials and the planet potential. The planet's role is then only to set up the basic state, assumed to be axisymmetric and defined by $\Sigma(r)$, $u_\phi(r)$ and $u_r = 0$. We begin with the linearized equations with a local isothermal equation of state.

We consider perturbations to the disc state variables with azimuthal and time dependence through a factor $\exp(i\sigma t + m\varphi)$ which is taken as real. Here $\sigma = \sigma_R - i\gamma$ is a complex frequency with σ_R and $-\gamma$ being the real and imaginary parts, respectively. The azimuthal mode number m is taken to be a positive integer. Denoting perturbations by a prime, the linearized equations of motion read

$$u'_r = -\frac{1}{D} \left[i\bar{\sigma} \left(c_s^2 \frac{dW}{dr} + \frac{d\Phi'}{dr} \right) + \frac{2im\Omega}{r} (c_s^2 W + \Phi') \right] \quad (8)$$

$$u'_\phi = \frac{1}{D} \left[\frac{\kappa^2}{2\Omega} \left(c_s^2 \frac{dW}{dr} + \frac{d\Phi'}{dr} \right) + \frac{m\bar{\sigma}}{r} (c_s^2 W + \Phi') \right], \quad (9)$$

where $D = \kappa^2 - \bar{\sigma}^2$, $\bar{\sigma} \equiv \sigma + m\Omega(r)$ is the Doppler shifted frequency, $W = \Sigma'/\Sigma$ is the relative surface density perturbation and Φ' is the disc gravitational potential perturbation which is given by the Poisson integral:

$$\Phi' = -G \int_{r_i}^{r_o} K_m(r, \xi) \Sigma(\xi) W(\xi) \xi d\xi, \quad \text{where} \quad (10)$$

$$K_m(r, \xi) = \int_0^{2\pi} \frac{\cos(m\varphi) d\varphi}{\sqrt{r^2 + \xi^2 - 2r\xi \cos(\varphi) + \epsilon_\xi^2(\xi)}}. \quad (11)$$

Using the linearized equations of motion to eliminate the velocity component perturbations from the linearized continuity equation,

$$i\bar{\sigma} W = -\frac{1}{r\Sigma} \frac{d}{dr} (r\Sigma u'_r) - \frac{im}{r} u'_\phi, \quad (12)$$

yields the governing equation:

$$\frac{d}{dr} \left[\frac{r\Sigma}{D} \left(c_s^2 \frac{dW}{dr} + \frac{d\Phi'}{dr} \right) \right] + \left[\frac{2m}{\bar{\sigma}} \left(\frac{\Sigma\Omega}{D} \right) \frac{dc_s^2}{dr} - r\Sigma \right] W + \left[\frac{2m}{\bar{\sigma}} \frac{d}{dr} \left(\frac{\Sigma\Omega}{D} \right) - \frac{m^2\Sigma}{rD} \right] (c_s^2 W + \Phi') = 0. \quad (13)$$

3.1 Modes leading to vortex formation: the effect of self-gravity

We shall be concerned with the vortex-forming instabilities commonly observed in disc–planet interactions in discs with low viscosity (Koller et al. 2003; Li et al. 2005, 2009; Lin & Papaloizou 2010). The unstable modes are found to be localized to gap edges and associated with minima in the vortensity (Papaloizou & Lin 1989; Papaloizou & Savonije 1991; Lovelace et al. 1999).

Here, we investigate the effect of self-gravity on these modes and show somewhat paradoxically, in view of the fact that increasing self-gravity in general destabilizes discs, that it tends to stabilize the modes. To do this we consider a small change to the strength of self-gravity and apply perturbation theory to determine the consequence of the change for these modes.

Non-linear simulations show more vortices develop as self-gravity is increased. This is in fact consistent with the stabilization effect demonstrated below, which is effective for small m . In other words, stabilization by self-gravity discourages low m vortex modes, supporting the observation that only higher m modes develop in simulations.

To make the analysis tractable, further simplifications are made. We take the fluid to be such that either c_s^2 is constant (strictly isothermal) or more generally the equation of state is barotropic such that $p = p(\Sigma)$. We do not expect a significant difference between adopting a strictly isothermal, or barotropic equation of state, and

the fixed c_s profile as used in simulations. This is because the modes of interest are driven by local features and disturbances are localized, whereas $c_s(r)$ for a locally isothermal equation of state varies on a global scale. For mathematical convenience we also take a softening prescription such that $K_m(r, \xi) = K_m(\xi, r)$ is symmetric, e.g. $\epsilon_g = \text{constant}$. Although convenient mathematically, this is not expected to lead to significant changes for the same reason as that given above.

Introducing the variable S

$$S \equiv c_s^2 W + \Phi' = c_s^2 W - G \int_{r_i}^{r_o} K_m(r, \xi) \xi \Sigma(\xi) W(\xi) d\xi, \quad (14)$$

the governing equation derived from equation (13) simply by taking c_s to be constant is

$$\begin{aligned} r \Sigma W &= \frac{d}{dr} \left[\frac{r \Sigma}{D} \left(\frac{dS}{dr} \right) \right] + \left\{ \frac{m}{\bar{\sigma}} \frac{d}{dr} \left[\frac{1}{\eta(1 - \bar{v}^2)} \right] - \frac{m^2 \Sigma}{rD} \right\} S \\ &\equiv rL(S), \end{aligned} \quad (15)$$

where we have used the expression for vortensity $\eta = \kappa^2/2\Omega\Sigma$ and $\bar{v} \equiv \bar{\sigma}/\kappa$. We remark that in fact these equations also hold for a general barotropic equation of state.

3.2 The limit of negligible self-gravity

When self-gravity is negligible, we may set $G = 0$ in (14), from which we obtain $W = S/c_s^2$. Substituting this into (15) gives the second-order ordinary differential equation for S that governs stability in the limit of zero self-gravity in the form

$$\frac{r \Sigma S}{c_s^2} = \frac{d}{dr} \left[\frac{r \Sigma}{D} \left(\frac{dS}{dr} \right) \right] + \left\{ \frac{m}{\bar{\sigma}} \frac{d}{dr} \left[\frac{1}{\eta(1 - \bar{v}^2)} \right] - \frac{m^2 \Sigma}{rD} \right\} S. \quad (16)$$

This equation was studied by Lin & Papaloizou (2010) for gap profiles of the type we consider. They found that depending on parameters such as m , there were neutral modes with corotation radius, r_c , where $\bar{\sigma}(r_c) = 0$ coincident with a vortensity minimum. This could be associated with either the inner or outer gap edge. Small variations of the disc parameters could then lead to an instability associated with vortex formation. When unstable these modes retain r_c to be close to the vortensity minimum with which they can be considered to be associated.

3.3 Localized low m modes

When m is small the neutral modes described above are localized around corotation and therefore insensitive to distant boundary conditions. For these modes, we may neglect \bar{v} and set $D = \kappa^2$ in (16) above to get the simpler equation:

$$\frac{r \Sigma S}{c_s^2} = \frac{d}{dr} \left[\frac{r \Sigma}{\kappa^2} \left(\frac{dS}{dr} \right) \right] + \left\{ \frac{m}{\bar{\sigma}} \frac{d}{dr} \left[\frac{1}{\eta} \right] - \frac{m^2 \Sigma}{r \kappa^2} \right\} S. \quad (17)$$

Localization occurs because the solutions of (17) can be seen to decay exponentially away from the corotation region, where there are large background gradients, on a length-scale comparable to H . However, for (17) to be a good approximation, we require $|\bar{v}^2| \ll 1$ in the region of localization which is also comparable to H in extent. This in turn requires $H \ll 2r/(3m)$, a condition which is satisfied for low m . For larger m the mode becomes de-localized with the excitation of density waves that propagate into the extended disc (Lin & Papaloizou 2010). In this case the boundary conditions can play a role. The analysis below assumes localization so that boundary conditions do not play a role. It therefore only applies for low m .

3.4 Evaluating the effect of small changes to low m modes using perturbation theory

In order to investigate the effect of self-gravity, we return to equations (14) and (15) and note that strengthening self-gravity by scaling up the surface density is equivalent to increasing G , provided that the background form remains fixed. Thus although we increase the effective disc gravity, we shall assume the background disc model remains unchanged. In fact, in our case the background model is structured by a perturbing planet and so its response to changing the disc gravity is non-trivial to evaluate analytically. Calculations presented below in Section 4.7 show that changes to the background surface density profile induced by incorporating the disc gravity tend to act to make the vortex modes we consider more unstable. However, the direct effect of self-gravity on the linear response considered below turns out to be more important for localized modes with low m , and acts to stabilize them.

Accordingly we assume a solution S corresponding to a neutral mode with corotation radius, r_c , located at a vortensity minimum, such that $d\eta/dr|_{r=r_c} = 0$. As the associated σ is real, for this value, the operator L is real and regular everywhere. We then perturb this solution by altering the strength of self-gravity via

$$G \rightarrow G + \delta G,$$

so that $\delta G > 0$ corresponds to increasing the importance of self-gravity and vice versa (note that the initial value, G , could be zero). This induces perturbations

$$S \rightarrow S + \delta S, \quad \Sigma' \rightarrow \Sigma' + \delta \Sigma', \quad \sigma \rightarrow \sigma + \delta \sigma,$$

$$L \rightarrow L + \delta L,$$

with

$$\delta \sigma = \delta \sigma_R - i\gamma$$

$$\delta L = \frac{\partial L}{\partial \sigma} \delta \sigma,$$

where $\delta \sigma_R$ and γ are real.

Noting that δ denotes a small change and γ is small, we linearize in terms of these quantities about the assumed original neutral mode and determine γ . The governing equations lead to

$$L(\delta S) + \delta L(S) = \delta \Sigma', \quad (18)$$

$$\begin{aligned} \delta S &= c_s^2 \frac{\delta \Sigma'}{\Sigma} - G \int_{r_i}^{r_o} K_m(r, \xi) \xi \delta \Sigma'(\xi) d\xi \\ &\quad - \delta G \int_{r_i}^{r_o} K_m(r, \xi) \xi \Sigma'(\xi) d\xi. \end{aligned} \quad (19)$$

Next, we define the inner product between two functions $U(r)$ and $V(r)$ as

$$\langle U, V \rangle = \int_{r_i}^{r_o} r U^*(r) V(r) dr, \quad (20)$$

Then, assuming localized functions corresponding to localized modes so that we can assume boundary values vanish when integrating by parts, we have

$$\langle U, L(V) \rangle = \langle L(U), V \rangle, \quad (21)$$

which used the fact that L corresponding to the neutral mode is real, which makes L self-adjoint. Equations (18)–(19) can be manipulated to yield

$$\begin{aligned} \langle S, \delta L(S) \rangle &= \delta G \int_{r_i}^{r_o} \int_{r_i}^{r_o} r \xi K_m(r, \xi) \Sigma'^*(\xi) \Sigma'(r) d\xi dr. \\ &\equiv \delta G E, \end{aligned} \quad (22)$$

where we have used the fact that the kernel K_m is symmetric. Note that $E > 0$ and is proportional to the magnitude of the gravitational energy of the mode. Following Papaloizou & Lin (1989), we separate out the contribution to the perturbed linear operator δL that is proportional to the vortensity gradient and is potentially singular at corotation and write

$$\langle S, \delta L(S) \rangle \equiv \langle S, \delta L_1(S) \rangle + \langle S, \delta L_2(S) \rangle, \quad (23)$$

where $\delta L_2(S)$ contains the potentially singular contribution, and we have

$$\langle S, \delta L_2(S) \rangle = -m\delta\sigma \left[\mathcal{P} \left(\int_{r_1}^{r_o} \frac{g(r)}{\bar{\sigma}} dr \right) + \int_{r_1}^{r_o} i\pi\delta(\bar{\sigma})g(r) dr \right], \quad (24)$$

where

$$g(r) \equiv \frac{|S|^2}{\bar{\sigma}} \frac{d}{dr} \left(\frac{1}{\eta} \right). \quad (25)$$

Note that the usual Landau prescription for positive γ has been used to deal with the pole at corotation and that \mathcal{P} outside the above integral indicates that the principal value is to be taken. Note too that δL_1 accounts for the remainder of δL and is not singular at corotation. The contribution from δL_2 is the only one that can lead to an imaginary contribution in the limit $\delta\sigma \rightarrow 0$ because of the pole at corotation. In this regard, we remark that one does not get such contributions arising from Lindblad resonances where $D = 0$. This is because, as is well known, these do not constitute effective singularities in a gaseous disc (e.g. Papaloizou & Savonije 1991).

Recalling that γ is small, equations (22)–(24) can be combined to give

$$\begin{aligned} \delta GE &= \delta\sigma \left[\frac{\partial \langle S, L_1(S) \rangle}{\partial \sigma} - m\mathcal{P} \left(\int_{r_1}^{r_o} \frac{g(r)}{\bar{\sigma}} dr \right) \right. \\ &\quad \left. - im\pi \int_{r_1}^{r_o} \delta(\bar{\sigma})g(r) dr \right] \\ &\equiv (\mathcal{A} - im\chi)\delta\sigma, \end{aligned} \quad (26)$$

where \mathcal{A} is the contribution from the L_1 term plus the principle value integral and

$$\chi \equiv \frac{\pi|S|^2}{m\Omega'|m\Omega'|} \frac{d^2}{dr^2} \left(\frac{1}{\eta} \right) \Big|_{r=r_c}. \quad (27)$$

In the limit $\gamma \rightarrow 0_+$, \mathcal{A} is real so we have

$$\gamma = -\frac{m\chi}{\mathcal{A}^2 + m^2\chi^2} \delta GE. \quad (28)$$

As we have remarked above, vortex-forming modes are associated with vortensity minima or at maxima of η^{-1} . Consider a marginally stable mode with corotation at $\max(\eta^{-1})$. Typical rotation profiles have $\Omega' < 0$, which means $\chi > 0$ for this mode. For consistency with the assumption of $\gamma > 0$ in deriving equation (28), we require $\delta G < 0$ since $E > 0$. Accordingly in order to destabilize this mode, the strength of self-gravity needs to be reduced.

This leads to the conclusion that increasing self-gravity *stabilizes* low m modes with corotation at a vortensity *minimum* (as has been borne out by our linear and non-linear calculations presented below); while increasing self-gravity would *destabilize* modes which had corotation at a vortensity *maximum*. This suggests that for sufficiently strong self-gravity, modes associated with vortensity maxima should be favoured. This has been found to be the case, but they are not modes leading to vortex formation (see Lin & Papaloizou 2011).

Note $\chi \propto 1/m^2$, suggesting that γ decreases for large m , so the stabilization/destabilization effect of self-gravity diminishes for increasing azimuthal wavenumber. This is only speculative because there are implicit dependencies on m through terms in the integrals and through the original eigenfunctions S . Nevertheless, a weakening effect of self-gravity through the potential perturbation is anticipated because increasing m decreases the magnitude of the Poisson kernel K_m .

3.5 Association of localized low m normal modes with vortensity minima for the strength of self-gravity below a threshold

We can also show that corotation radii for localized neutral modes with low m must be at vortensity minima unless self-gravity (or an appropriate mean value of Q) is above (below) a threshold level. Let us consider a disc with localized steep surface density gradients and a non-axisymmetric disturbance with corotation radius in this region. Multiplying equation (15) by S^* and integrating over the disc, assuming most of the contribution is from near corotation, as is expected for low m modes (see Section 3.3), so that the term that is potentially singular at corotation and is proportional to the vortensity gradient is dominant on the right-hand side, we have the balance

$$\begin{aligned} &\int_{r_1}^{r_o} rc_s^2 \frac{|\Sigma'|^2}{\Sigma} dr - G \int_{r_1}^{r_o} K_m(r, \xi) r \xi \Sigma'^*(\xi) \Sigma'(r) d\xi dr \\ &\simeq \int_{r_1}^{r_o} \frac{m|S|^2}{\bar{\sigma}} \frac{d}{dr} \left(\frac{1}{\eta} \right) dr. \end{aligned} \quad (29)$$

If the left-hand side can be shown to always be positive then corotation for a localized neutral mode must lie at a vortensity minimum or $\max(\eta^{-1})$. This holds if the maximum possible value of Λ for any Σ' is less than unity, where

$$\Lambda = \frac{G \int_{r_1}^{r_o} K_m(r, \xi) r \xi \Sigma'^*(\xi) \Sigma'(r) d\xi dr}{\int_{r_1}^{r_o} rc_s^2 \frac{|\Sigma'|^2}{\Sigma} dr}. \quad (30)$$

Now from the Cauchy–Schwartz inequality it follows that

$$\Lambda \leq G \sqrt{\int_{r_1}^{r_o} K_m^2(r, \xi) \frac{\Sigma^*(r)\Sigma(\xi)}{c_s^2(r)c_s^2(\xi)} r \xi d\xi dr}. \quad (31)$$

Thus if the right-hand side of the above is less than unity, corotation of a neutral mode localized at a vortensity extremum must be localized at a vortensity minimum. This condition requires that the strength of self-gravity be below a threshold, and this implies a lower bound on an appropriate mean Q value. The fact that this fails for sufficiently strong self-gravity is consistent with the discussion above that led to the conclusion that increasing self-gravity tends to stabilize the vortex-forming instability.

Indeed when self-gravity increases further, instability is transferred to modes with corotation associated with vortensity maxima. These are different in character to vortex-forming modes being more global and are referred to as edge modes, and they are studied in Lin & Papaloizou (2011).

4 LINEAR CALCULATIONS

We now present numerical solutions to equation (13) with a local isothermal equation of state for consistency with numerical simulations used to set up the basic state. We regard the governing equation

as the requirement that an operator \mathcal{L} acting on W should be zero, thus

$$\mathcal{L}(W) = 0. \quad (32)$$

One form of \mathcal{L} is

$$\begin{aligned} \mathcal{L} = & r c_s^2 \frac{d^2}{dr^2} + r \mathcal{I}_2 + \left[\frac{1}{r} + \frac{\Sigma'}{\Sigma} - \frac{D'}{D} \right] \left(r c_s^2 \frac{d}{dr} + r \mathcal{I}_1 \right) \\ & + \left\{ \frac{2m}{\bar{\sigma}} \left[\frac{\Sigma' \Omega}{\Sigma} + \Omega' - \frac{\Omega D'}{D} \right] - \frac{m^2}{r} \right\} (c_s^2 + \mathcal{I}_0) \\ & + r c_s^2 \frac{d}{dr} + \left[\frac{2m \Omega c_s^2}{\bar{\sigma}} - r D \right]. \end{aligned} \quad (33)$$

It is understood that primes in equation (33) denote d/dr . The integro-differential operators \mathcal{I}_n are such that

$$\mathcal{I}_n(W) = \frac{d^n \Phi'}{dr^n}. \quad (34)$$

First and second derivatives of the perturbed potential are performed by replacing $K_m(r, \xi)$ by $\partial K_m / \partial r$ and $\partial^2 K_m / \partial r^2$ in the Poisson integral, respectively.

4.1 Quadratic approximation

Vortex modes are localized about corotation at a vortensity minimum. They extend over a region characterized by small $|\bar{\sigma}|$ in the neighbourhood of r_c and are expected to be insensitive to boundary conditions. If we multiply equation (33) by $\bar{\sigma} D$ and expand the resulting equation in powers of $\bar{\sigma}$ and neglect terms proportional to $\bar{\sigma}^3$ and higher powers, we can derive an equation of the form

$$(\sigma^2 \mathcal{L}_2 + \sigma \mathcal{L}_1 + \mathcal{L}_0)W = 0, \quad (35)$$

where the operators \mathcal{L}_i are real and only depend on the basic state. The eigenfrequency now appears explicitly. We call this the quadratic approximation. Although the above procedure is not expected to be valid in general, it should be valid for modes localized about their corotation radii with the scale of localization being much less than r_c itself. Modes can be found from equation (35) using standard numerical methods. The eigenvalues are then used as starting values in order to obtain an iterative solution for the eigenvalues for the full equation (32). The concept of the vortex-forming mode as a localized mode *is confirmed* if the final solution of (32) is not significantly changed from the solution of (35). Note that since the operators \mathcal{L}_i are real, eigenfrequencies are either real or occur in complex conjugate pairs.

4.2 Distinction between including and excluding self-gravity

Since we are concerned with effects of self-gravity, we need to clearly define self-gravitating (SG) and non-self-gravitating (NSG) cases. The background state on which we perform linear stability analysis was obtained from non-linear hydrodynamic simulations (see Section 4.4), which can be run with self-gravity (SGBG) or without self-gravity (NSGBG). In the linear calculations, self-gravity can be disabled by setting $\Phi' = 0$. Therefore, we have two disturbance types: an SG response (SGRSP) and an NSG response (NSGRSP).

The analytical discussion in Section 3.1 applies to the effect of self-gravity through the linear response, assuming the form of the background remains unchanged. In the context of hydrodynamic simulations, including or neglecting self-gravity simply means whether or not the disc gravitational potential is included. If it is included, the background state set up by simulations also

depends on self-gravity. Hydrodynamic simulations therefore correspond either to the combination SGBG+SGRSP or to the combination NSGBG+NSGRSP. We call these fully SG and fully NSG cases. The advantage of performing linear calculations is that we can distinguish between the effects of self-gravity arising through its effects on the background state and the effects resulting from its influence on the linear response.

4.3 Numerical approach

We discretized the linear operators on a grid that divides the radial range $[r_i, r_o]$ into typically $N_r = 385$ equally spaced grid points at which W is evaluated as $W_j, j = 1, 2, \dots, N_r$. The governing equation thus becomes a system of N_r simultaneous equations for the W_j which determine an eigenvalue problem for σ . The system of equations incorporates the boundary conditions. For simplicity we impose $dW/dr = 0$ at boundaries in discretized form. We remark that it is known from other linear calculations and simulations that vortex-forming modes are localized and insensitive to boundary conditions (de Val-Borro et al. 2007; Lin & Papaloizou 2010). Tests have shown that the boundary condition does not influence the essential effect of self-gravity on localized modes.

The discretized quadratic approximation obtained from equation (35) leads to a quadratic eigenvalue problem that is equivalent to a $2N_r \times 2N_r$ standard linear matrix eigenvalue problem. This is solved by methods which give all eigenvalues σ . The most unstable eigenvalue is then used as a trial solution for the full system which yields a discriminant that is then solved iteratively for the actual eigenvalue, σ , using the Newton–Raphson method. We restrict attention to modes with eigenfrequencies such that $|\sigma_R/m\Omega_e + 1| < 0.1$ and $|\gamma| < \Omega_e$, where Ω_e is the rotation frequency of the vortensity minimum at the outer gap edge.

4.4 Background state

The background states used for linear calculations were set up by running disc–planet simulations for the models described in Section 2.1. Numerical details will be described in Section 5. We allow the planet to open a gap, then take azimuthal averages to obtain one-dimensional profiles.

Fig. 1 shows the gap profile, in terms of the inverse vortensity $1/\eta$, opened up by a Saturn mass planet in the $Q_0 = 4$ disc. The

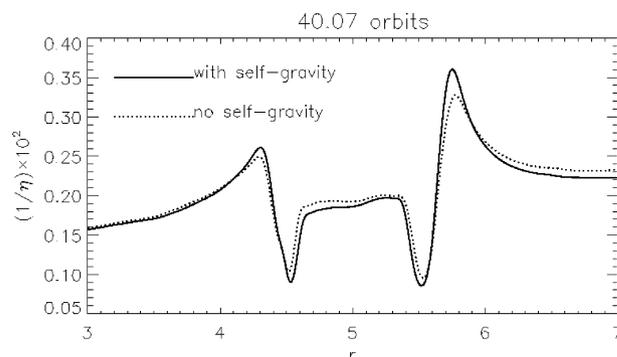


Figure 1. Gap profiles opened by a Saturn mass planet in an initial disc model with $Q_0 = 4$. The inverse vortensity, η^{-1} , obtained with (solid) and without (dotted) self-gravity included is shown. Profiles were obtained from azimuthal averages taken from disc–planet simulation outputs. The planet is at the fixed location $r = 5$.

formation of vortensity peaks is due to the generation of vortensity as material passes through shocks induced by the giant planet which extend into its co-orbital region (Lin & Papaloizou 2010). We shall see that in hydrodynamic simulations vortices predominantly develop at the outer gap edge, so we focus on modes associated with the outer vortensity extrema. The outer vortensity maximum or $\min(\eta^{-1})$ is located at $r \simeq 5.5$ while the outer vortensity minimum or $\max(\eta^{-1})$ is located at $r \simeq 5.75$. These extrema are separated by about $0.89H$. As we increase the strength of self-gravity, the discussion in Sections 3.4 and 3.5 indicates that eventually vortex modes associated with vortensity minima are suppressed and modes associated with vortensity maxima become favoured instead. This is consistent with self-gravity stabilizing low m vortex modes through the linear response.

The gap profiles set up with and without self-gravity are similar. When self-gravity is included, the peaks and troughs have slightly larger amplitudes due to increased effective planet mass (see Section 5.3).

4.5 Solution in the quadratic approximation

We first present solutions for linear modes for the $Q_0 = 4$ disc in the quadratic approximation. Fig. 2 compares the growth rates, γ , and the eigenfunctions, W , obtained for $m = 5$ for the fully SG case and the fully NSG case. Growth rates are such that the most unstable mode shifts to higher m when self-gravity is included. Without self-gravity, the dominant mode has $m = 3$, whereas with self-gravity the dominant mode has $m = 6-7$. The combination of self-gravity acting through the background and the linear response stabilizes modes with $m \leq 5$ and destabilizes modes with $m \geq 6$. Thus higher m vortex modes are enabled by self-gravity.

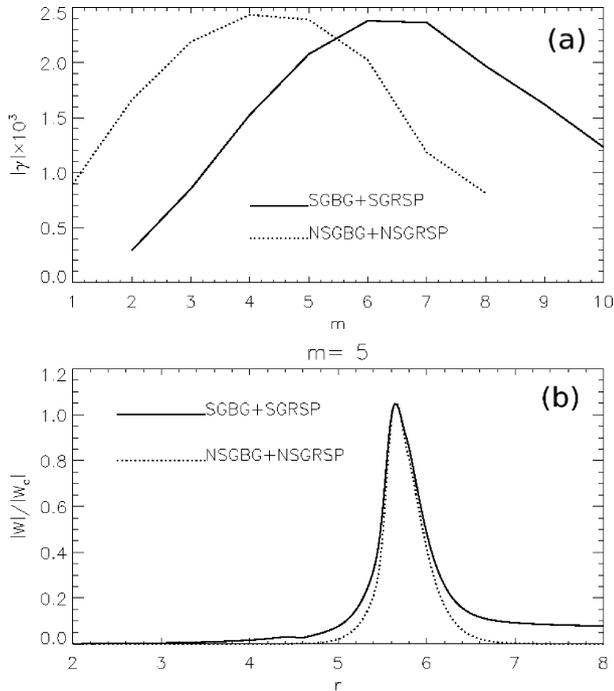


Figure 2. Growth rates (a) obtained from the quadratic approximation to the governing equation and the modulus of the $m = 5$ eigenfunction (b). Here $W_c = W(r = 5.7)$. The solid (dotted) lines correspond to the case with (without) self-gravity in both the background state and response. The initial disc model had $Q_0 = 4$.

The eigenfunctions W with and without self-gravity are similar, and are for the most part localized about corotation. This suggests the nature of the instability is unchanged in this case. However, the eigenfunction amplitude in $r > 6$ and $r < 5.2$ is larger when self-gravity is included. Note too that the amplitude is larger for $r > 6$ than for $r < 5.2$. This is not unexpected since the outer disc has smaller Toomre Q values it is expected to be more responsive.

It is important to remember that the quadratic approximation assumes modes are localized about corotation. The global nature of some disturbances may invalidate this approximation. Increasing m eventually quenches modes in the NSG case, thus we did not find modes for $m \geq 9$ that fit the description of being localized modes. For the full governing equation (13), high m modes are expected to have increasing wave-like behaviour and be less focused around corotation, contradicting the quadratic approximation, and so it is not surprising that they are not found here. Similarly, we did not find a localized $m = 1$ vortex mode in the SG case because it had been stabilized by the inclusion of self-gravity, according to earlier analysis. However, there exist other types of low m mode not captured by the quadratic approximation. Modes with extreme values of m are of less relevance since they are not observed to develop in hydrodynamic simulations of interest, which typically show the number of vortices in agreement with the most unstable m .

4.6 Solutions to the full governing linear equation

Solutions to the full governing equation (13) for the fiducial cases with $Q_0 = 4$ above are shown in Fig. 3. In Fig. 3 we compare growth

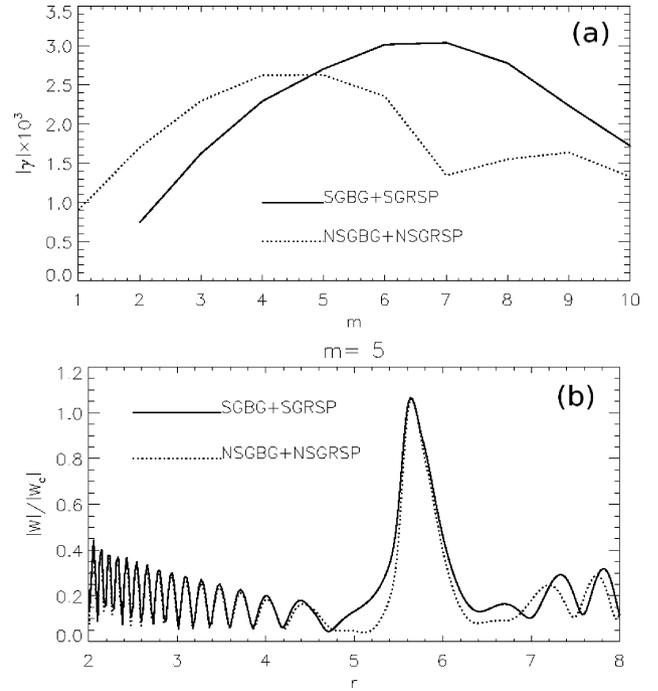


Figure 3. Growth rates obtained for the full governing equation (a) and the eigenfunction $|W|$ for $m = 5$ (b). Here $W_c = W(r = 5.7)$. Solid lines are for the fully SG case and dotted lines are for the fully NSG case. The initial disc had $Q_0 = 4$. In (a), the local maximum around $m = 9$ in the dotted curve is most likely caused by a boundary condition effect. Modes with $m \geq 7$ without self-gravity are not seen in non-linear simulations and are not relevant.

rates and the $m = 5$ eigenfunction. The corotation radii for the SG and NSG cases are $r_c = 5.88$ and 5.79 , respectively. These radii are close to the local vortensity minimum or $\max(\eta^{-1})$. In the NSG case, corotation almost coincides with the extremum.

The NSG case has maximum growth rate for $m = 4$, and the SG case is most unstable for $m = 6-7$. Modes with $m < 5$ are stabilized while $m > 5$ are destabilized by self-gravity. This results in a shift to higher m modes when self-gravity is included in non-linear simulations of the model. Modes with the larger m values are destabilized, but this is not in contradiction with earlier analysis which assumed small m and did not take account of variations in the form of the background state. It is shown in Section 4.7 below that the destabilization of higher m modes is in fact due to the change in the background state.

Results for $m < 5$ are in essential agreement with those obtained in the quadratic approximation, giving us confidence in our view of the importance of the dominance of the corotation region. As with the quadratic approximation, we did not find a $m = 1$ mode with dominant disturbance about the gap edge in the SG case. This suggests that such modes are the most easily stabilized by increasing self-gravity. In the NSG case, the growth rates for $m \geq 8$ are larger than $m = 7$ and do not follow the trend of decreasing growth rates seen from $m = 5 \rightarrow 7$. Note in Fig. 3(a) (and Fig. 4a below) the local maxima at $m = 9$ for NSG cases. The NSG $m \geq 8$ modes are unlikely to be the same type of vortex modes as $m \leq 7$ because the former have significant wave-like regions in W and are not concentrated near corotation as for $m \leq 7$. In addition, the $m \geq 8$ region is also where the quadratic approximation appears to fail. These wave-dominated modes demand radiative boundary conditions rather than the simplistic conditions applied here, which

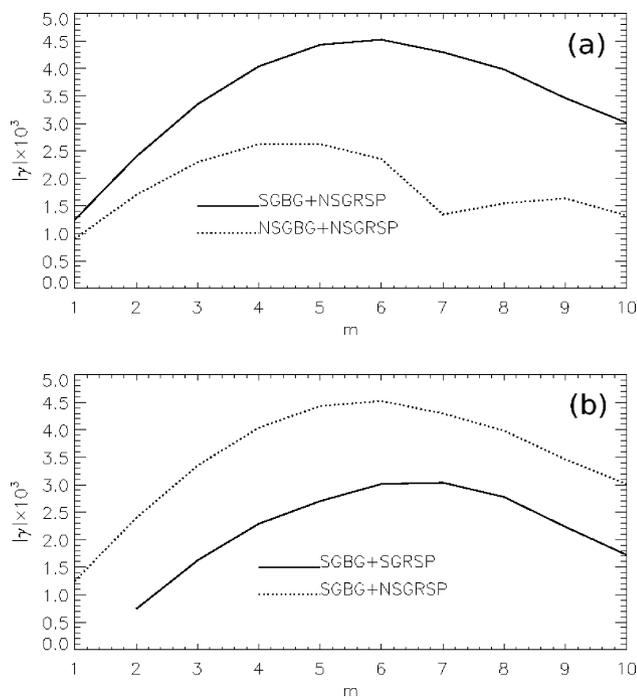


Figure 4. The effect of self-gravity on growth rates (as a function of azimuthal wavenumber m) of vortex-forming, gap edge instabilities through the background (a) and through the response (b). The disc model is $Q_0 = 4$. In (a), the local maximum around $m = 9$ in the dotted curve is most likely caused by a boundary condition effect. Modes with $m \geq 7$ without self-gravity are not seen in non-linear simulations and are not relevant.

are appropriate for boundary condition insensitive modes such as local vortex-forming modes. The NSG $m \geq 8$ modes identified here are thus likely to be artefacts of the boundary condition. Fortunately, these modes are irrelevant because they are not the most unstable, nor are they observed in the corresponding non-linear simulations. By considering the behaviour for $m \leq 7$, we can expect a cut-off for vortex modes around $m = 8$ for this model.

The $m = 5$ eigenfunctions shown in Fig. 3 are similar. Both have dominant disturbance around corotation. Behaviour in the region $r \leq 4.6$ is essentially identical. The largest difference is found in the region $r \geq 6.4$. The disturbance around corotation is also somewhat wider in the SG case. Comparing with Fig. 2, we see that the quadratic approximation captures the main feature of the mode in the corotation region. However, it removes the wave-like behaviour in the exterior disc.

4.7 The role of self-gravity

The calculations above can be compared to hydrodynamic simulations where self-gravity is either enabled or not. The linear problem allows one to examine separately the effect of self-gravity through its influence on the basic state and through its influence on the linear response.

We continue with the disc model with $Q_0 = 4$ and compare growth rates for a pair of models, one with and the other without self-gravity in setting up the basic state (i.e. the simulations were run with and without self-gravity enabled), but both without self-gravity implemented in the linear mode calculation (i.e. setting $\Phi' = 0$). In addition we compared another pair of models, one with and the other without self-gravity implemented in the linear response, but with both having the background state calculated with the disc self-gravity incorporated.

Fig. 4(a) examines the influence of self-gravity through its modification of the background state. It shows that this modification is destabilizing. This is not surprising because a deeper gap is set up when self-gravity is included in the simulation. In going from the case without self-gravity to the one with self-gravity, the most unstable mode shifts from $m = 4-5$ to $m = 6$. The vortex modes do not require self-gravity to operate. However, the difference in growth rates decreases as m decreases, so the effect due to the modification of the background is smallest for small m . In other words, high m modes can be made more unstable by including self-gravity in the base state.

Fig. 4(b) compares the growth rates obtained from linear calculations for the same background state but with and without self-gravity implemented in the response. The analysis in Section 3.1 applies to this comparison. Consistent with that, enabling self-gravity in the response decreases $|\gamma|$ and stabilizes the system against modes with corotation associated with a vortensity minimum for *any value of* m . Unlike the effect acting through the background, the influence of self-gravity is more significant for low m and can lead to stabilization. The $m = 1$ mode has been stabilized. The most unstable mode without self-gravity has $m = 5$ and including self-gravity shifts it to $m = 6$.

The fact that self-gravity acting in the linear response and background state has opposite effects on growth rates is consistent with the comparison between fully SG and fully NSG cases (Fig. 3). The effect of self-gravity through changes to the background and through direct influence on the linear response both contribute to favouring higher m . However, the background effect is achieved by increasing high m growth rates, whereas the effect via the response works by stabilizing low m modes in accordance with the discussion

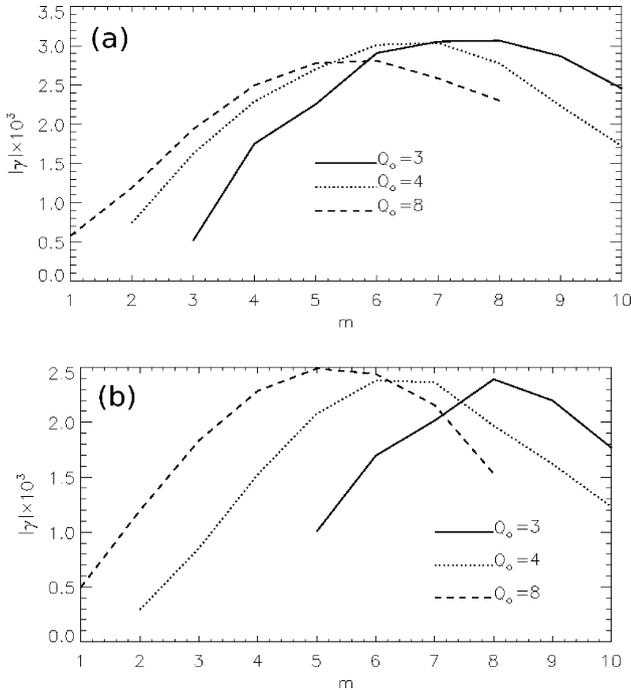


Figure 5. Growth rates for unstable modes in background discs with $Q_0 = 3, 4, 8$. Panel (a) is obtained from solving the full linear equation, while panel (b) is from the quadratic approximation. Self-gravity is fully incorporated throughout.

in Section 3.4. Thus low m vortex modes become disfavoured. Overall, we expect more vortices to form, corresponding to increasing m , with increasing self-gravity.

4.8 Models with different Q_0

In the calculations presented below, self-gravity is included both in setting up the background and in evaluating the linear response. We compare the fiducial calculation to cases with $Q_0 = 3$ (a more massive disc) and $Q_0 = 8$ (a less massive disc). The $Q_0 = 8$ case has a mass $M_d = 0.012M_*$ which is usually considered to be NSG in disc–planet simulations.

Growth rates are shown in Fig. 5(a). The plot demonstrates a clear shift of the most rapidly growing mode to higher m as Q_0 is lowered (increasing disc mass). For $Q_0 = 8$ the most unstable mode has $m = 5$ –6 and for $Q_0 = 3$ it shifts to $m = 7$ –8. The shift is accompanied by the stabilization (or loss) of low m modes. For example, when the disc mass is halved as Q_0 changes from $Q_0 = 4$ to 8, the $m = 3$ growth rate decreases by a factor of 3. Modes with the lowest m are stabilized by strong self-gravity, as we did not find $m = 1, 2$ modes when $Q_0 = 3$.

These calculations have also been performed in the quadratic approximation. Fig. 5(b) shows a similar dependence of the growth rates on Q_0 in this case. The shift to higher m is more apparent. For $Q_0 = 3$ we did not find modes with $m \leq 4$. Since the approximation reinforces the localized property of vortex-forming modes, it means that localized modes for low m become increasingly unlikely as we lower Q_0 . Hence, for massive discs only high m vortex-forming modes can develop. The agreement between results obtained using the quadratic approximation and the full governing equation indicates a lack of sensitivity to boundary conditions and so is reassuring.

5 NON-LINEAR HYDRODYNAMIC SIMULATIONS OF VORTEX-FORMING INSTABILITIES

We present hydrodynamic simulations of vortex formation and evolution at edges of gaps opened by a giant planet for discs of varying masses with self-gravity self-consistently included. The disc–planet models have been already described in Section 2.1. We consider a Saturn mass planet with $M_p = 3 \times 10^{-4}M_*$. The kinematic viscosity $\nu = 10^{-6}$ corresponds to an α viscosity of $\alpha = \nu/(c_s H) = 1.8 \times 10^{-4}$ at the planet’s fixed orbital radius at $r_p = 5$.

Higher viscosities corresponding to $\alpha = O(10^{-3})$, which is commonly assumed for protoplanetary discs, inhibit vortex formation (de Val-Borro et al. 2007; Lin & Papaloizou 2010). If planetary masses appropriate for type I migration are used, then a much lower viscosity would be required for vortex formation to occur (e.g. Li et al. 2009, where $\alpha \leq 10^{-5}$ was needed). Since we consider a giant planet, such low-viscosity regimes are not needed in order to develop vortices.

5.1 Numerical scheme

The hydrodynamic equations are evolved using the FARGO code (Masset 2000). FARGO is an explicit finite-difference code similar to ZEUS (Stone & Norman 1992) with second-order accuracy in space. It employs a modified azimuthal transport to achieved large time-steps, otherwise limited by the Keplerian velocity near the inner boundary of the domain. Self-gravity for FARGO was implemented and tested by Baruteau & Masset (2008). The gravitational acceleration due to the disc is calculated directly using Fast Fourier Transforms in both azimuth and radius. The latter requires the radial domain to be doubled when calculating the self-gravity potential.

The disc is divided into $N_r \times N_\phi = 768 \times 2304$ computational grid cells in radius and azimuth. The computational grid in the radial direction is logarithmically spaced. We impose an open boundary at $r = r_i$ and non-reflecting boundary as used by Zhang et al. (2008) at $r = r_o$ (Godon 1996). Since vortices are localized features, as long as gap edges are far from boundaries, boundary conditions can only have a limited effect. We also performed some simulations with damping boundary conditions (de Val-Borro et al. 2007) and open outer boundaries. We found similar results to those presented below.

5.2 The effect of self-gravity

We compare cases where disc gravitational potential is either included or excluded in the total potential calculation. This is the standard distinction between SG and NSG disc–planet simulations (Nelson & Benz 2003a; Zhang et al. 2008). We consider the $Q_0 = 4$ ($Q_p = 7$) disc. The disc mass is $M_d = 0.024M_*$.

Fig. 6 shows vortensity contours that illustrate vortex formation at the outer gap edge. There is a local vortensity maximum at about $r = 5.5$ that has been produced by flow through shocks induced by the planet. This maximum remains largely undisturbed indicating that instability is associated with structure outside it and associated with the vortensity minimum. More vortices are excited when self-gravity is included. In that case, the $m = 6$ vortex mode dominates whereas the $m = 3$ mode dominates in the NSG case. This behaviour is consistent with linear calculations.

Fig. 6 also shows that vortices have radial length-scales comparable to the local scaleheight [$\simeq H(6) = 0.3$]. The vortices in these cases have radial sizes of $\sim 2.2H(6)$ without self-gravity and

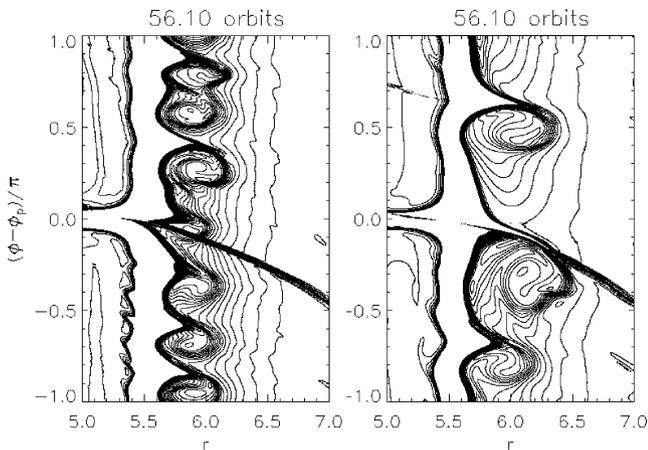


Figure 6. Vortensity contours for $Q_0 = 4$ with (left) and without (right) self-gravity. The planet is located at $r = 5, \phi = \phi_p$. The thick lines crossing the outer radial boundaries correspond to spiral shocks induced by the planet.

$\sim 1.3H(6)$ with self-gravity. The vortices are of smaller radial extent when self-gravity is included because of the preference for higher m . A decrease in radial size can be expected from the decrease in width of the WKBJ evanescent zone centred on corotation. This is determined by the condition:

$$(\sigma + m\Omega)^2 = \kappa^2(1 - 1/Q^2). \quad (36)$$

Writing $\sigma = -m\Omega(r_c)$, it is straightforward to show that the radial width of the evanescent zone approximately scales as $1/m$. Since the preferred m is approximately a factor of 2 smaller, we expect vortices in the case without self-gravity to be double the size of those in the case where self-gravity is included.

In the case with self-gravity, the vortices are approximately centred along the radius $r = 5.9$, close to the corotation radius expected from linear calculations, $r_c = 5.88$. In the NSG case, linear calculation gives $r_c = 5.81$ for $m = 3$, but the vortices are approximately centred along $r = 6$. However, perturbations here are already in the non-linear regime and interaction between vortices or with the spiral shock may shift the vortices around. Note too that in this regard there is more variation in the radial locations of the vortices as compared to the case with self-gravity.

5.3 Varying disc mass: gap profiles

We present simulations of discs models with $1.5 \leq Q_0 \leq 8$, equivalently $2.6 \leq Q_p \leq 14$ or $0.063 \geq M_d/M_* \geq 0.012$. The equilibrium gap profiles have a range of Q values with local extrema of $1.5 \leq \min(Q) \leq 9.5$ and $4.8 \leq \max(Q) \leq 21.6$ near the outer gap edge. The gap profiles opened by the planet are given in Fig. 7 for a range of Q_0 that develops vortices. Outside the plotted region the curves are indistinguishable. The gap deepens with decreasing Q_0 and gap edges become steeper. In going from $Q_0 = 8 \rightarrow 2$ the gap depth $|\Delta\Sigma/\Sigma|$ increases by about 0.05–0.08, similarly the bumps near gap edges increase by 0.05–0.06. On a global scale, self-gravity becomes more important with increasing radius. However, we observe no trend in the difference between gap profiles with respect to radius.

Self-gravity affects gap structure on a local scale by increasing the effective planet mass so that $M_p \rightarrow M'_p$. A straightforward estimate, based on the unperturbed disc model, of the expected mass within the Hill radius, r_H , of a point mass planet with $M_p = 3 \times 10^{-4} M_*$ is $M_H = 0.047 M_p$ for $Q_0 = 8$ and $M_H = 0.17 M_p$ for

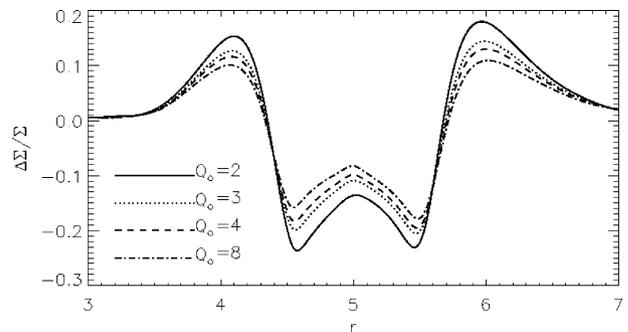


Figure 7. Gap profiles opened by a Saturn mass planet in SG discs as a function of disc mass, parametrized by Q_0 . The azimuthally averaged relative surface density perturbation is shown. The planet located at $r = 5$.

$Q_0 = 2$. It is likely that M_H or at least some significant fraction of it adds to the effective mass of the planet acting on the disc when self-gravity is self-consistently included. Thus M'_p , and therefore also the gap depth, is expected to increase with disc mass. Thus we note that without carefully tuning M_p , SG disc–planet calculations with different surface density scales will always differ in M'_p . Since the gap profiles differ, as we have seen above, stability is affected also. Note that the gap width w in Fig. 7 does not change greatly with Q_0 , which is consistent with the scaling $w \propto r_H \propto M_p^{1/3}$. However, the peaks/troughs for $Q_0 = 2$ do lie slightly closer to the orbital radius r_p than other cases. This is because shocks are induced closer to the planet due to increased M'_p (Lin & Papaloizou 2010).

5.4 Varying disc mass: gap stability

Fig. 8 shows snapshots of the relative surface density perturbation as instability sets in. Consider $Q_0 \geq 2$ first. The instability is associated with the outer gap edge while the inner edge remains relatively stable. In the least massive disc $Q_0 = 8$, three to four vortices form at the outer gap edge, similar to what happens in the NSG disc in Section 5.2. As we increase the disc mass, more vortices develop. By $Q_0 = 2$, eight surface density maxima can be identified. Note that a vortex may be obscured if it coincides with the planetary wake. In moving from $Q_0 = 8 \rightarrow Q_0 = 2$, vortices become smaller and their centres move inwards. When $Q_0 = 8$ local surface density maxima lie just outside $r = 6$, while for $Q_0 = 2$ they lie just interior to $r = 6$.

Increasing M_d makes the perturbation more global. Vortices in massive discs can gravitationally perturb parts of the disc further out, similar to a planet. Lowering Q_0 , vortices develop longer, more prominent trailing spirals exterior to them. This is most notable with $Q_0 = 2$, where the vortex spirals can have comparable amplitudes as the planet wake. However, increasing self-gravity even further, we expect a global instability to eventually develop as the vortices perturb the disc strongly via self-gravity.

The $Q_0 = 1.5$ disc does not develop vortices. This is consistent with the stabilization effect of self-gravity on the vortex instability through the linear response. Instead, $Q_0 = 1.5$ develops global spiral instabilities associated with the gap edge. We call these edge modes. They have been suggested to occur in SG discs with surface density depression (Meschiari & Laughlin 2008). Here, we have shown that they can indeed develop in gaps self-consistently opened by a giant planet. Detailed discussions of edge modes are beyond the scope of this paper and are presented in Lin & Papaloizou (2011), but for reference we note some important differences to the vortex modes. Edge modes here are associated with local vortensity *maxima*, which

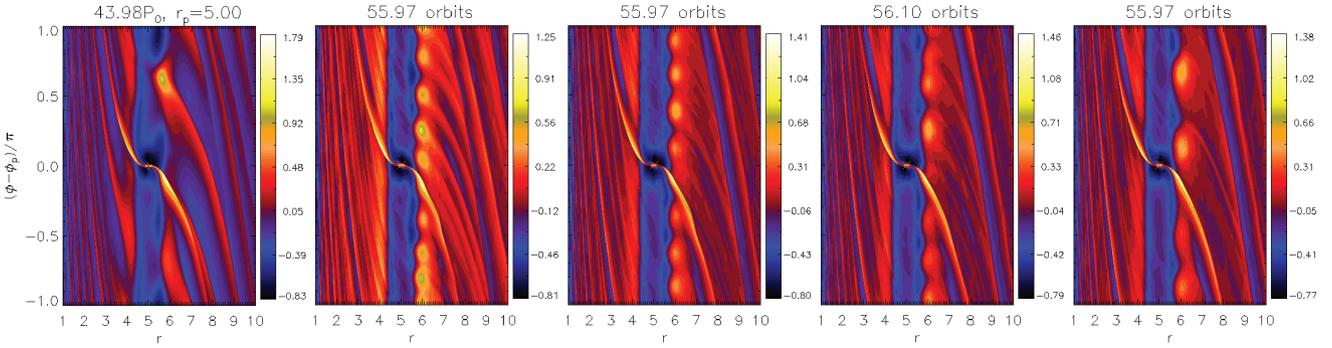


Figure 8. Instability at the outer gap edge of a Saturn mass planet, in discs with minimum Toomre parameter, from left to right, of $Q_0 = 1.5, 2, 3, 4, 8$. The relative surface density perturbation is shown.

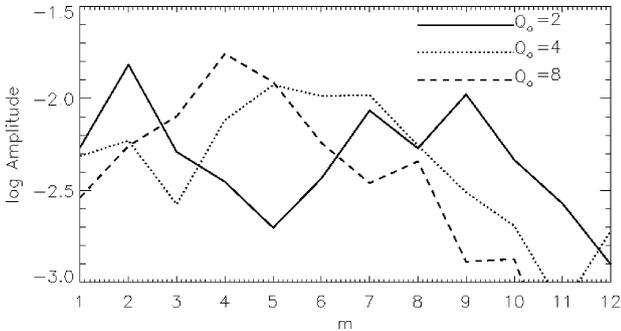


Figure 9. Fourier amplitudes of the surface density in the outer disc $r \in [5, 10]$, normalized by the $m = 0$ component, for some of the models illustrated in Fig. 8.

lie closer to the planet than vortensity minima. Perturbations at the inner gap edge are also identified and correlate with those at the outer gap edge, where the overdensity is largest (and stronger than for vortex disturbances). Communication between the two sides of the gap is only possible with sufficient self-gravity.

The Fourier amplitudes of the surface density in $r \in [5, 10]$, as a function of m , is shown in Fig. 9 for $Q_0 = 2, 4, 8$ at $t = 56P_0$. Amplitudes have been normalized by the axisymmetric component. The shift to higher m vortex modes with increasing disc mass is evident as expected from linear calculations. For $Q_0 = 8, 4$ and 2 , the preferred vortex modes have respectively $m = 4, 5-7$ and $7-9$, with average amplitudes that decrease with Q_0 . The latter may reflect the stabilization effect of self-gravity on linear modes with low m .

An important region in Fig. 9 is $m \leq 3$. There is a peak in amplitude at $m = 2$ for $Q_0 = 2, 4$. These are the edge modes described above. They were not seen in linear calculations because there we focused on finding vortex modes, which have corotation at or close to local vortensity minima. The loss of low m vortex modes with increasing self-gravity observed in linear calculations is replaced by the increasing prominence of global edge modes. Fig. 9 shows the $m = 2$ amplitude becomes more significant with increasing self-gravity.

In the $Q_0 = 2$ simulations, we do see evidence of an $m = 2$ disturbance hindering vortex evolution. This transition case is difficult to analyse since both types of instabilities develop. Fig. 9 shows the edge and vortex modes have comparable amplitudes in $r > 5$. However, considering the region $r \geq 7$ for $Q_0 = 2$ we found $m = 2$ is dominant, because the edge mode is global, whereas vortex mode disturbances are localized to the edge ($r \simeq 6$). Increasing self-gravity

further, we expect eventually edge modes become dominant, this is seen in the $Q_0 = 1.5$ case in Fig. 8.

5.5 Evolution and merging of self-gravitating vortices

We examine the evolution of vortices under the influence of self-gravity. Fig. 10 compares surface density perturbations for a range of disc masses at $t = 100P_0$. Since the vortices emerge at roughly $t = 56P_0$, they have evolved for a similar time. Typical growth rates for vortex modes are $\tau \sim 6P_0$, which implies the vortices have evolved for about 7τ , well into the non-linear regime.

The snapshot for $Q_0 = 8$ shows a single vortex, resulting from the merging of the initial vortices. The vortex disturbance is largely confined to within a local scale-height of the gap edge. Such a result is typical for simulations with no self-gravity (de Val-Borro et al. 2007). The case with $Q_0 = 4$ shows vortex merging taking place, as individual surface density maxima can still be identified in the large vortex behind the shock. Note there is a smaller vortex just passing through the shock. When $Q_0 = 4$, a single vortex forms at $t \simeq 110P_0$. Fig. 10 shows that increasing the disc mass delays vortex merging. For $Q_0 = 3.5$, five vortices remain and for $Q_0 = 3$ and $Q_0 = 2.5$, seven vortices remain. They have not merged into a single vortex as happened when $Q_0 = 8$. A 25 per cent increase in disc mass as $Q_0 = 4 \rightarrow 3$ causes merging to be delayed by $50P_0$. This suggests that increased gravitational interaction between vortices opposes merging.

As Q_0 is lowered the inter-vortex distance increases. When $Q_0 = 2.5$ their azimuthal separation can be larger than the vortex itself. Also, vortices become less elongated and more localized being symptomatic of gravitational condensation. Trailing wakes from vortices also become more prominent as the vortices more strongly perturb the disc via their self-gravity. In fact, the planetary wake becomes less identifiable among the vortex wakes. Note vortex wakes are mostly identified with vortices upstream of the planetary wake, rather than just downstream. The vortex wakes appear to detach from the vortex after passing through the planetary shock. There is a sharp contrast in gap structure between the $Q_0 = 2.5$ and 8 cases. The single vortex that results when $Q_0 = 8$ is aligned along the outer gap edge, which is still approximately identified as circle $r = 6$. However, when $Q_0 = 2.5$ the vortices and their wakes intersect the circle $r = 6$ making the radius of the outer gap edge less well defined.

An important feature of vortices is that the wakes are associated with the excitation of density waves which transport angular momentum transport outwards (see e.g. Paardekooper, Lesur &

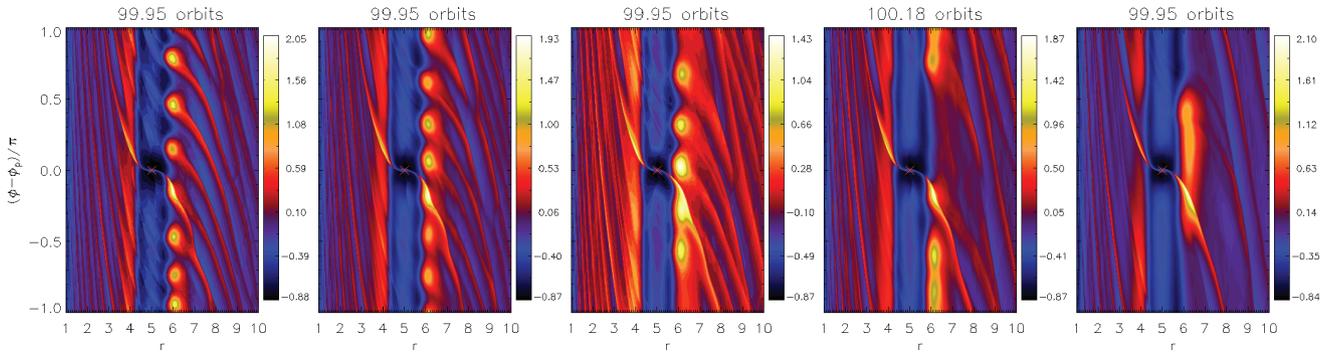


Figure 10. Non-linear evolution of vortex instabilities at the outer gap edge of a Saturn mass planet as a function of disc mass, parametrized by the minimum Toomre parameter Q_0 , from left to right, of $Q_0 = 2.5, 3.0, 3.5, 4.0, 8.0$.

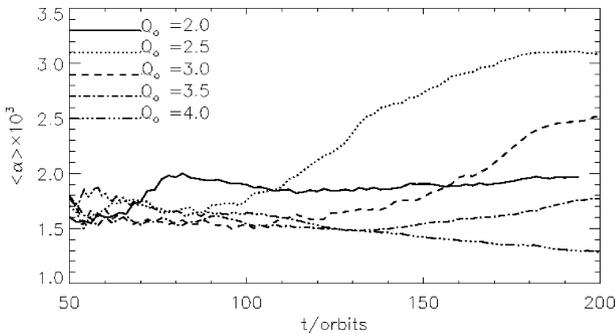


Figure 11. The running time average of the viscosity parameter α averaged over the vortex region for a range of disc models.

Papaloizou 2010). A measure of this is made through the standard α viscosity parameter. We define it here through

$$\alpha(r, \varphi) \equiv \Delta u_r \Delta u_\varphi / c_s^2(r), \quad (37)$$

where Δ denote deviation from azimuthal averaged values. As we are interested in the vortex region, α is spatially averaged over the annulus $r \in [5.7, 7.1]$ and its running-time average $\langle \alpha \rangle$ is plotted in Fig. 11. The parameter $\langle \alpha \rangle$ associated with vortices is $O(10^{-3})$, an order of magnitude larger than the imposed value associated with ν . When $Q_0 = 4$, $\langle \alpha \rangle$ decays steadily after an initial transient growth. The case $Q_0 = 2$ is also shown. In that case $\langle \alpha \rangle \simeq 1.9 \times 10^{-3}$ for $t \gtrsim 80P_0$. Interestingly, for $Q_0 = 2.5\text{--}3.5$ there is growth in $\langle \alpha \rangle$ over several tens of orbits.

The behaviour of $\langle \alpha \rangle$ as a function of Q_0 is consistent with the general picture of vortex formation and evolution (Fig. 10). We see below that the decay in $\langle \alpha \rangle$ is associated with vortex merging leading to fewer vortices. This is the case for $Q_0 = 4$. Vortex merging happens more readily for lower disc masses, hence, although multiple vortices develop from the instability, this phase does not last long enough for the multiple-vortex configuration to significantly transport angular momentum. As we increase self-gravity when $Q_0 = 3.5 \rightarrow 2.5$, vortices become less prone to merging. Hence, the multiple-vortex phase lasts longer. They have time to evolve into compact objects that further perturb the disc. This is consistent with fact that $\langle \alpha \rangle$ -growth is prolonged with increased disc mass, as merging is delayed. However, if self-gravity is too strong, such as when $Q_0 = 2.0$, growth is again limited, because the $m = 2$ edge mode develops and hinders vortex evolution.

5.6 Long-term evolution of gap edge vortices

We consider the long-term evolution of vortices in the disc with $Q_0 = 3$. Fig. 12 shows the instantaneous α viscosity parameter, average overdensity (defined in regions where the relative surface density perturbation is positive) and surface density contour plots of the vortex region. The viscosity parameter α grows from $t = 100P_0$ to $t = 170P_0$ with $\max(\alpha) \simeq 8 \times 10^{-3}$, which is 50 times larger than the contribution from the background viscosity. At $t = 150P_0$ there remains six distinct vortices, one of which just passing through the planetary wake. The multiple-vortex configuration has been maintained for a further $\sim 50P_0$ since Fig. 10. With weak self-gravity, a single vortex would have formed through merging. Delayed merging allows individual vortices to evolve and become planet-like. The typical overdensity in a vortex is ~ 1 at $t = 100P_0$ and increases to >1.5 by $150P_0$. As a consequence we expect increased disturbance in the surrounding disc. The phase of α -growth correlates with a linear increase in the average overdensity in the vortex region.

At $t = 170P_0$, Fig. 12 shows six vortices still remain, but the pair just upstream of the planet is merging. The parameter α is a maximum. However, after a burst of vortex merging events, α decreases rapidly. The snapshot at $t = 180P_0$ shows a quieter disc with only three vortices of similar size to those before merging. This is unlike the cases with weak self-gravity where a larger vortex results from merging. Fig. 13 compares post-merging vortices in discs with $Q_0 = 3$ and 4. For $Q_0 = 3$, the post-merging vortices are localized in azimuth, with $Q \sim 1$ and the contour plot shows the densest vortex has an overdensity of $\simeq 3.75$. However, for the disc with $Q_0 = 4$, a single vortex forms that extends about half the total azimuth and has $Q > 2$.

The mass of the disc with $Q_0 = 3$ is $M_d = 0.031M_*$, usually considered insufficient for gravitational collapse by itself. However, the comparison above shows that addition of self-gravity to the vortex instability induced by an embedded planet may enable collapse into compact objects that survive against shear.

We found that vortices in the disc with $Q_0 = 3$ noticeably affect the gap structure. Fig. 14 shows several snapshots of the gap structure from $t = 100P_0$, when there were multiple vortices, to the end of the simulation when a vortex pair remained. The gap profile for $Q_0 = 4$ is also shown for comparison. Neither the single vortex for the disc with $Q_0 = 4$ nor multiple vortices with $Q_0 = 3$ affect the one-dimensional gap profile at $t = 100P_0$, because the disturbances only redistribute mass in the azimuthal direction. However, the original bump at the outer edge is diminished after vortex merging takes place in the disc with $Q_0 = 3$ at $t = 200P_0$. A surface density

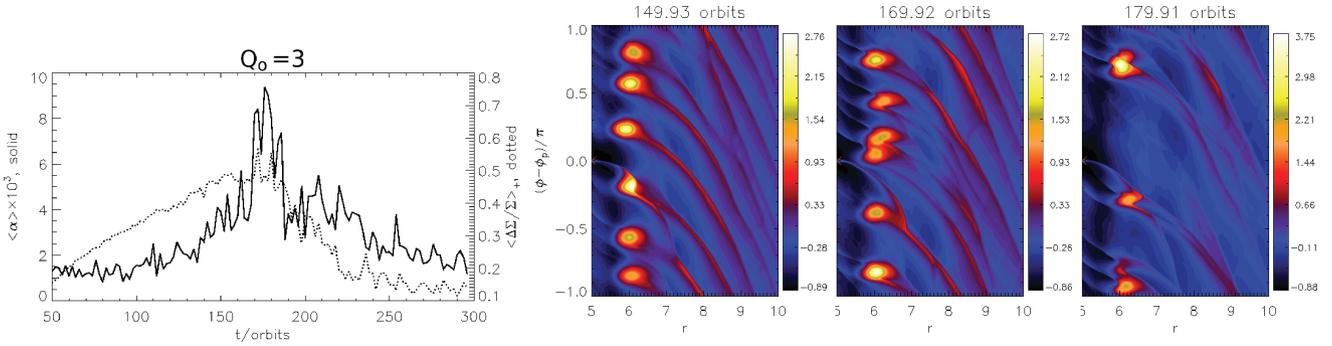


Figure 12. Vortex evolution for the disc with $Q_0 = 3$. Line plot: instantaneous α viscosities (solid) and average overdensity (dotted). Contour plots: relative surface density perturbations before and after vortex merging.

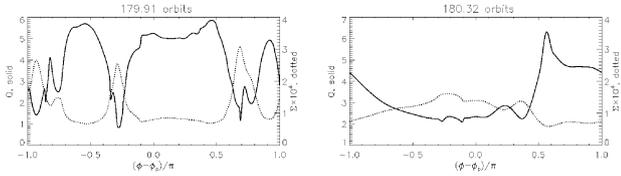


Figure 13. Comparison of post-merging vortices for discs with $Q_0 = 3$ (left) and $Q_0 = 4$ (right). The Toomre Q (solid) and surface density (dotted) averaged over $r \in [6, 6.5]$ is shown as a function of azimuth relative to the planet.

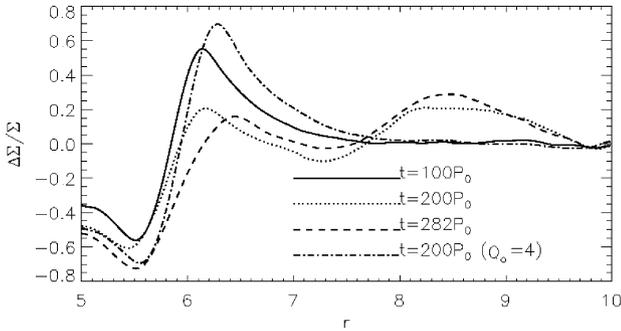


Figure 14. The gap profile for the $Q_0 = 3$ disc at different times (solid, dotted and dashed) as shown through relative surface density perturbation. A snapshot of the profile for the disc with $Q_0 = 4$ (dot-dashed) is also given for comparison.

depression of $\Delta\Sigma/\Sigma \simeq -0.1$ then develops at $r = 7.3$ and a bump develops at $r = 8.5$. These features last to the end of the simulation. Self-gravitating vortices behave like planets and gap opening is to be expected. Assuming vortices lie near the surface density maximum at $r = 6$, the creation of the surface density deficit for $r \in [7, 7.5]$ and the new surface density maximum at $r = 8.5$ could be induced, in a similar manner as it would be by a planet, through the outward transport of angular momentum by the density waves launched by the vortices. The material removed from the region of the deficit, which has gained angular momentum, ends up contributing to the new surface density maximum at $r = 8.5$.

5.7 Anticyclonic vortices

For the disc with $Q_0 = 3$, a vortex-pair forms at $t \simeq 230P_0$ and lasts until the end of the simulation ($t \simeq 300P_0$) and is reminiscent of co-orbital planets. The pair survives on a time-scale beyond which merging occurs in the simulations without self-gravity. A snapshot is shown in Fig. 15(a). Two blobs can be identified along the gap

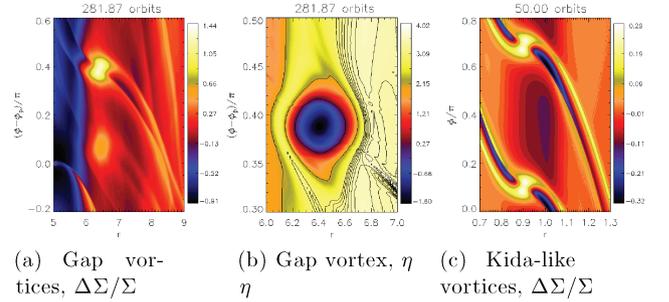


Figure 15. The final vortex pair at the end of the disc–planet simulation for $Q_0 = 3$. Panel (a): relative surface density contour plot. Panel (b): vortensity contours for the upper vortex. Shown in panel (c) is the relative surface density for a pair of vortices formed by imposing Kida vortex solutions as a perturbation in a standard power-law disc.

edge, the upper vortex being more overdense than the lower one. They lie at a radius $r_v = 6.4$, corresponding to local surface density maximum in azimuthally averaged one-dimensional profiles, which is expected to be neutral for vortex migration (Paardekooper et al. 2010).

The upper vortex in Fig. 15(a) is different to the pre-merging vortices or those with weak self-gravity. It has two spiral disturbances extending from the vortex to $(r, \phi - \phi_p) = (9, -0.2\pi)$, whereas the pre-merging vortices have one trailing spiral. Its vortensity field is shown in Fig. 15(b). The vortex core has $\eta < 0$, whereas the final large vortex in the disc with $Q_0 = 4$ has $\eta > 0$ in its core, though it is still a local minimum.

From the vortensity field the region where $\eta < 0$ has a mass of $M_v \simeq 5.46 \times 10^{-5} M_* \simeq 0.18 M_p$ and average radius $\bar{r} \simeq 0.92H(r_v)$. This is about 18 Earth masses if $M_* = 1M_\odot$. Denoting the mean square relative velocity (to the vortex centre) as Δv^2 , we found $GM_v/\bar{r}\Delta v^2 \sim 3.9$. This region is gravitationally bound. Although planets of such mass are not expected to open gaps, the surface density deficit in $r = [7, 7.5]$ in Fig. 15(a) indicates vortices may do so. This may be because the process is assisted by the fact that a vortex of size H produces a perturbation of a magnitude similar what would be produced by Saturn when $h = 0.05$, even without self-gravity (Paardekooper et al. 2010). We estimate by inspection that the region with negative vorticity has semimajor and semiminor axis of $a \sim 1.57H(r_v)$, $b \simeq 0.55H(r_v)$, respectively, corresponding to aspect ratio of 2.9.

Finally, Fig. 15(c) shows $\Delta\Sigma/\Sigma$ for a pair of Kida-like vortices placed in a disc. We note the similarity between the Kida-like vortex and the upper vortex in Fig. 15(a), particularly the double wake structure, the tilted core and the surface density deficit just outside

the vortices. This is a surprising coincidence, given that the simulation setups that produced them were completely different (for more details see below).

6 SIMULATIONS OF CO-ORBITAL KIDA VORTICES

In the simulations above, we observed one of the effects of self-gravity is to delay vortex merging. This effect is significant when vortices develop into compact structures. Furthermore, the similarity between the final vortex in the disc with $Q_o = 3$ and a Kida-like vortex (Kida 1981) motivates us to consider the effect of self-gravity on the interaction between two Kida vortices.

The simulations presented here are supplementary and focus on the interaction between vortices without the perturbation of the planet. Thus, we isolate effects due to vortex–vortex interactions and the influence of self-gravity on those without interference from the planet. In this way we obtain a better understanding of some of the results above.

6.1 Numerical setup

We use a standard power-law disc with initial surface density profile $\Sigma = \Sigma_0 r^{-1/2}$, where Σ_0 is a scaling constant for $r \in [0.25, 2.5]$. The value of Σ_0 is chosen to provide a specified Q_{Kep} at a specified radius as before. The initial azimuthal velocity is chosen so that the disc is in hydrostatic equilibrium. The initial radial velocity is zero. The disc is locally isothermal with $h = 0.05$. The viscosity is $\nu = 10^{-9}$ which is essentially the inviscid limit. No planets are introduced, but in this section we use (r_p, φ_p) to denote a vortex’s centroid.

To set up vortices in the disc, we follow Lesur & Papaloizou (2009) and introduce velocity perturbations $(\delta u_r, \delta u_\varphi)$ that correspond to the elliptical vortices of Kida (1981) in an incompressible shear flow. Details of the implementation is described in Appendix B. Two vortex perturbations are imposed with initial angular separation $\theta = \pi/2$. Table 1 summarizes our numerical experiments. We consider switching self-gravity on or off, a range of disc masses [here specified through $Q_1 \equiv Q_{\text{Kep}}(1)$] and initial radial separations X_0 of the two vortices (so that for one of the vortices, $r_p \rightarrow r_p + X_0$).

We use the FARGO code with resolution $N_r \times N_\varphi = 800 \times 2400$ giving $\simeq 17$ grid points per scaleheight. The vortices are typically of

Table 1. Parameters for two-vortex interaction simulations. The first column gives the nomenclature for the simulations. The initial radial separation is X_0 and Q_1 is the Keplerian Toomre parameter Q_{Kep} at unit radius. The fourth column indicates whether self-gravity was included.

Case	X_0/H	Q_1	Self-gravity
Fnsg	0	8.0	NO
Fsg	0	8.0	YES
M1	0	15.9	YES
M2	0	5.3	YES
S1	0.1	8.0	YES
S2	0.2	8.0	YES
S3	0.3	8.0	YES
S4	0.6	8.0	YES

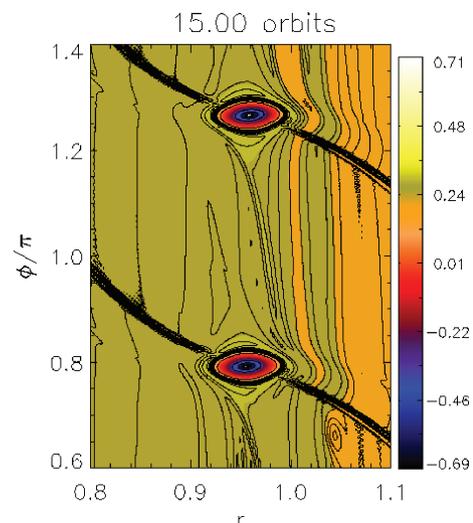


Figure 16. Vortices formed by imposing Kida vortex solutions as a perturbation to a global disc. The vortensity field (scaled) is shown.

that size. Damping boundary conditions are applied (de Val-Borro 2006).

6.1.1 Vortex pair interactions

As an example of vortices formed by the above procedure, Fig. 16 shows the vortensity field for the case Fnsg. The vortex pair circulates at $r \simeq 0.96$ and are localized in radius and azimuth. The vortex centroids have vorticity -1.57 (dimensionless units) in the non-rotating frame, close to the local Keplerian shear ($-1.5r_p^{-3/2} \simeq -1.59$). The vortices are not symmetric with respect to reflection in a co-orbital radius. The upper vortex has a long tail reaching the outer part of the lower vortex rather than its centroid. The wakes associated with each vortex are similar to those induced by a planet and responsible for vortex migration (Paardekooper et al. 2010).

Focusing on one vortex, the region with negative absolute vorticity has half width in radius and azimuth of $\simeq 0.6H$ and $1.3H$, respectively. This region has mass $m_v \sim 1.2 \times 10^{-5} M_*$, or about 4 Earth masses if $M_* = 1M_\odot$. The estimate is not far from the assumption that the final vortex has size H with average density that of the unperturbed disc at the location it was set up, which gives $m_v \sim 1.6 \times 10^{-5}$. If self-gravity is enabled, we can expect gravitational interaction between vortices to behave like that between two co-orbital planets of at least a few Earth masses.

6.2 Simulation results

We first compare cases Fsg and Fnsg. These are for the disc with $Q_1 = 8$ with and without self-gravity, respectively. The distance between vortex centroids as a function of time is shown in Fig. 17(a). It takes almost five times as long for the SG vortices to merge. Vortensity evolutionary plots show that NSG vortices begin to merge at $t = 28P_0$, but SG vortices at $t = 126P_0$. For Fnsg, vortices approach each other within $\sim 23P_0$ of formation and merge. Assuming they move relative to each other because of Keplerian shear, the time taken to merge implies that their radial separation at vortex formation must be $\simeq 0.14H$. This length-scale is resolved by about 2.5 grid cells indicating that a non-zero initial separation is generated by grid effects, despite the vortex perturbations being imposed at the same radius. In a pair of lower resolution runs ($N_r = N_\varphi/3 = 400$), vortex merging without self-gravity occurs at $t = 23P_0$, whereas

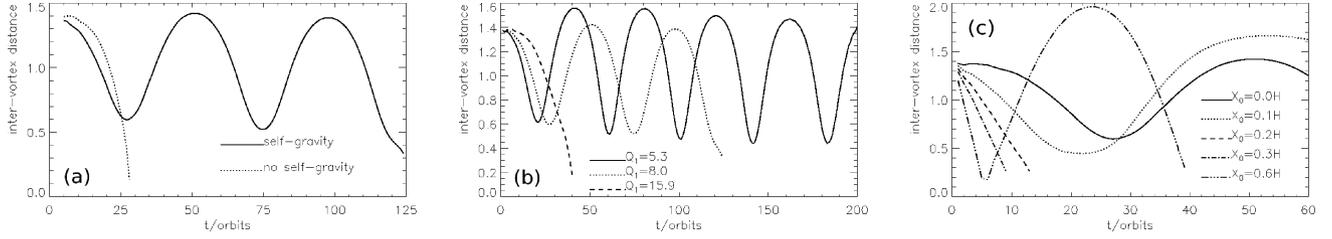


Figure 17. Inter-vortex distances for vortex pair simulations. Panel (a): with and without self-gravity for $Q_1 = 8$ (simulations Fsg and Fnsf). Panel (b): for discs of different masses characterized by Q_1 , the Keplerian Toomre parameter at $r = 1$, self-gravity being included (simulations Fsg, M1 and M2). Panel (c): vortices input with varying initial radial separation, X_0 , with self-gravity included for disc models with $Q_0 = 8$ (simulations Fsg, S1, S2, S3 and S4).

with self-gravity the vortex pair survive until the end of the simulation. These are consistent if the initial radial separation is increased due to lowered resolution, giving stronger differential rotation, hence earlier merging in the non-self-gravity run and increased impact parameter in terms of horseshoe orbits in the SG run (see below).

In the SG case Fsg, the inter-vortex distance oscillates with period $50P_0$ and there are two close encounters before merging. This is analogous to the survival mode of the vortex pair towards the end of the disc–planet simulation for $Q_0 = 3$. Given a maximal separation of $\simeq 1.4$ and that the vortices are circulating near unit radius, the maximal angular separation is $\simeq \pi/2$. The minimal separation during the first two close encounters is $\simeq 0.6$, or about $12H$ which implies that the vortices are on tadpole orbits. As the minimum separation is much larger than the typical vortex size, merging does not occur during the first two close encounters.

Next, in Fig. 17(b) we vary the disc mass via Q_1 . The reference case has $Q_1 = 8$. Doubling Q_1 to $Q_1 = 15.9$ (case M1) weakens self-gravity and vortices merge within few tens of P_0 . For $Q_1 = 5.3$ (case M2), the oscillation period is $\simeq 40P_0$ and maximal separation is 1.55, larger than for $Q_1 = 8$, as is the first minimum separation; the increased self-gravity has enhanced the mutual repulsion of vortices. The M2 two-vortex configuration lasts until the end of the simulation, but there is a secular decrease in the minimum separation (of 0.6 at $t = 20P_0$ and 0.45 at $t = 180P_0$) due to vortex migration. We expect merging to occur eventually. Consistent with the behaviour seen for gap edge vortices, increasing self-gravity delays merging.

Our final experiment varies the initial radial separation of the vortex perturbations. Results are shown in Fig. 17(c). Increasing X_0 to $0.1H$ from the reference case, the first minimum separation decreases to 0.45 from 0.6: vortices approach each other more closely, but still repel and undergo co-orbital dynamics. The increased oscillation amplitudes imply a larger tadpole orbit. For $X_0 = 0.2H$ and $0.3H$, vortex separation decreases more rapidly and becomes small enough for vortex merging. However, for $X_0 = 0.6H$, vortices simply circulate past by each other and no merging occurs, despite reaching a similar minimal separation as when $X_0 = 0.2H, 0.3H$. This implies merging requires that vortices should collide head on.

6.3 Vortices as co-orbital planets

The simulations presented above indicate that SG vortex pairs behave like co-orbital planets. This means there exists an initial radial separation (or impact parameter) below which vortices execute horseshoe orbits relative to each other. Analytical and numerical work indicates that vortices merge if their centroids are within $d \sim 3s$ of each other (e.g. Zabusky, Hughes & Roberts 1979; Melander, Zabusky & McWilliams 1988), where s is the vortex size. We cannot

assume the same critical d/s apply to our simulations because the situations are very different, but it is reasonable to expect merging if vortices can reach within some critical distance of one another.

The results above can be anticipated from existing treatments of co-orbital dynamics, which we discuss here for completeness. The first is Murray & Dermott (2000) model of the co-orbital satellites of Saturn, the Janus–Epimetheus system. The governing equation from Murray & Dermott (2000) gives a relationship between two configurations, the ‘final’ configuration (subscript f) and the ‘initial’ configuration (subscript i), in the form:

$$\left(\frac{\Delta r_f}{r_0}\right)^2 - \left(\frac{\Delta r_i}{r_0}\right)^2 = -\frac{4}{3}q[H(\theta_i) - H(\theta_f)], \quad (38)$$

$$H(\theta) = [\sin(\theta/2)]^{-1} - 2\cos\theta - 2,$$

where Δr is the radial separation of the satellites, r_0 the average orbital radius (assumed fixed), θ their angular separation and $q = \mathcal{M}/M_*$, with \mathcal{M} being the sum of the satellite masses. It is assumed $q \ll 1$.

Let us apply equation (38). We assume zero final radial separation, $\Delta r_f = 0$ when the vortices are at their minimum separation. Inserting $\theta_i = \pi/2$, $\Delta r_i = 7.1 \times 10^{-3}$ (corresponding to the initial conditions deduced for Fnsf), $r_0 = 1$ and $q = 2.3 \times 10^{-5}$ (corresponding to the measured mass of the negative absolute vorticity region in Fsg), equation (38) gives $\theta_f \simeq 0.41$ or a minimal angular separation of $\sim 8H$ so merging is not expected if vortices have size H . This estimate is lower than the observed value of 0.6, but we have used a lower limit on q . Inserting $q = 1.2 \times 10^{-4}$ gives $\theta_f = 0.6$, assuming all other parameters remain the same. This implies the gravitational mass of the vortex should be 20 Earth masses.

Equation (38) is illustrated in Fig. 18(a). For fixed X_0 , as we increase q (e.g. as a by-product of increasing the disc mass), the minimal distance Y increases, eventually becoming too large for merging to occur. For fixed q , Y decreases as X_0 increases. This is similar to a test particle on a horseshoe orbit in the restricted

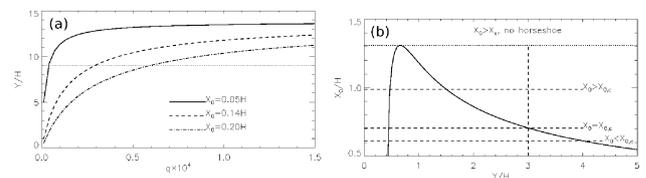


Figure 18. Merging conditions based on Murray & Dermott model of co-orbital satellites (a) and shearing sheet dynamics (b). In (a), Y is the minimum inter-vortex separation and the horizontal line is a hypothetical critical separation below which merging occurs. Thus merging is less likely for larger q . Fig. 18(b) illustrates equation (39); for a given X_0 , the allowed values of minimum separation Y lies between the intersection of the horizontal line $X_0/H = \text{constant}$ and the solid curve. The vertical dashed line is a hypothetical critical separation, below which merging occurs.

three-body problem. The larger its impact parameter (while still in libration), the closer it approaches the secondary mass.

We may also view the horseshoe turns in a shearing sheet approximation. In Appendix C we derive the inequality

$$\hat{x}_0 < \frac{1}{h} \left(\frac{8q}{3\hat{y}h} - \frac{q^2}{3\hat{y}^4 h^4} \right)^{1/2}, \quad (39)$$

where $\hat{x}_0 = X_0/H$, $\hat{y} = Y/H$. Given X_0 , equation(39) can be inverted to give the range of possible Y . This inequality is displayed in Fig. 18(b). There exists a critical $X_0 = X_s$, beyond which vortices do not execute horseshoe turns and instead circulate past each other. There may also exist a value $X_{0,c}$ such that $X_{0,c} < X_s$ and for $X_{0,c} < X_0 < X_s$, vortices execute U-turns but the values of Y attained are sufficiently small to allow merging. This happens only when the vortices have small enough mass. When $X_0 < X_{0,c}$, Y can be larger than critical, so that merging does not occur.

These simple models give qualitatively the same results as vortex pair and disc–planet simulations, and serve to explain the effect of self-gravity delaying merging by interpreting vortices as co-orbital planets that execute mutual horseshoe turns, thereby imposing a minimum inter-vortex separation (mainly in the azimuthal direction). If this minimum separation is still larger than a critical separation known to exist for vortex merging, then merging cannot occur. Hence, multiple-vortex configurations can be sustained longer as the strength of self-gravity is increased.

7 IMPLICATIONS ON VORTEX-INDUCED MIGRATION

Vortices at gap edges can lead to brief phases of rapid inwards migration. The case for Saturn and Jupiter mass planets in NSG discs was investigated by Lin & Papaloizou (2010). In their simulations, gap edges become unstable to vortex modes leading to vortex mergers on dynamical time-scales, in turn resulting in a large-scale vortex circulating at either gap edge. Upon approaching the planet, the inner vortex can execute a horseshoe turn and move outwards across the gap, gaining angular momentum. Thus, the planet loses angular momentum and is scattered inwards.

For comparison purposes, we repeated simulations of Lin & Papaloizou with self-gravity. These correspond to inviscid discs with initially uniform surface density. Fig. 19 shows the orbital migration of a Saturn mass planet in discs with total mass $M_d = 0.035M_*$ and $0.025M_*$. In the $M_d = 0.035M_*$ disc with self-gravity neglected, vortex-induced migration occurs at $t = 60P_0$, whereas in the SG case it is delayed to $t = 85P_0$. The delay is consistent with both the stabilization of low m modes, and the slower vortex merging

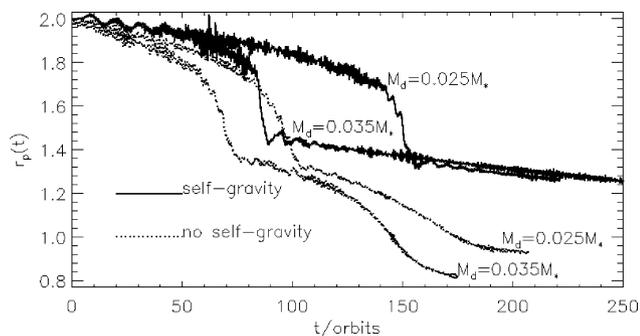


Figure 19. Vortex-induced migration with (solid) and without (dotted) self-gravity.

induced by self-gravity. Furthermore, there could be gravitational influence of the outer gap edge vortices on the inner gap edge.

Indeed, although somewhat inhibited,² we found the formation of an azimuthally extended coherent vortex at the inner gap edge when self-gravity was included, just as in the NSG case. However, it takes a longer time for the inner vortex to build up, disrupt the co-orbital region and flow across the gap, therefore induced rapid migration is delayed. For $M_d = 0.025M_*$ it is delayed by $\simeq 50P_0$, or twice of the higher mass disc, indicating stabilization is more effective for the lower disc mass. However, the extent of rapid migration is unaffected by self-gravity. Note also the increased oscillations when self-gravity is included. This is because of the sustained multi-vortex configuration at the outer gap edge (causing large oscillations in disc–planet torques), whereas without self-gravity these vortices would have merged.

After the first scattering event, migration stalls while the planet opens a gap at its orbital radius and vortex formation recurs. In the NSG case, a large inner vortex develops (taking about $15P_0$), and each time it passes by the planet some vortex material splits off and flows across the planet’s orbital radius, leading to inward planet migration. With self-gravity included, the formation of a single vortex takes longer (about $35P_0$ in the higher mass disc) and it is narrower in radial extent than the NSG case. We found that in this case, when the thinner vortex passes by the planet, little vortex material splits off from the main vortex and flows across the gap. This may be due to the self-gravity of the vortex. Hence, there is a longer stalling period when self-gravity is included. Thus the net effect of self-gravity is to slow the migration in this example.

For the setup of Lin & Papaloizou (2010), we do not observe a second fast migration episode within the simulation time-scale, but it may eventually occur. The total practical simulation time of a few hundred orbits is still very short compared to disc lifetimes. However, for the $Q_0 = 4$ disc model used in previous parts of this work, two episodes of rapid migration occur. Self-gravity does not change the physical nature of vortex-induced migration, but we comment that for discs of even larger mass where vortex formation no longer occurs, being replaced by global spiral instabilities, the migration may be even faster than in the NSG case (Lin & Papaloizou 2011).

A detailed numerical study of the effect of self-gravity on vortex-induced migration and its consequences for disc structure will be deferred to a future paper, but the simple experiment described here indicates that episodes of fast migration punctuated by periods of slower migration are still expected as a consequence of vortex-induced migration.

8 CONCLUSIONS

We have studied the effect of self-gravity on vortex formation and evolution in protoplanetary discs. We specifically considered vortex production through instability associated with edges of surface density dips or gaps opened by a giant planet. It was shown analytically that vortex-forming modes are stabilized by self-gravity through its effect on the linear mode when the background remains fixed. This aspect has been confirmed by linear calculations, which also showed that self-gravity has a destabilizing effect through its effect on the background state that is set up by the perturbing action of an embedded planet. Linear calculations showed that the vortex-forming modes with the highest growth rate shift to higher

² This is also seen in disc–planet simulations in previous sections; the inner gap edge is relatively more stable than outer gap edge.

m with increasing disc mass or equivalently decreasing Toomre Q value.

We performed hydrodynamic simulations that compare simulations with and without self-gravity for a range of disc masses. It was found that more vortices form as the disc mass increased and minimum Q value decreased in accordance with linear calculations. However, for sufficiently strong self-gravity, the vortex modes are suppressed. In this case, global spiral modes develop instead. Self-gravity was found to delay vortex merging. In addition, multiple-vortex configurations can be sustained for longer with increasing disc mass, allowing vortices to evolve individually.

The nature of post-merging vortices is also affected by self-gravity. With weak self-gravity (i.e. $M_d \leq 0.024M_*$), a single vortex extended in azimuth forms and circulates at the outer gap edge. However, for $M_d = 0.031M_*$, the final configuration is a vortex pair at the outer gap edge, each localized in azimuth. In this case, a vortex, despite containing on the order of $20M_{\oplus}$, can be gravitationally bound. The effect of their gravitational influence on the rest of the disc is to redistribute mass radially, whereas for lower disc masses, redistribution is restricted to the same azimuth.

We note that the internal flow in these SG localized vortices adjusts so that they are not destroyed by the background shear, taking on a structure similar to that of a Kida vortex (Kida 1981). We found such vortices form in discs much less massive than that required for direct disc fragmentation. Classical gravitational instability would produce clumps of much larger mass than vortices in our models and would have difficulty surviving against shear beyond a few orbits. However, higher resolution calculations in three dimensions that relax the assumption of a locally isothermal equation of state are needed to assess the viability of the self-gravitational collapse of a vortex.

We performed supplementary simulations of Kida-like vortices to understand the effect of self-gravity on inter-vortex interactions in a simpler setting where it was not contaminated by the presence of the embedded planet. The effectiveness of merging as a function of disc mass may be understood from existing descriptions of co-orbital horseshoe dynamics. In essence, it appears that pressure forces do not play a role, so that in an SG disc, vortex pairs behave like co-orbital planets and as a consequence there exists a minimum inter-vortex distance. Merging is then avoided if this minimal separation is still larger than a critical separation, below which vortex merging occurs.

Finally, we briefly considered the consequence of self-gravity on vortex-induced migration. With self-gravity, the resistance to forming a single large vortex, which in NSG simulations is responsible for scattering the planet inwards, results in vortex-induced migration being delayed. The vortices are less effective in scattering the planet, because they are smaller and do not disrupt the co-orbital region as significantly as their NSG counterparts. Thus, it can be said that in the regime of disc masses where vortices form and are significantly affected by self-gravity, vortex-induced migration is slowed down.

8.1 Outlook and caveats

A possible issue in this work is that the introduction of a planet over short time-scales of a few orbits may favour instability. However, de Val-Borro et al. (2007) have considered NSG discs with a pre-defined gap profile and found similar growth rates for unstable modes as models without an initial gap. In addition, our linear calculations based on gap profiles attained in a quasi-steady state are in good agreement with simulations. This indicates the instability

operates on such attained profiles and is not significantly affected by how the gap is formed.

For the most part planetary migration has been neglected in this work. The role of the embedded planet was to create a structured background state from which vortices develop. Hence, we expect our findings on the effect of self-gravity to be applicable to other types of structured features in a protoplanetary disc that could support vortex-forming instabilities. Such a possibility is the boundary region between a dead zone and active region of the disc (Lyra et al. 2009).

Although migration was briefly considered, we note that there are many technical issues associated with a numerical study. These include the treatment of the Roche lobe and the time and the nature of the release of the planet. The issue of the effects of the choice of softening length and disc viscosity should also be investigated. A detailed parameter survey is beyond the scope of this paper, but we hope to investigate migration in more detail in the future.

Vortices can trap dust particles due to their association with pressure maxima. We have seen that increasing self-gravity leads to vortices of stronger density contrast (Fig. 13). We then anticipate SG vortices to be more effective at collecting solid particles. It would be interesting to consider a two-component gas and dust disc model to investigate the specific role of self-gravity on dust trapping.

Another issue is that we adopted the two-dimensional approximation. Clearly our studies should be extended to three dimensions. It is known that Kida vortices with aspect ratio $\lesssim 4$ are strongly unstable to elliptic instability in three-dimensional, unstructured non-self-gravitating discs (Lesur & Papaloizou 2009). However, results may change when vortices are produced by an instability arising from a background structure and self-gravity is important. These should be investigated in the future. We point out that recently Meheut et al. (2010) demonstrated vortex formation in three-dimensional simulations without self-gravity via the vortex-forming instability. Because the importance of self-gravity is sensitive to the disc thermodynamics and equation of state, it would be desirable to incorporate a realistic energy equation in future studies.

ACKNOWLEDGMENTS

MKL acknowledges support from St John's College, Cambridge, the Isaac Newton Trust and an Overseas Research Award.

REFERENCES

- Armitage P. J., Hansen B. M. S., 1999, *Nat*, 402, 633
 Baruteau C., Masset F., 2008, *ApJ*, 678, 483
 de Val-Borro M. et al., 2006, *MNRAS*, 370, 529
 de Val-Borro M., Artymowicz P., D'Angelo G., Peplinski A., 2007, *A&A*, 471, 1043
 Godon P., 1996, *MNRAS*, 282, 1107
 Goldreich P., Tremaine S., 1980, *ApJ*, 241, 425
 Kida S., 1981, *J. Phys. Soc. Japan*, 50, 3517
 Koller J., Li H., Lin D. N. C., 2003, *ApJ*, 596, L91
 Lesur G., Papaloizou J. C. B., 2009, *A&A*, 498, 1
 Li H., Finn J. M., Lovelace R. V. E., Colgate S. A., 2000, *ApJ*, 533, 1023
 Li H., Colgate S. A., Wendroff B., Liska R., 2001, *ApJ*, 551, 874
 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, 1003
 Li H., Lubow S. H., Li S., Lin D. N. C., 2009, *ApJ*, 690, L52
 Lin D. N. C., Papaloizou J., 1986, *ApJ*, 309, 846
 Lin M., Papaloizou J. C. B., 2010, *MNRAS*, 405, 1473
 Lin M., Papaloizou J. C. B., 2011, *MNRAS*, in press (doi:10.1111/j.1365-2966.2011.18797.x) (this issue)

- Lovelace R. V. E., Li H., Colgate S. A., Nelson A. F., 1999, *ApJ*, 513, 805
 Lyra W., Johansen A., Zsom A., Klahr H., Piskunov N., 2009, *A&A*, 497, 869
 Marcy G., Butler R. P., Fischer D., Vogt S., Wright J. T., Tinney C. G., Jones H. R. A., 2005, *Progress Theor. Phys. Suppl.*, 158, 24
 Masset F., 2000, *A&AS*, 141, 165
 Masset F. S., 2002, *A&A*, 387, 605
 Mayor M., Queloz D., 1995, *Nat*, 378, 355
 Meheut H., Casse F., Varniere P., Tagger M., 2010, *A&A*, 516, A31
 Melander M. V., Zabusky N. J., McWilliams J. C., 1988, *J. Fluid Mech.*, 195, 303
 Meschiari S., Laughlin G., 2008, *ApJ*, 679, L135
 Murray C. D., Dermott S. F., 2000, *Solar System Dynamics*. Cambridge Univ. Press, Cambridge
 Nelson A. F., Benz W., 2003a, *ApJ*, 589, 556
 Nelson A. F., Benz W., 2003b, *ApJ*, 589, 578
 Ou S., Ji J., Liu L., Peng X., 2007, *ApJ*, 667, 1220
 Paardekooper S.-J., Papaloizou J. C. B., 2009, *MNRAS*, 394, 2297
 Paardekooper S., Lesur G., Papaloizou J. C. B., 2010, *ApJ*, 725, 146
 Papaloizou J. C. B., Lin D. N. C., 1989, *ApJ*, 344, 645
 Papaloizou J. C. B., Pringle J. E., 1985, *MNRAS*, 213, 799
 Papaloizou J. C. B., Pringle J. E., 1987, *MNRAS*, 225, 267
 Papaloizou J. C., Savonijie G. J., 1991, *MNRAS*, 248, 353
 Papaloizou J. C. B., Nelson R. P., Kley W., Masset F. S., Artymowicz P., 2007, in Reipurth B., Jewitt D., Reid K., eds, *Protostars and Planets V*. Univ. Arizona Press, Tucson, p. 655
 Stone J. M., Norman M. L., 1992, *ApJS*, 80, 753
 Udry S., Santos N. C., 2007, *ARA&A*, 45, 397
 Yu C., Li H., Li S., Lubow S. H., Lin D. N. C., 2010, *ApJ*, 712, 198
 Zabusky N. J., Hughes M. H., Roberts K. V., 1979, *J. Comput. Phys.*, 30, 96
 Zhang H., Yuan C., Lin D. N. C., Yen D. C. C., 2008, *ApJ*, 676, 639

APPENDIX A: PARAMETRIZATION OF DISC MODELS FOR THE VORTEX INSTABILITY

Disc models used to study the vortex instability are labelled by Q_0 , which corresponds to values of Q_p and disc-to-star mass ratio given in Table A1.

APPENDIX B: ARTIFICIAL VORTICES IN AN ACCRETION DISC

The artificial Kida-like vortices used in Section 6 are set up as follows. Consider a small patch of the disc, whose centre (r_p, φ_p) rotates at angular speed Ω_p about the primary and set up local Cartesian coordinates $x = r_p(\varphi - \varphi_p)$, $y = r_p - r$. In the (x, y)

Table A1. Relationship between the initial Keplerian Toomre stability parameter at the outer boundary, Q_0 , at the planet's initial orbital radius, Q_p , and the disc-to-star mass ratio for disc models used in disc-planet interactions.

Q_0	Q_p	M_d/M_*
1.5	2.62	0.063
2.0	3.49	0.047
2.5	4.36	0.038
3.0	5.23	0.031
3.5	6.11	0.027
4.0	6.98	0.024
8.0	14.0	0.012

frame, there exists a Kida vortex solution for incompressible flow, whose velocity field (u_x, u_y) is

$$u_x = \frac{3\Omega_p \zeta y}{2(\zeta - 1)}, \quad u_y = -\frac{3\Omega_p x}{2\zeta(\zeta - 1)} \quad (\text{B1})$$

inside the vortex core. The ratio of the vortex semimajor to semiminor axis being $\zeta = a/b$ is a free parameter. This velocity field is such that the vorticity ω is constant in the rotating frame. The elliptical boundary of the vortex is defined such that $\omega = -3\Omega_p(1 + \zeta^2)/[2\zeta(\zeta - 1)] = \omega_v - 3\Omega_p/2$ inside the boundary of the vortex and $\omega = -3\Omega_p/2$ outside. The quantity $\omega_v = -3\Omega_p(1 + \zeta)/[2\zeta(\zeta - 1)]$ is then the vorticity of the vortex core relative to the background. Being negative, this corresponds to an anticyclonic vortex. In order to introduce perturbations corresponding to Kida vortices, we impose perturbations $\delta u_r \equiv -v_y$, $\delta u_\varphi \equiv v_x$ inside a specified elliptical boundary with an exponential decay outside. The boundary is fixed by specifying $\zeta = 8$ and its semimajor axis $a = H(r_p)$, where the reference radius is $r_p = 1$.

APPENDIX C: MUTUAL HORSESHOE TURNS IN THE SHEARING SHEET

We describe the gravitational interactions between two vortices in the shearing sheet. It is assumed that they behave like point masses and that pressure forces may be neglected. We indicate below why this is a reasonable assumption.

Consider a local Cartesian coordinate system (x, y) that corotates with a small patch of fluid with angular velocity Ω about the primary, at a distance r_p . We have $x = r - r_p$, $y = r_p(\varphi - \Omega t)$. Let (x_j, y_j) denote the coordinates of the centroid of the j th vortex and m_j be its mass ($j = 1, 2$). Defining $X \equiv x_2 - x_1$, $Y \equiv y_2 - y_1$ and $\mathcal{M} \equiv m_2 + m_1$, the equations of motion give

$$\ddot{X} - 2\Omega\dot{Y} = 3\Omega^2 X - \frac{G\mathcal{M}X}{R^3}, \quad (\text{C1})$$

$$\ddot{Y} + 2\Omega\dot{X} = -\frac{G\mathcal{M}Y}{R^3}, \quad (\text{C2})$$

where $R^2 \equiv X^2 + Y^2$. These equations imply the constancy of the Jacobi constant

$$J \equiv \frac{1}{2}(\dot{Y}^2 + \dot{X}^2) - \frac{3}{2}\Omega^2 X^2 - \frac{G\mathcal{M}}{R}. \quad (\text{C3})$$

Let the initial conditions be $X = X_0$, $Y = \infty$, $\dot{X} = 0$, $\dot{Y} = -3\Omega X_0/2$. We assume the point of closest approach occurs when $X = \dot{Y} = 0$. Equating J at the initial time and at the time of closest approach, we obtain

$$-\frac{3}{8}\Omega^2 X_0^2 = \frac{1}{2}\dot{X}^2 - \frac{G\mathcal{M}}{Y} \quad (\text{C4})$$

at the time of closest approach. Since the vortices are then at minimum separation, $\dot{Y} > 0$. The y component of the equation of motion then implies

$$\frac{1}{2}\dot{X}^2 > \frac{1}{8} \left(\frac{G\mathcal{M}}{\Omega Y^2} \right)^2. \quad (\text{C5})$$

Substituting \dot{X} from (C4) the inequality becomes

$$\frac{3}{8}\hat{X}_0^2 < \frac{q}{\hat{Y}} - \frac{q^2}{8\hat{Y}^4} \quad (\text{C6})$$

at minimum separation, where $\hat{X}_0 = X_0/r_p$, $\hat{Y} = Y/r_p$, $q = \mathcal{M}/M_*$, and we have assumed $\Omega^2 = GM_*/r_p^3$.

Equation (C6) is useful for the case there vortices are just able to undergo U-turns. For fixed q , the function

$$f(\hat{Y}; q) = \frac{q}{\hat{Y}} - \frac{q^2}{8\hat{Y}^4} \quad (\text{C7})$$

has a maximum value at $\hat{Y} = (q/2)^{1/3}$, corresponding to the maximum conceivable initial separation $X_0 = X_s$, where

$$X_s = 2^{2/3} q^{1/3} r_p. \quad (\text{C8})$$

If $X_0 > X_s$ then equation (C6) cannot be satisfied and there can be no horseshoe turns. For initial separations $X_0 < X_s$, (C6) implies that the minimal inter-vortex distance must exceed $q^{1/3}/2$ (so that $f > 0$). Now for sufficiently large q , $q^{1/3}/2$ will be larger than the critical separation for merging, so merging is avoided during the encounter.

It is interesting to compare equation (C8) to the estimate of the horseshoe half-width x_s of Paardekooper & Papaloizou (2009). They found $x_s = 1.68(q/h)^{1/2} r_p$ based on hydrodynamic simulations for

low-mass planets. Equating x_s and X_s with $h = 0.05$, we find $q = 8.9 \times 10^{-5}$. This should give the minimum q for which pressure effects could be ignored. Inserting this value in Murray & Dermott (2000) model of co-orbital satellites gives a minimal separation of 0.58 (see the estimate in Section 6.3), close to simulation results. Hence, if the vortex-pair interaction is purely gravitational, a single vortex behaves in a similar way to ~ 15 Earth masses, i.e. a low-mass protoplanet.

Considering a vortex size of order H , the vortex-to-star mass ratio is $q \sim \pi H^2 \Sigma / M_* \simeq h^3 / Q$. For an SG vortex where $Q \sim 1$, we have $q \simeq h^3 = 1.25 \times 10^{-4}$ for $h = 0.05$, slightly exceeding the threshold value above. Hence, we expect the pressureless treatment of SG vortex–vortex interactions to be acceptable for the purpose of explaining the resisted merging of SG vortices.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.