Lecture notes on the formation of nuclear rings in barred galaxies

Mattia Sormani
University of Insubria

September 17, 2024

Contents

These are my lecture notes for the HERA24 Conference of the European Astronomical Society "The Nature and the Dynamics of Structures Observed in Galactic Disks" held at the Academy of Athens, Greece, September 15-20 2024 (http://astro.academyofathens.gr/hera24.html). I take full responsibility of any mistakes that you might find. Should you have any questions or comments please email mattiacarlo.sormani@gmail.com.

1 Introduction

Nuclear rings are remarkable structures found in the central regions of many barred galaxies (Fig. 1). They have typical radii of few hundred of pc and are often sites of vigorous star formation. They are promising reservoirs of gas for the feeding of central black holes in active galactic nuclei. Our own Galaxy, the Milky Way, is a barred galaxy and has a nuclear ring known as the "central molecular zone".

The goal of these notes is to provide a self-contained account of the theory for the formation of nuclear rings that we developed in [4]. The dynamical process by which nuclear rings form turns out to be a surprisingly difficult problem in disk dynamics. I will tackle this problem in an idealised form, ignoring the physical effects that I believe are not essential.

2 Defining the problem

Nuclear rings are *easy* to form in simulations. This has been well-known since the 1990s (see Fig. 2). The morphological similarity between the rings obtained in simulations and real galaxies is striking (compare Fig. 1 with Fig. 2). Thus there is little doubt that the simulations capture the essential physics involved. However, watching rings forming in simulations does not necessarily mean that we understand the dynamical mechanism by which they form. The goal of these notes is to understand the ring formation mechanism at a more fundamental level.

The simplest simulations, such as those in Fig. 2, demonstrate that nuclear rings form even if we ignore quite a lot of physics: the gas self-gravity, magnetic fields, the vertical structure of the gas, star formation and their associated feedback (e.g., supernova explosions, stellar winds), and the multi-phase nature of the interstellar medium are all absent in these simulations. All is really needed is gas flowing in a non-axisymmetric barred gravitational potential, and a nuclear ring will spontaneously form.

The equations of motion that govern the gas dynamics in these simple simula-

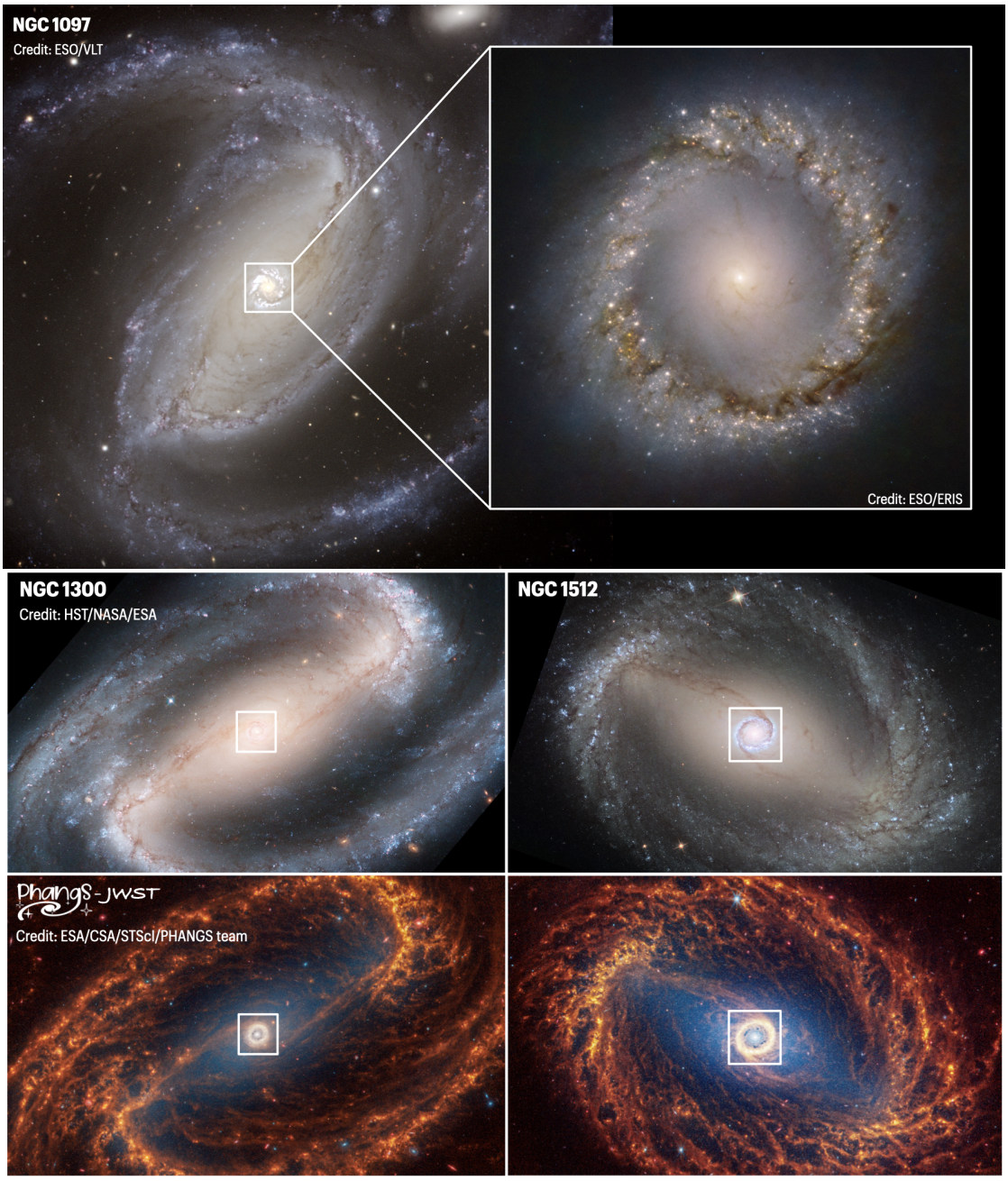

Figure 1: Examples of barred galaxies with nuclear rings. The nuclear ring in each galaxy is marked by a white square.

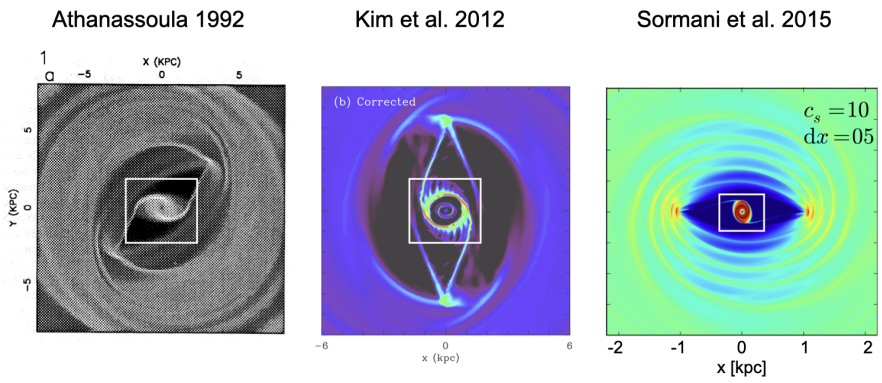

Figure 2: Examples of nuclear rings in simple hydrodynamical simulations from the literature. The nuclear rings are marked by a white square.

tions are:

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0\,, \tag{1}$$

$$\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{\nabla P}{\rho} - \nabla \Phi\,, \tag{2}$$

where $\rho(\mathbf{x}, t)$ is the gas density, $\mathbf{v}(\mathbf{x}, t)$ is the gas velocity, and $P(\mathbf{x}, t)$ is the pressure. There are just the familiar continuity and Euler equations. The pressure is assumed to be related to the density by a simple isothermal equation of state, $P = c_{\rm s}^2 \rho$, where $c_{\rm s}$ is a constant called the sound speed. The gas is assumed to be 2D (i.e., razor-thin). The quantity $\Phi(\mathbf{x}, t)$ is an external, rigidly-rotating non-axisymmetric barred gravitational potential. This is generated by the stars and dark matter that compose the galaxy. We ignore the contribution of the gas self-gravity as the gas often contributes only a few % to the total mass budget.

Since simulations based on Eqs. (1)-(2) successfully reproduce nuclear rings, it ought to be possible to understand their formation starting from these equations in a more analytic way. This is the goal of the remainder of these notes. The analyses in the next sections will be fairly general, but for the sake of concreteness we will always use a specific model for the external gravitational potential $\Phi(\mathbf{x}, t)$ whenever we need to make a plot or provide numerical values. I hope this will make the notes easier to follow. The next section describes this model and can be skipped by the reader not interested in too much detail.

3 Galaxy model

The purpose of this section it to provide a simple yet reasonably realistic analytic expression for the gravitational potential $\Phi(\mathbf{x}, t)$ of a barred spiral galaxy. We

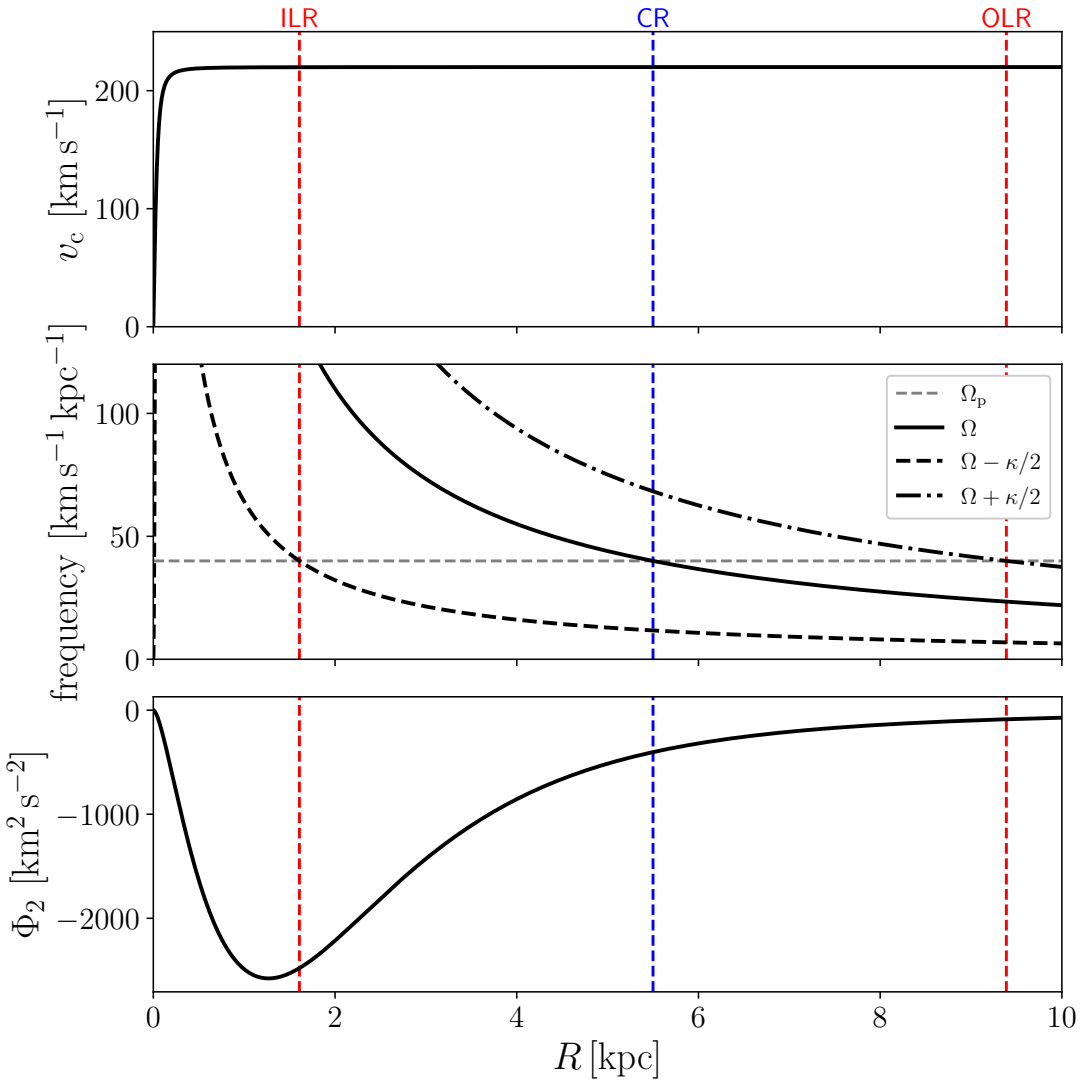

Figure 3: The gravitational potential used in this notes. *Top*: the circular velocity, defined as $v_c = (R\mathrm{d}\Phi_0/\mathrm{d}R)^{1/2}$. *Middle:* the angular velocity Ω and the curves $\Omega \pm \kappa/2$, where $\Omega = v_c/R$ and κ is the epicyclic frequency (Eq. 49). The intersection of these curves with the horizontal line at Ω_p gives the position of the resonances, indicated by vertical dashed lines. The inner Lindblad resonance is at $R_{\rm ILR} = 1.61$ kpc, the corotation resonance at $R = 5.5\,$kpc, and the outer Lindblad resonance at $R_{\rm OLR} = 9.39$ kpc. *Bottom*: the quadrupole (Eq. 6).

require it to include both the axisymmetric part generated by the galactic disc and the dark matter halo and the non-axisymmetric part generated by the bar.

What is the simplest possible bar potential? Bars typically have the shape of a cigar that lies in the plane of the galaxy. The figure of the bar rotates at a rate called the pattern speed $\Omega_{\rm p}$.[1] We are only interested in the potential in the midplane $z = 0$ since we approximate the gas layer as two-dimensional. Any rigidly rotating potential in a plane can be expanded as:

$$\Phi(R, \theta, t) = \Phi_0(R) + \sum_{n=1}^{\infty} \Phi_n(R) \cos\left[n(\theta - \Omega_{\rm p} t) + \phi_n\right] \tag{3}$$

where (R, θ) are polar coordinates in the rotating frame, ϕ_n are constants, and $\Omega_{\rm p}$ is the angular rotation speed. The term $\Phi_0(R)$ is called the monopole, $\Phi_1(R)$ the dipole, $\Phi_2(R)$ the quadrupole, and so on. Bars are bi-symmetric around the Galactic centre [i.e. their stellar density is the same at the four points (x, y), $(x, -y)$, $(-x, y)$ and $(-x, -y)$], so $\Phi_n = 0$ for n odd. N-body simulations of bars show that the quadrupole ($n = 2$) is by far the most significant of the even-n terms, while higher order terms rapidly get smaller. Thus to a good approximation a bar potential can be written as

$$\boxed{\Phi(R, \theta, t) = \Phi_0(R) + \Phi_2(R) \cos\left[2(\theta - \Omega_{\rm p} t)\right]} \tag{4}$$

where without loss of generality we have chosen the origin of time so that $\phi_2 = 0$.

Let us now choose specific forms for $\Phi_0(R)$ and $\Phi_2(R)$. For the monopole, we take the logarithmic potential

$$\Phi_0(R) = \frac{v_0^2}{2} \log\left(R^2 + R_{\rm c}^2\right) , \tag{5}$$

where $v_0 = 220\,{\rm km\,s^{-1}}$ and $R_{\rm c} = 0.05\,{\rm kpc}$. The logarithmic potential is convenient because the rotation curve is rising at small R and is flat at $R \gg R_{\rm c}$, roughly consistent with the rotation curves observed in many disk galaxies (see Fig. 3). For the quadrupole, we take the following potential:[2]

$$\Phi_2(R) = -A(v_0 e)^2 f\left(R/R_q\right) \tag{6}$$

where $A = 0.4$ is a dimensionless parameter that quantifies the bar strength, $e = 2.71[\ldots]$ is Euler's number, $v_0 = 220\,{\rm km\,s^{-1}}$ is the same as in Eq. (5), $R_q = 1.5\,{\rm kpc}$ is a radial scalelength, and f is the following function:

$$f(x) = \frac{3 - e^{-2x}\left(2x^4 + 4x^3 + 6x^2 + 6x + 3\right) + 4x^5 {\rm E}_1(2x)}{20x^3} , \tag{7}$$

[1] Note that the invidual stars that compose the bar rotate at a faster rate than the pattern speed, i.e. they rotate faster than the shape of the figure they collectively generate.

[2] The 3D density distribution that generates this potential is proportional to the Y_2^2 spherical harmonic and drops exponentially in radius with scalelenght $R_q/2$.

where $\mathrm{E}_1(x)$ is the exponential integral function, a special function defined as

$$\mathrm{E}_1(x) = \int_x^\infty \frac{e^{-t}}{t}\,\mathrm{d}t\,. \tag{8}$$

This quadrupole has been shown to reproduce well those generated by N-body exponential bars. Finally, for the pattern speed we take the value $\Omega_\mathrm{p} = 40\,\mathrm{km\,s^{-1}\,kpc^{-1}}$.

To summarise, the gravitational potential used in the rest of the notes is given by Eqs. (4), (5) and (6). These are the only equations from this section that we will need for the remainder of the notes. Figure 3 gives an overview of the basic properties of the potential.

4 Waves in a disc

The goal of this section is to understand the propagation and generation of waves in a gas disc that is described by Eqs. (1)-(2). As we shall see below, this will be important to understand the formation of nuclear rings.

Imagine a 2D axisymmetric rotating gas disc in equilibrium. What happens if we perturb it a little bit by "touching" it here and there? This will generate waves propagating through the disc, in a way similar a plucked cord generates sound waves that propagate through air. Introducing a time-dependent rotating non-axisymmetric barred potential has a similar effect: it stirs the gas disc like a giant galactic spoon and generates waves that propagate through the disc. As we will see below, these waves will be responsible for redistributing the angular momentum and creating the accumulation of gas that we call a nuclear ring. Our task in this section is to understand how waves are generated by the external bar potential and how they propagate on top of the gas disc.

We will limit ourselves to study waves in the linear approximation. In practice, this means that we will linearise the equations of motion around the equilibrium state and neglect quadratic and higher order terms in the perturbed quantities. This approximation is necessary to make the problem analytically tractable and formally requires that the wave amplitude is "small". In reality, waves excited by the quadrupole described in Sect. 3, whose non-axisymmetric force reaches $\sim 5\%$ of the axisymmetric force at some radii, will not be small at all. However, such non-linear waves are very difficult to study analytically. They can be studied using a simulation, but as described in Sect. 2 we then need to face the question of how to extract basic physical understanding from it. Here we choose a different approach. We focus on the linear regime aiming to gain insight on the basic dynamical processes. This will provide us with understanding at a more fundamental level, but numerical values should not expected to be accurate when compared to a full non-linear calculation.

This section is structured as follows. In Sect. 4.1 we lay out the starting equations and write them in cylindrical coordinates. In Sect. 4.2 we describe the axisymmetric equilibrium discs that form the background on top of which waves travel. In Sect. 4.3 we linearise the equations of motion around this equilibrium state and reduce the problem of studying general perturbations to that of solving a single second-order ordinary differential equation. In Sect. 4.4 we find approximate solutions to this equation in the absence of a perturbing external potential. These solutions describe "free density waves" propagating in the gas disc, analog to sound waves propagating through air. In Sect. 4.5 we solve the same equation in the presence of a forcing external potential, and study the generation of waves by such external potential.

4.1 Basic equations

Our starting point are the continuity and Euler equations (1)-(2):

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0\,, \tag{9}$$

$$\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{\nabla P}{\rho} - \nabla \Phi\,. \tag{10}$$

where ρ is the surface density, $\mathbf{v} = v_x \hat{\mathbf{e}}_x + v_y \hat{\mathbf{e}}_y$ is the velocity, P is the pressure, and $\Phi(\mathbf{x}, t)$ is the external gravitational potential given by Eq. (4). Although our stated goal in Sect. 2 is to study an isothermal gas, we consider a slightly more general polytropic equation of state since the calculations will be almost identical:

$$P = K\rho^{\gamma}, \tag{11}$$

where $\gamma \geq 1$ and K is a constant. The isothermal case is recovered as $\gamma = 1$. To simplify the calculations it is convenient to introduce the enthalpy h defined by:

$$\nabla h = \frac{\nabla P}{\rho}\,, \tag{12}$$

substituting (11) into (12) and integrating we find:

$$h = \begin{cases} K\left(\frac{\gamma}{\gamma-1}\right)\rho^{\gamma-1} & \text{if} \quad \gamma > 1\,, \\ K \log \rho & \text{if} \quad \gamma = 1\,. \end{cases} \tag{13}$$

Using (12), the equations of motion (9) and (10) can be expanded in polar coordinates (R, θ) as:

$$\partial_t \rho + \frac{1}{R}\partial_R \left(R\rho v_R\right) + \frac{1}{R}\partial_\theta \left(\rho v_\theta\right) = 0\,, \tag{14}$$

$$\partial_t v_R + \left(v_R\partial_R + \frac{v_\theta}{R}\partial_\theta\right) v_R - \frac{v_\theta^2}{R} = -\partial_R h - \partial_R \Phi\,, \tag{15}$$

$$\partial_t v_\theta + \left(v_R\partial_R + \frac{v_\theta}{R}\partial_\theta\right) v_\theta + \frac{v_R v_\theta}{R} = -\frac{1}{R}\partial_\theta h - \frac{1}{R}\partial_\theta \Phi\,. \tag{16}$$

4.2 Axisymmetric equilibrium state

The equilibrium axisymmetric solutions on top of which waves travel are as follows. Assume $\Phi_2 = 0$ in Eq. (4) and look for steady-state solutions of Eqs. (14)-(16) of the form:

$$\rho = \rho_0(R), \tag{17}$$

$$h = h_0(R), \tag{18}$$

$$\mathbf{v} = \Omega(R) R\, \hat{\mathbf{e}}_\theta. \tag{19}$$

Substituting these into (14)-(16) and assuming steady-state and axisymmetry ($\partial_t = \partial_\theta = 0$), we see that the continuity equation (14) and the azimuthal Euler equation (16) are already satisfied, while the radial Euler equation (15) gives:

$$\Omega^2 R = \frac{\mathrm{d}(h_0 + \Phi_0)}{\mathrm{d}R}. \tag{20}$$

When studying waves in the following sections, h_0, Φ_0 and Ω will be considered prescribed functions of R that satisfy Eq. (20). Note that given $\Phi_0(R)$, there formally exists an equilibrium solution $h_0(R)$ for any arbitrary rotation profile $\Omega(R)$. However not all possible profiles are physical. To avoid instability, the unperturbed state must satisfy the Rayleigh stability criterion, which states that a necessary and sufficient condition for the local axisymmetric stability of an inviscid differentially rotating fluid disc is that the specific angular momentum monotonically increases with R, i.e.[3]

$$\frac{\mathrm{d}(R^2\Omega)}{\mathrm{d}R} > 0\,. \qquad \text{(Rayleigh criterion)} \tag{21}$$

In these notes we will consider two main types of density profiles:

[3]The Rayleigh criterion is equivalent to the condition that the epicyclic frequency is real ($\kappa^2 > 0$, see Eqs. 33 and 49).

1. **Constant density**. This is simply:

$$\rho_0(R) = \bar{\rho} = \text{constant}\,. \tag{22}$$

2. **Truncated disc density**. This is given by

$$\rho_0(R) = \frac{\bar{\rho}}{2}\left[1 - \frac{a}{(1+a^2)^{1/2}}\right], \tag{23}$$

where

$$a = \frac{R - R_{\rm edge}}{\Delta R}\,, \tag{24}$$

This density is roughly constant at $R \ll R_{\rm edge}$, has a relatively sharp transition at $R_{\rm edge}$ during which it drops at a much lower value, and is then roughly constant again at $R > R_{\rm edge}$. Note that the edge cannot be too thin, otherwise it would violate the Rayleigh criterion (21).

The quantity $\bar{\rho}$ in both profiles is a constant that essentially defines the units used for density. Physically meaningful results do not depend on the particular value of this quantity since the equations of motion (9) and (10) are invariant under density rescaling. Without loss of generality, we set $\bar{\rho} = 1$ hereafter.

4.3 General equation of linear disc dynamics

Here we linearise the equations of motion around the equilibrium state of the previous section, and then reduce the problem of studying the most general linear perturbation to solving a second-order ODE. We expand (14)-(16) around the unperturbed state described in Sect. 4.2 as:

$$\rho = \rho_0 + \rho_1\,, \tag{25}$$
$$h = h_0 + h_1\,, \tag{26}$$
$$\mathbf{v} = \mathbf{v}_0 + \mathbf{v}_1\,, \tag{27}$$
$$\Phi = \Phi_0 + \Delta\Phi\,. \tag{28}$$

Here, $\Delta\Phi$ is the non-axisymmetric part of the potential, which we will eventually put equal to the Φ_2 term in Eq. (4) but for the moment we keep general. Substituting equations (25)-(28) into (14)-(16), the zeroth-order terms of the equilibrium solution simplify, and linearising by keeping only first-order terms in the quantities

with subscript 1, we obtain:

$$\frac{D}{Dt}\left(\frac{\rho_1}{\rho_0}\right) + \frac{\mathrm{d}\log(R\rho_0)}{\mathrm{d}R} v_{R1} + (\partial_R v_{R1}) + \frac{1}{R}(\partial_\theta v_{\theta 1}) = 0\,, \tag{29}$$

$$\frac{Dv_{R1}}{Dt} - 2\Omega v_{\theta 1} = -\partial_R\left[h_1 + \Delta\Phi\right]\,, \tag{30}$$

$$\frac{Dv_{\theta 1}}{Dt} + 2Bv_{R1} = -\frac{1}{R}\partial_\theta\left[h_1 + \Delta\Phi\right]\,, \tag{31}$$

where we have defined the convective derivative of the unperturbed state

$$\frac{D}{Dt} = \partial_t + \Omega\partial_\theta\,, \tag{32}$$

and the Oort parameter

$$B(R) = \Omega + \frac{R}{2}\frac{\mathrm{d}\Omega}{\mathrm{d}R}\,. \tag{33}$$

Substituting Eq. (25) into (13) we can relate h_1 to ρ_1:

$$h_1 = c_\mathrm{s}^2\left(\frac{\rho_1}{\rho_0}\right) \tag{34}$$

where we have introduced the sound speed of the unperturbed medium:

$$c_\mathrm{s}^2 = \gamma K \rho_0^{\gamma-1}\,. \tag{35}$$

Eq. (34) is valid for $\gamma \geq 1$ ($\gamma = 1$ corresponds to the isothermal case).

Equations (29)-(31) govern the time evolution of general linear perturbations. Given initial conditions $\rho_1(\mathbf{x}, t=0)$, $v_{R1}(\mathbf{x}, t=0)$, $v_{\theta 1}(\mathbf{x}, t=0)$, these equations completely determine the subsequent evolution. We can simplify these equations further using the superposition principle. The coefficients of the various terms of these equations do not depend on θ. This means that if we Fourier-expand the angular part, different modes will evolve independently without mixing with each other. We can solve for each mode separately and then put together the results at the end if needed. Therefore, without loss of generality, we write:

$$\rho_1 = \tilde{\rho}_1(R)\exp[i(m\theta - \omega t)], \tag{36}$$

$$v_{1R} = \tilde{v}_{1R}(R)\exp[i(m\theta - \omega t)], \tag{37}$$

$$v_{1\theta} = \tilde{v}_{1\theta}(R)\exp[i(m\theta - \omega t)], \tag{38}$$

$$h_1 = \tilde{h}_1(R)\exp[i(m\theta - \omega t)], \tag{39}$$

$$\Delta\Phi = \Delta\tilde{\Phi}(R)\exp[i(m\theta - \omega t)], \tag{40}$$

where $\tilde{\rho}_1$, $\tilde{v}_{1R}$ etc. are complex numbers, and the "physical" quantity is the real part. In these notes, we will only be concerned with $m = 2$, corresponding to the

quadrupole term in Eq. (4). With the substitutions (36)-(40), we have $\partial_t = -i\omega$ and $\partial_\theta = im$. We define the pattern speed of each mode as:

$$\Omega_{\rm p} = \frac{\omega}{m}\,. \tag{41}$$

This is the angular frequency with which each mode appears to rotate, as can be understood by noting that

$$e^{i(m\theta-\omega t)} = e^{im(\theta-\Omega_{\rm p}t)}\,. \tag{42}$$

In these notes, $\Omega_{\rm p}$ will always coincide with the pattern speed of the bar potential (4). Only modes at this frequency and $m = 2$ can be excited by the external potential Φ_2 in the linear approximation.

Hereafter, we drop the $\tilde{}$ symbol in all equations to avoid cluttering. Substituting (36)-(40) into (29)-(31) we obtain:

$$im\,(\Omega-\Omega_{\rm p})\left(\frac{\rho_1}{\rho_0}\right) + \frac{{\rm d}\log(R\rho_0)}{{\rm d}R}v_{R1} + \frac{{\rm d}v_{R1}}{{\rm d}R} + \frac{im}{R}v_{\theta 1} = 0\,, \tag{43}$$

$$im\,(\Omega-\Omega_{\rm p})\,v_{R1} - 2\Omega v_{\theta 1} = -\frac{\rm d}{{\rm d}R}\left[h_1+\Delta\Phi\right]\,, \tag{44}$$

$$im\,(\Omega-\Omega_{\rm p})\,v_{\theta 1} + 2Bv_{R1} = -\frac{im}{R}\left[h_1+\Delta\Phi\right]\,, \tag{45}$$

Isolating v_{R1} and $v_{\theta 1}$ from (44) and (45) we find:

$$v_{R1} = -\frac{im}{D}\left(\frac{2\Omega}{R} + (\Omega-\Omega_{\rm p})\frac{\rm d}{{\rm d}R}\right)\left[h_1+\Delta\Phi\right] \tag{46}$$

$$v_{\theta 1} = \frac{1}{D}\left(\frac{m^2(\Omega-\Omega_{\rm p})}{R} + 2B\frac{\rm d}{{\rm d}R}\right)\left[h_1+\Delta\Phi\right] \tag{47}$$

where we have defined

$$D = \kappa^2 - m^2(\Omega-\Omega_{\rm p})^2, \tag{48}$$

$$\kappa^2 = 4B\Omega \tag{49}$$

The quantity κ is called the **epicyclic frequency**. It is the frequency at which a fluid parcel in circular orbit that is radially displaced will oscillate. The points where $D = 0$ define the **Lindblad resonances**.[4] These are the radii at which the radial oscillation frequency κ is equal to the "forcing frequency" $m(\Omega-\Omega_{\rm p})$ seen by a particle. We call inner Lindblad resonances (ILR) those that occur where

[4]Note that the position $D = 0$ also depends on the unperturbed density profile ρ_0 because of the contribution from the pressure term h_0 to Ω in Eq. (20).

$\Omega > \Omega_\mathrm{p}$, and outer Lindblad resonances (OLR) those that occur where $\Omega < \Omega_\mathrm{p}$. The point where $\Omega = \Omega_\mathrm{p}$ defines the **corotation resonance** (CR). The middle panel in Fig. 3 shows where the resonances occur for the potential described in Sect. 3. Fig. 4 plots κ and D for the same potential.

Now we substitute (46) and (47) into (43) and use (34) to eliminate ρ_1. We obtain an equation in the variable h_1:

$$\frac{\mathrm{d}^2 h_1}{\mathrm{d}R^2} + 2H(R)\frac{\mathrm{d}h_1}{\mathrm{d}R} + W(R)h_1 = F(R) \tag{50}$$

where

$$H(R) = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}R}\left[\log\left(\frac{R\rho_0}{|D|}\right)\right], \tag{51}$$

$$W(R) = C(R) - \frac{D(R)}{c_\mathrm{s}^2}, \tag{52}$$

$$C(R) = \left(\frac{2\Omega}{R(\Omega-\Omega_\mathrm{p})}\right)\frac{\mathrm{d}}{\mathrm{d}R}\left[\log\left(\frac{\rho_0\Omega}{|D|}\right)\right] - \frac{m^2}{R^2}, \tag{53}$$

$$F(R) = -\left\{\frac{\mathrm{d}^2}{\mathrm{d}R^2} + 2H(R)\frac{\mathrm{d}}{\mathrm{d}R} + C(R)\right\}\Delta\Phi(R). \tag{54}$$

Eq. (50) coincides with Eq. (13) of [3]. It is a second order ordinary differential equation with non-constant coefficients $H(R)$ and $W(R)$. The term $F(R)$ is a forcing term. Note that $H(R)$ and $W(R)$ diverge where $(\Omega - \Omega_\mathrm{p}) = 0$ and where $D = 0$, i.e. at the corotation and Lindblad resonances.

In order to eliminate the first order derivative from Eq. (50), it is convenient to define a new variable g_1 such that

$$g_1(R) = \exp\left[\int^R H(s)\,\mathrm{d}s\right] h_1(R) = \left(\frac{R\rho_0}{|D|}\right)^{1/2} h_1\,. \tag{55}$$

Substituting Eq. (55) into Eq. (50), one finds

$$\boxed{\frac{\mathrm{d}^2 g_1}{\mathrm{d}R^2} + K^2(R)g_1 = Q(R)}, \tag{56}$$

where

$$K^2(R) = W - H^2 - \frac{\mathrm{d}H}{\mathrm{d}R} \tag{57}$$

$$Q(R) = \left(\frac{R\rho_0}{|D|}\right)^{1/2} F(R)\,. \tag{58}$$

Eq. (56) is the fundamental equation that governs linear modes in the disc. The only approximation made to derive it is the linear approximation. A good portion of the remainder of these notes will be spent finding approximate solutions of this equation. We will see below (Eq. 69) that there is a conserved quantity associated with this equation, which physically is interpreted as the flux of angular momentum. To follow the calculations in the following sections more intuitively, it might be useful to note that Eq. (56) is like that of a forced harmonic oscillator, $m\ddot{x} + k^2(t)x = q(t)$ with a time-dependent spring constant $k(t)$ and a time-dependent external force $q(t)$ (in this analogy, t replaces R, m is the mass attached to the spring).

There are two types of points where Eq. (56) requires special attention:

- *Turning points.* These are the points $R_\star$ where $K^2(R_\star) = 0$. At these points, K^2 changes sign and the character of the solutions changes from oscillatory to exponential.

- *Singular points.* These are points where $K(R)$ diverges. As can be seen from Eqs. (57) and (51)-(54), this happens at the Lindblad and Corotation resonances.

Figure 5 shows the coefficients of Eq. (56) for the galaxy model in Sect. 3, for a uniform and a truncated disc profile and for sound speed $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$. In the region relevant for the formation of nuclear rings ($R < 2\,\mathrm{kpc}$) there is one turning point $R_\star$ and one singular point at R_ILR, with $R_\star < R_\mathrm{ILR}$. As we shall discuss below, $R_\star$ is where the medium becomes absorbing and leading waves incident from $R < R_\star$ are reflected into trailing waves that subsequently travel inwards. The position of $R_\star$ depends on both the sound speed c_s and the shape of the unperturbed density profile $\rho_0(R)$. In the limit $c_\mathrm{s} \to 0$ we have $R_\star \to R_\mathrm{ILR}$. However, for a finite value of the sound speed, the two points are distinct.

4.4 Free waves

Free perturbations in the fluid disc are described by Eq. (56) with $Q = 0$ (no external forcing). This is a second order differential equation, so it has two linearly independent solutions. We can think of these two solutions as are a more complicated version of the two sound waves at a given frequency that can propagate in a uniform medium.

In general, Eq. (56) is hard to solve exactly. Approximate solutions can be found using the WKB method. Appendix A presents a self-contained exposition of this method. The general WKB solution for an equation of type (56) is given

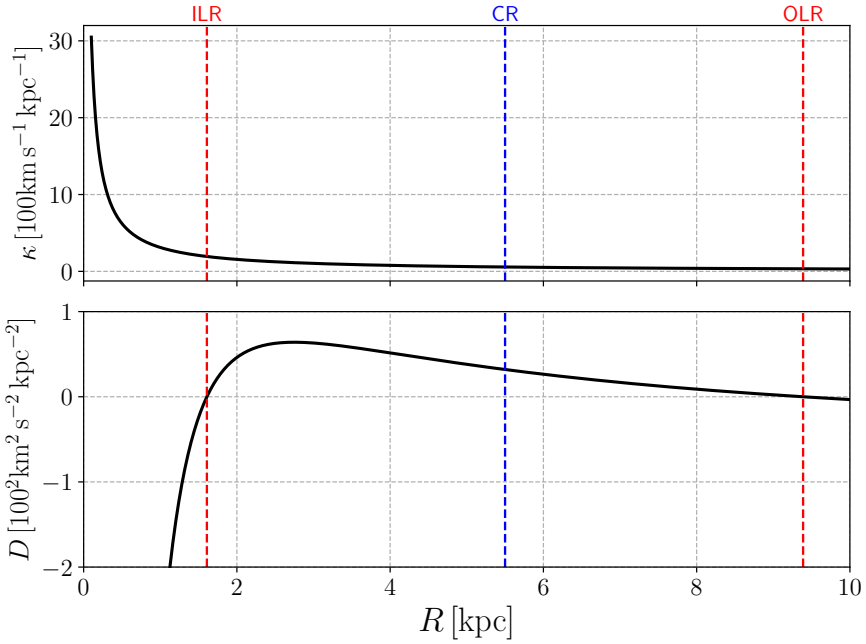

Figure 4: *Top*: the epicyclic frequency κ given by Eq. (49) for the gravitational potential described in Sect. 3. *Bottom*: the quantity D defined by Eq. (48). These plots assume a constant density profile ρ_0.

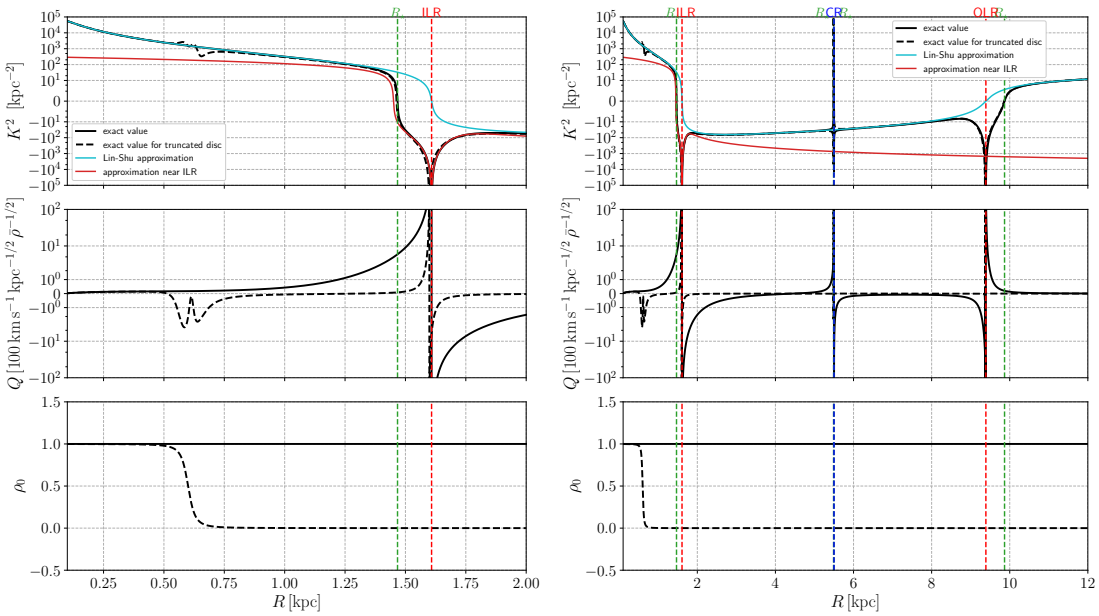

Figure 5: *Top*: the wavenumber $K(R)$ i(Eq. 56). Full black line: for constant density profile. Dashed black line: for a truncated disc profile (Eq. 23). Cyan line: Lin-Shu approximation (Eq. 63). Red line: approximation near the ILR (Eq. 90). *Middle*: forcing term $Q(R)$ (Eq. 56). *Bottom*: the uniform ($\rho_0 = 1$) and truncated disc (Eq. 23 with $R_{\rm edge} = 0.6\,{\rm kpc}$ and $\Delta R = 0.03\,{\rm kpc}$) density profiles assumed in this figure. The red and blue dashed lines mark the resonances. The green dashed lines mark the turning points $R_\star$ (see Sect. **??**). The left panels are a zoom of the right panel for $R < 2\,{\rm kpc}$. All panels assume a sound speed $c_{\rm s} = 10\,{\rm km\,s^{-1}}$.

by Eq. (147), which adapted to the present notation reads:

$$g_1(R) = \frac{C_1}{\sqrt{K(R)}} \exp\left[i \int_{R_0}^{R} K(s)\mathrm{d}s\right] + \frac{C_2}{\sqrt{K(R)}} \exp\left[-i \int_{R_0}^{R} K(s)\mathrm{d}s\right] , \qquad (59)$$

where C_1 and C_2 are arbitrary complex constants and R_0 is an arbitrary radius.

The two terms on the right-hand side of Eq. (59) represent two waves travelling in opposite directions, similar to the two sound waves that are possible in a uniform medium at a given frequency (to see this, reattach the $e^{im\theta - i\omega t}$). The quantity K is a radius-dependent wavenumber. When $K^2 > 0$, the solution (59) has oscillatory character and waves can travel, when $K^2 < 0$ it has exponential character and the medium is absorbing. Fig. 5 shows that travelling waves can exist essentially only inside the ILR or outside the OLR.

Equation (57) implicitly contains ω, and therefore for fixed R this expression can be seen as a **dispersion relation** $K = K(\omega)$ (Fig. 6).

The two waves defined by Eq. (59) typically have a spiral shape. This can be seen reattaching the $e^{im\theta}$ term to (59). Taking for example the C_1 wave, we see that the location of the maxima of g_1 (and therefore of density ρ_1) as a function of θ for fixed R and t occur along lines

$$\int_{R_0}^{R} K(s)\mathrm{d}s + m\theta = \text{constant} \qquad (\text{mod } 2\pi) \qquad (60)$$

where n is an integer. The curve defined by Eq. (60) in the plane (R, θ) typically traces a spiral shape (try for example K =constant). The sign of K determines whether the wave is leading or trailing. In our case, the C_1 wave is **leading**, while the C_2 wave is **trailing**. With a little bit of geometry (see for example Chapter 6 of [1]) one can find that the pitch angle (i.e., the angle between the tangent to the arm and the circle R =constant) is:

$$\cot i = \frac{|K|R}{m} . \qquad (61)$$

What is the range of validity of the WKB approximation? The WKB approximation is expected to work well when the following parameter is small (see Eq. 139):

$$\epsilon = \left| \frac{\mathrm{d}K/\mathrm{d}R}{K^2} \right| , \qquad (62)$$

Thus, the WKB approximation works when K varies 'slowly'. This means that Eq. (59) represent rapidly oscillating waves whose amplitude is slowly modulated by the term $K^{-1/2}$. Figure 7 plots the parameter (62) for our galaxy model. It becomes large near the turning points and the resonances. At these points the

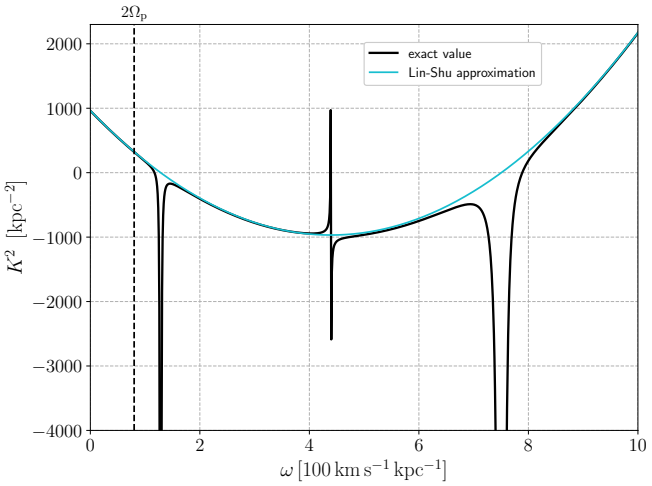

Figure 6: The dispersion relation $K(\omega)$ at $R = 1$ kpc for our galaxy model. The vertical dashed line indicates the frequency $\omega = 2\Omega_p$ of the bar potential. The frequencies at which K diverges are those for which the ILR, CR and OLR would be at $R = 1$ kpc (resonances move inwards as frequency is increased). We assumed $c_s = 10\,\mathrm{km\,s^{-1}}$ and constant unperturbed density $\rho_0(R) = 1$.

WKB approximation fails. The WKB approximation also fails near sharp edges of ρ_0, because $\mathrm{d}K/\mathrm{d}R$ becomes large. Figure 8 compares an exact (numerically calculated) and a WKB solution of Eq. (56), and shows that the WKB approximation works exceptionally well at $R < R_\star$, but breaks down near $R = R_\star$.

The WKB approximation used here is not completely equivalent to the more well-known Lin-Shu approximation. The Lin-Shu dispersion relation in the absence of self-gravity ($G = 0$) is given by (Eq. 6.55 of [1]):

$$K^2_{\text{Lin-Shu}} = -\frac{D}{c_s^2}\,, \tag{63}$$

where D is given by Eq. (48). The top of Fig. 5 compares the Lin-Shu dispersion relation (cyan line) with the dispersion relation given by Eq. (57). The two differ mainly around the resonances. Differences become larger for larger values of the sound speed c_s, because the Lin-Shu dispersion relation assumes very small c_s, while the dispersion relation (57) takes into account the effect of finite sound speed. Indeed, in the limit of vanishing sound speed we recover the Lin-Shu dispersion relation from our relation (57). This can be shown by noting that in this limit $W(R) \simeq -D/c_s^2$ (Eq. 52), while $H^2 \ll W$ and $\mathrm{d}H/\mathrm{d}R \ll W$ (Eq. 57).

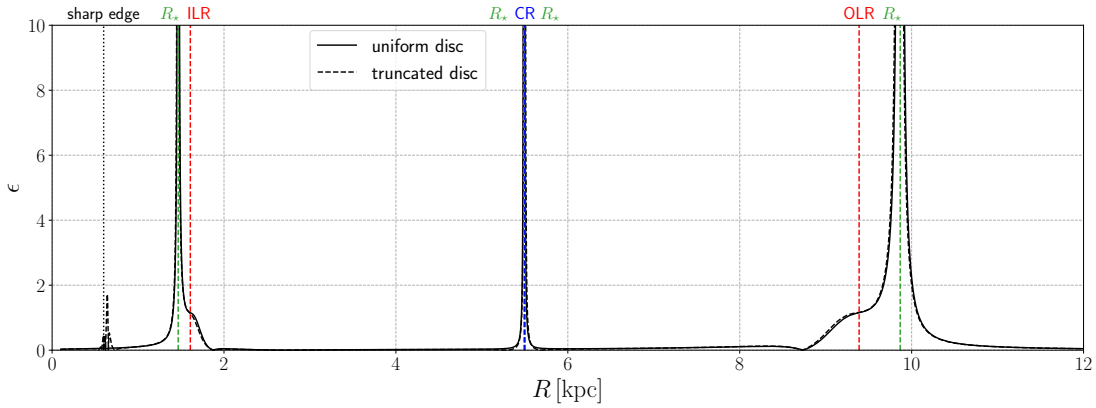

Figure 7: The "small parameter" of the WKB approximation for a uniform and truncated disc (Eq. 62). The parameters used (sound speed, truncated profile parameters, gravitational potential) are the same as in Fig. 5.

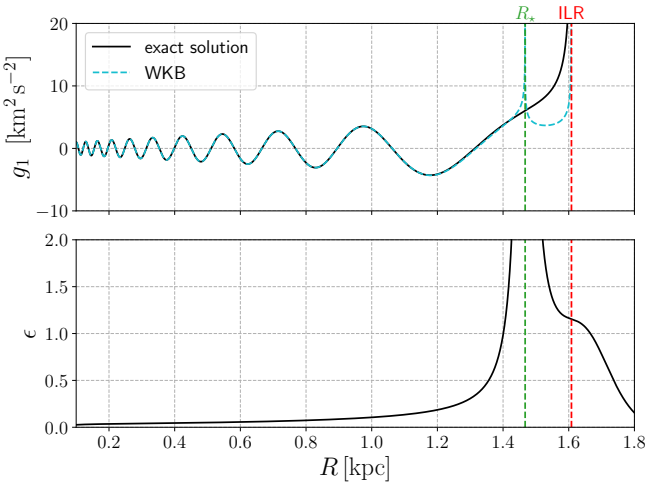

Figure 8: *Top*: comparison between the exact solution of Eq. (56) and the WKB approximation. The full black line shows the solution obtained numerically integrating Eq. (56) from $R = 0.1\,\mathrm{kpc}$ with initial conditions $g_1 = 1$, $\mathrm{d}g_1/\mathrm{d}R = 0$ (full black line). The cyan dashed line shows the WKB approximation (59). We assumed $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$ and constant unperturbed density $\rho_0(R) = 1$. *Bottom*: the "small" parameter of the WKB approximation (Eq. 62). At $R_\star$ it diverges. The WKB approximation works well only at $R < R_\star$.

4.4.1 Group velocity

The crests of the waves described by Eq. (59) travel at a **phase velocity** (to see this, reattach the $e^{i\omega t}$), defined by

$$v_p = \frac{\omega}{K} \tag{64}$$

However, it is well known if waves are dispersive (i.e., ω/K is not constant as a function of ω at a given R, as in our case, see Fig. 6), then the envelope shape of a wave packet travels with the **group velocity**:

$$v_g = \frac{\partial \omega}{\partial K} \tag{65}$$

The group velocity is also the velocity at which quantities such as energy and angular momentum are transported along the wave. Figure 9 plots the phase and group velocities for $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$. We can see a few things. At $R < R_\mathrm{ILR}$ the group velocity of trailing waves ($C_1 \neq 0$ and $C_2 = 0$) is negative, meaning that these waves travel inward. Leading waves instead travel outwards. The phase and group velocities have opposite signs, but the latter is the 'real' direction of propagation of wave packets. The group velocity loses meaning and becomes imaginary between the ILR and the OLR, when the medium becomes absorbing. At $R > R_\mathrm{OLR}$, trailing (leading) waves propagate outwards (inwards), and the phase and group velocity have the same sign.

The cyan line compares our group velocity with that obtained from the Lin-Shu dispersion relation (63) (see Equation 20 of [3]):

$$c_{\mathrm{g;Lin\text{-}Shu}} = -\frac{K_\mathrm{Lin\text{-}Shu} c_\mathrm{s}^2}{m\left(\Omega - \Omega_\mathrm{p}\right)}\,. \tag{66}$$

4.4.2 Angular momentum flux of WKB waves

Density waves transport angular momentum. This ability will be important for the formation of nuclear rings. The general equations that express angular momentum conservation in fluid discs are reviewed in Appendix B. The wave angular momentum is given by

$$F_\mathrm{w} = R^2 \rho_0 \int_0^{2\pi} \mathrm{d}\theta v_{\theta 1} v_{R1}\,. \tag{67}$$

Using Eqs. (46) and (47), we can rewrite this, after some algebra, as:

$$F_\mathrm{w} = \mathrm{Re}\left\{\frac{\pi R \rho_0 i m}{D}\left[\left(\Phi_1 + h_1\right)^* \frac{\mathrm{d}}{\mathrm{d}R}\left(\Phi_1 + h_1\right)\right]\right\}\,. \tag{68}$$

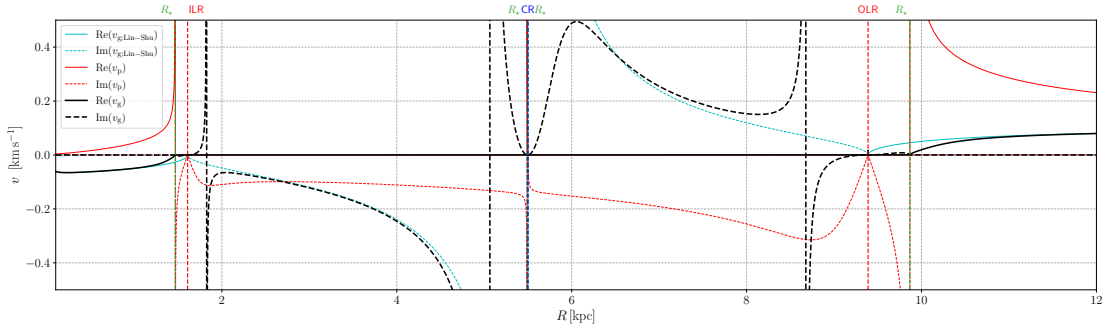

Figure 9: Group velocity v_g and phase velocity v_p of WKB waves for $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$ and constant unperturbed density $\rho_0(R) = 1$. The cyan line shows the group velocity according to the Lin-Shu dispersion relation (66). Resonances and turning points are marked by vertical dashed lines.

where $*$ denotes the complex conjugate. Setting $\Phi_1 = 0$ (free waves) and using (55) this becomes

$$F_\mathrm{w} = \mathrm{Re}\left\{ i\pi m g_1^* \frac{\mathrm{d}g_1}{\mathrm{d}R} \right\} \mathrm{sgn}\,(D)\ . \tag{69}$$

To evaluate the flux of angular momentum associated with the WKB waves (Eq. 59), we substitute Eq. (59) into Eq. (69). We obtain:

$$F_\mathrm{w} = m\pi \left(|C_2|^2 - |C_1|^2 \right) \mathrm{sgn}\,(D)\ . \tag{70}$$

Since C_1 and C_2 are constant for a given WKB wave, this equation shows that the flux of angular momentum is constant with R. In fact, the quantity (69) is conserved *exactly* by Eq. (56), not just in the WKB approximation (you can prove this by direct differentiation and using 56). The angular momentum flux is also the adiabatic invariant associated with the WKB solution (Appendix A). Eq. (70) shows that the trailing (leading) wave has $F_w > 0$ ($F_w < 0$). Thus, trailing (leading) waves transport angular momentum outwards (inwards).

4.5 Forced waves

An oscillating external potential can generate waves in a similar way a vibrating rod can generate sound waves in air. If the potential oscillates with a given frequency ω,[5] the excited waves will have the same frequency ω in the linear regime. The wavelength of waves at this frequency is determined by the dispersion relation to be $\lambda = 2\pi/K(\omega)$ (top panel of Fig. 5).[6] What is the amplitude of the generated

[5] ω is related to the pattern speed of the external potential by Eq (41).

[6] From this we can already tell their pitch angle generated by the bar potential, Eq. (61).

waves? The coupling between the external forcing and the fluid will be strongest when λ is of the same order of the typical scalelength of the external forcing. Therefore the excited waves will be largest when this condition is satisfied. In appendix C we present an exactly solvable toy problem that illustrates this concept in more detail with a concrete example.

The external forcing in our problem is represented by F (or equivalently Q) on the right-hand side of Eq. (50) (or 56). Waves will be effectively excited whenever the typical scalelength of the forcing $L \sim F/(\mathrm{d}F/\mathrm{d}R)$ is comparable to the wavelength $\lambda = 2\pi/K$. However, as can be seen in Fig. 5, at most locations we have $\lambda \ll L$, so the coupling is weak and the the excited waves negligibly small. This is because the forcing F varies on the scale of the barred potential $\Delta\Phi$ and of the background state (Eq. (7)), and both of these typically vary on a scalelength $L \sim R$, while $K \ll R$ for small sound speeds.

There are two exceptions in which the forcing can efficiently couple to the waves:

- **Near the resonances**. In the vicinity of a resonance, $\lambda = 2\pi/K$ becomes large (more precisely, it diverges at the turning points $R_\star$, top panel in Fig. 5). Thus, we can have $\lambda \sim R$, and the forcing can couple to the waves.

- **At sharp edges**. When ρ_0 changes abruptly, the forcing F varies on the same scale as ρ_0 (see for example the 'bumps' in Q at the sharp edge of ρ_0, dashed line in the middle panel of Fig. 5). If this scale is small enough, it can be of the same order of λ, and the forcing can couple to the waves.

In this section, we will solve the non-homogeneous Eq. (56) (or equivalently 50) with the forcing term included. In Sect. 4.5.1 we solve it away from resonances and sharp edges. In Sect. 4.5.2 and 4.5.3 we solve it near resonances and sharp edges, and calculate the amplitude of the waves excited at these locations.

4.5.1 Non-wave solution away from resonances and sharp edges

Away from resonances and sharp edges, the following is an approximate particular solution of Eq. (56):

$$g_Q(R) = \frac{Q}{K^2}\,. \tag{71}$$

To see the condition under which this solution is valid, we substitute $g_1 = g_Q$ in Eq. (56), and impose that the first term on the left-hand side is small, i.e. $\mathrm{d}^2 g_1/\mathrm{d}R^2 \ll K^2 g_1$. This gives the following condition:

$$\frac{\mathrm{d}^2}{\mathrm{d}R^2}\left(\frac{Q}{K^2}\right) \ll Q\,. \tag{72}$$

This condition essentially states that the forcing should vary slowly with respect to the wavelength. This is verified away from resonances and sharp edges (Fig. 5). This condition is more accurately verified at low sound speed, since $K \to \infty$ as $c_s \to 0$ at fixed R (see Eq. 57).

Eq. (71) represents a non-wave solution which is the analogue of the black dashed solution for the toy problem in Fig. 12 in Appendix C an is equivalent to Eq. (15) of [3]. In the analogy between (56) and the harmonic oscillator, this solution corresponds to following the "instantaneous" equilibrium position of the oscillator as the external force slowly varies. It is be valid when the "force" $Q(R)$ varies slowly enough compared to the frequency of the harmonic oscillator.

The full solution of the non-homogeneous (56) away from resonances and sharp edges is the sum of the particular solution (72) and of the WKB homogeneous solution (59):

$$\boxed{g_1(R) = \frac{C_1}{\sqrt{K(R)}} \exp\left[i \int_{R_0}^{R} K(s)\mathrm{d}s\right] + \frac{C_2}{\sqrt{K(R)}} \exp\left[-i \int_{R_0}^{R} K(s)\mathrm{d}s\right] + g_Q(R)} \tag{73}$$

In the next two sections we calculate the approximate solutions near the resonances and at sharp edges and "match" these to the solution (59).

4.5.2 Excitation of waves near the Lindblad resonances

To solve Eq. (56) near the Lindblad resonances, we introduce the dimensionless variable

$$x = \frac{R - R_{\mathrm{ILR}}}{R_{\mathrm{ILR}}}\,. \tag{74}$$

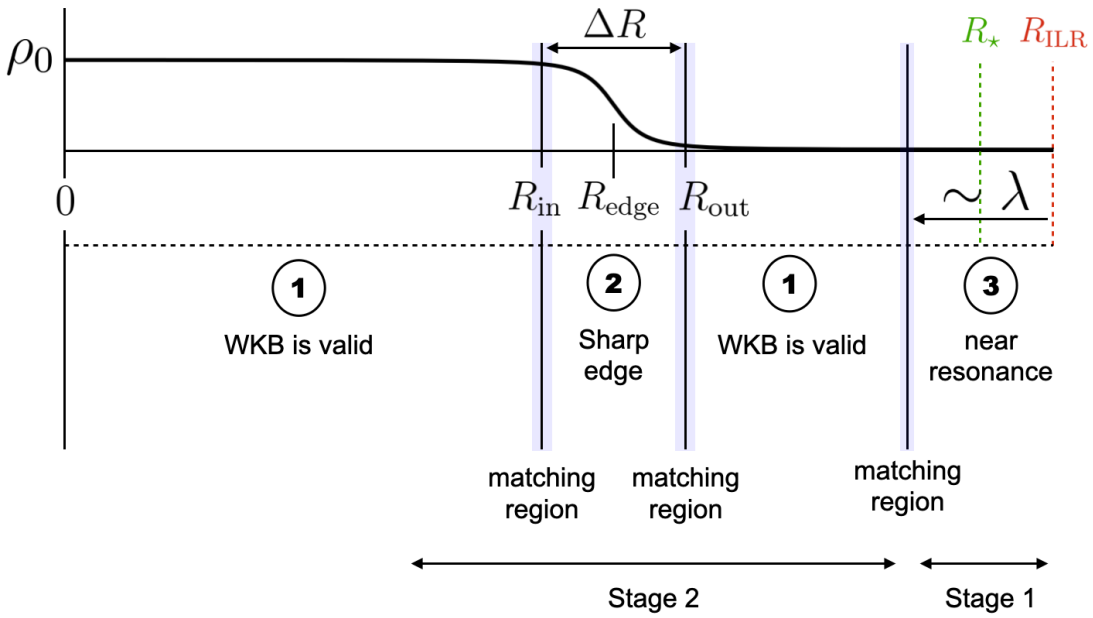

Figure 10: Schematic diagram of where the various approximate solutions of Eq. (56) apply. (1) denotes regions away from resonances and sharp edges where the solution is well approximated by (73). (2) denotes the region near the edge where Eq. (127) is a good approximation. (3) denotes the region within approximately one wavelength λ from the ILR, where the solution (122) is a good approximation. The shaded "matching regions" denote where two neighbouring solutions are simultaneously good approximations and we can apply the method of matched asymptotic expansions. "Stage 1" and "Stage 2" denote the regions corresponding to the two stages of our picture for ring formation described in Sect. 5.

We approximate the coefficients of Eqs. (50) and (56) for $x \ll 1$ and $c_{\rm s}/\Omega R \ll 1$:

$$D \simeq \mathcal{D} x \tag{75}$$

$$H \simeq -\left(\frac{1}{R_{\rm ILR}}\right)\frac{1}{2x} \tag{76}$$

$$C \simeq -\left(\frac{1}{R_{\rm ILR}}\right)^2\frac{\alpha}{x} \tag{77}$$

$$W \simeq -\left(\frac{1}{R_{\rm ILR}}\right)^2\left[\frac{\alpha}{x}+\beta x\right] \tag{78}$$

$$F \simeq \left(\frac{1}{R_{\rm ILR}}\right)^2\frac{\psi}{x} \tag{79}$$

$$Q \simeq \mathrm{sgn}\,(x)\left(\frac{1}{R_{\rm ILR}}\right)^2\frac{\gamma\psi}{|x|^{3/2}} \tag{80}$$

$$K^2 \simeq -\left(\frac{1}{R_{\rm ILR}}\right)^2\left[\frac{\alpha}{x}+\beta x+\frac{3}{4x^2}\right] \tag{81}$$

where we have defined the following constants

$$\mathcal{D} = \left[\frac{\mathrm{d}D}{\mathrm{d}R}R\right]_{R_{\rm ILR}} \tag{82}$$

$$\alpha = \left[\frac{2\Omega}{(\Omega-\Omega_{\rm p})}\right]_{R_{\rm ILR}} \tag{83}$$

$$\beta = \frac{\mathcal{D}\,R_{\rm ILR}^2}{c_{\rm s}^2} \tag{84}$$

$$\psi = \left[R\Delta\Phi' + \left(\frac{2\Omega}{(\Omega-\Omega_{\rm p})}\right)\Delta\Phi\right]_{R_{\rm ILR}} \tag{85}$$

$$\gamma = \left(\frac{R\rho_0}{\mathcal{D}}\right)^{1/2} \tag{86}$$

Thus Eqs. (56) becomes

$$\frac{\mathrm{d}^2 g_1}{\mathrm{d}x^2} - \left[\frac{\alpha}{x}+\beta\,x+\frac{3}{4x^2}\right]g_1 = \mathrm{sgn}\,(x)\,\frac{\gamma\psi}{|x|^{3/2}}\,, \tag{87}$$

while Eq. (50) becomes

$$\frac{\mathrm{d}^2 h_1}{\mathrm{d}x^2} - \frac{1}{x}\frac{\mathrm{d}h_1}{\mathrm{d}x} - \left[\frac{\alpha}{x}+\beta x\right]h_1 = \frac{\psi}{x}\,. \tag{88}$$

These two equations are equivalent, and can be transformed into each other using the change of variables (55), which approximated near the resonance is

$$g_1 = \mathrm{sgn}(x)\frac{\gamma}{|x|^{1/2}}h_1\,. \tag{89}$$

We wrote both versions of the equation as from (87) it is easier to see that the term α/x can be neglected, while Eq. (88) will be easier to solve near the ILR (h_1 does not diverge has an extra factor g_1 has an extra factor $1/x^{1/2}$).

Let us first check that we can neglect the α/x term. Consider the three terms in the square parentheses in Eq. (87). Very close to $x = 0$, the $3/(4x^2)$ term clearly dominates over the other two as it tends to ∞ faster. Now consider the other two terms. In our problem, α is of order unity, while β is of order of the circular velocity divided by the sound speed, $\Omega R/c_\mathrm{s}$, so $|\alpha| \ll |\beta|$. For example, for the galaxy model in Sect. 3, we have $|\alpha| \simeq 3$ and $|\beta| \sim 800$ assuming $c_\mathrm{s} = 10\,\mathrm{km\,s^{-1}}$. The consequence is that as we move away from $x = 0$, the term βx becomes comparable to $3/(4x^2)$ before the α/x term. Thus, the α/x is always subdominant compared to either of the other two terms, and can be neglected. Thus we can approximate the wavenumber (81) as

$$K^2 = -\left(\frac{1}{R_\mathrm{ILR}}\right)^2\left[\beta x + \frac{3}{4x^2}\right] \tag{90}$$

and Eqs. (87)-(88) as:

$$\frac{\mathrm{d}^2 g_1}{\mathrm{d}x^2} - \left[\beta x + \frac{3}{4x^2}\right] g_1 = \mathrm{sgn}\,(x)\,\frac{\gamma\psi}{|x|^{3/2}}\,, \tag{91}$$

$$\frac{\mathrm{d}^2 h_1}{\mathrm{d}x^2} - \frac{1}{x}\frac{\mathrm{d}h_1}{\mathrm{d}x} - \beta x\, h_1 = \frac{\psi}{x}\,. \tag{92}$$

As a sanity check, the red line in the top panel of Fig. 5 compares the approximated wavenumber given by (90) with the full equation (57), and shows that the approximation works very well.

Now we solve Eq. (92). First, we eliminate β with the substitution:[7]

$$x = t/\beta^{1/3} \tag{93}$$

$$h_1 = z/\beta^{2/3} \tag{94}$$

Eq. (92) becomes

$$\boxed{z'' - \frac{1}{t}z' - tz = \frac{b}{t}} \tag{95}$$

[7] This t is not time!

where $b = \beta^{1/3}\psi$. It turns out that this equation is closely related to a well-studied equation known as the Airy equation. The general solution of the homogeneous Eq. (92) (i.e. with $\psi = 0$) is

$$z(t) = C_1 \operatorname{Ai}'(t) + C_2 \operatorname{Bi}'(t) \tag{96}$$

where Ai and Bi are special functions known as the **Airy functions**, which have the following fundamental properties:

$$\operatorname{Ai}'' = t \operatorname{Ai} \tag{97}$$

$$\operatorname{Bi}'' = t \operatorname{Bi} \tag{98}$$

their values in zero are

$$\operatorname{Ai}(0) = \frac{1}{3^{2/3}\,\Gamma\left(\frac{2}{3}\right)}, \qquad \operatorname{Ai}'(0) = -\frac{1}{3^{1/3}\,\Gamma\left(\frac{1}{3}\right)}, \tag{99}$$

$$\operatorname{Bi}(0) = \frac{1}{3^{1/6}\,\Gamma\left(\frac{2}{3}\right)}, \qquad \operatorname{Bi}'(0) = \frac{3^{1/6}}{\Gamma\left(\frac{1}{3}\right)}. \tag{100}$$

where Γ is the Gamma function. Using these properties we can calculate the Wroskian of the Airy functions and their derivatives to be

$$W_{\operatorname{Ai}} = \operatorname{Ai}\operatorname{Bi}' - \operatorname{Ai}'\operatorname{Bi} = \frac{1}{\pi} \tag{101}$$

$$W_{\operatorname{Ai}'} = \operatorname{Ai}'\operatorname{Bi}'' - \operatorname{Ai}''\operatorname{Bi}' = -\frac{t}{\pi} \tag{102}$$

Now that we have the general solution of the homogeneous equation, we can use the method of variation of parameters to find a particular solution of the non-homogeneous equation.[8] Applying this method to Eq. (95) and using the particular solutions (96) with Wroskian (102) we obtain the following particular solution:

$$z_0(t) = \pi b \left[\operatorname{Ai}' \int \frac{\operatorname{Bi}'}{t^2}\mathrm{d}t - \operatorname{Bi}' \int \frac{\operatorname{Ai}'}{t^2}\mathrm{d}t\right] \tag{105}$$

[8] The method of variation of parameters allows one to find the solution of a non-homogeneous equation once the solution of the homogeneous equations are known. Given the following ODE:

$$p(t)y'' + q(t)y' + r(t)y = g(t) \tag{103}$$

and given that $y_1(t)$ and $y_2(t)$ are solutions of the homogeneous equation, then a particular solution of the non-homogeneous equation is:

$$y_0(t) = -y_1 \int \frac{y_2 g(t)}{W(y_1, y_2)}\mathrm{d}t + y_2 \int \frac{y_1 g(t)}{W(y_1, y_2)}\mathrm{d}t \tag{104}$$

where $W(y_1, y_2) = y_1 y_2' - y_1' y_2$ is the Wroskian.

Integrating by part and using properties (97) we can write

$$\int \frac{\mathrm{Ai}'}{t^2}\mathrm{d}t = -\frac{\mathrm{Ai}'}{t} + \int \frac{\mathrm{Ai}''}{t}\mathrm{d}t \tag{106}$$

$$= -\frac{\mathrm{Ai}'}{t} + \int \mathrm{Ai}\,\mathrm{d}t \tag{107}$$

And analogously

$$\int \frac{\mathrm{Bi}'}{t^2}\mathrm{d}t = -\frac{\mathrm{Bi}'}{t} + \int \mathrm{Bi}\,\mathrm{d}t \tag{108}$$

Substituting Eqs.(107) and (108) into (105) we find

$$z_0(t) = \pi b\left[\mathrm{Ai}'\int \mathrm{Bi}\,\mathrm{d}t - \mathrm{Bi}'\int \mathrm{Ai}\,\mathrm{d}t\right] \tag{109}$$

This expression does not look too bad. In fact, it turns out that the expression within parentheses in Eq. (109) is the derivative of a function known as the **Scorer function**, defined as:

$$\mathrm{Hi}(x) = \mathrm{Bi}(x)\int_{-\infty}^{x}\mathrm{Ai}(t)\,dt - \mathrm{Ai}(x)\int_{-\infty}^{x}\mathrm{Bi}(t)\,dt \tag{110}$$

So we can rewrite (109) as

$$\boxed{z_0(t) = -\pi b\,\mathrm{Hi}'(t)} \tag{111}$$

The general solution to Eq. (95) can then be written

$$z(t) = C_1\,\mathrm{Ai}'(t) + C_2\,\mathrm{Bi}'(t) - \pi b\,\mathrm{Hi}'(t) \tag{112}$$

We now need to find C_1 and C_2. To do this, we see how the solution behaves for $|t| \to \infty$. Using the NIST digital library of mathematical functions[9] we find the following series expansions for $t \to +\infty$:

$$\mathrm{Hi}'(t) \sim \frac{e^{\frac{2t^{3/2}}{3}}t^{1/4}}{\sqrt{\pi}}; \qquad t \to +\infty \tag{113}$$

$$\mathrm{Ai}'(t) \sim \frac{e^{-\frac{2t^{3/2}}{3}}t^{1/4}}{\sqrt{\pi}}; \qquad t \to +\infty \tag{114}$$

$$\mathrm{Bi}'(t) \sim \frac{e^{\frac{2t^{3/2}}{3}}t^{1/4}}{\sqrt{\pi}}; \qquad t \to +\infty \tag{115}$$

[9] https://dlmf.nist.gov

If we require that the solution does not explode exponentially at $t = +\infty$ then we need to set $C_2 = \pi b$ in the homogeneous solution (Eq. 96).

Next we look at the series expansions around $t \to -\infty$. We have

$$\mathrm{Hi}'(t) \sim \frac{1}{\pi t^2}; \qquad t \to -\infty \tag{116}$$

$$\mathrm{Ai}'(t) \sim -\frac{(-t)^{1/4}}{\pi^{1/2}} \cos\left(\frac{2}{3}(-t)^{3/2} + \frac{\pi}{4}\right); \qquad t \to -\infty \tag{117}$$

$$\mathrm{Bi}'(t) \sim \frac{(-t)^{1/4}}{\pi^{1/2}} \sin\left(\frac{2}{3}(-t)^{3/2} + \frac{\pi}{4}\right); \qquad t \to -\infty \tag{118}$$

Setting $C_2 = \pi b$ as found above, the solution at $t = -\infty$ can be written:

$$z(t) \sim i\pi b\alpha(t) \left\{ \frac{iC_1}{\pi b} \cos[\phi(t)] - i \sin[\phi(t)] \right\} - \frac{b}{t^2}; \qquad t \to -\infty \tag{119}$$

where

$$\phi(t) = \frac{2}{3}(-t)^{3/2} + \frac{\pi}{4}, \tag{120}$$

$$\alpha(t) = \frac{(-t)^{1/4}}{\pi^{1/2}}. \tag{121}$$

To find C_1, the appropriate boundary conditions are "radiation" boundary conditions. Causality requires that waves propagate away from the region where they are generated (a solution in which waves come towards it would require a source of waves outside this region). The direction of propagation of the waves can be understood from their group velocity as described in Sect. 4.4.1. Since inside the ILR trailing waves move inwards and leading waves move outwards (Sect. 4.4.1), we require that at $t = -\infty$ our solution resembles a trailing wave. Eq. (119) looks like a trailing wave if we choose $C_1 = -i\pi b$. This is an informal guess for now but we will verify in a moment that it is correct. So the final solution to our problem with the correct boundary conditions is:

$$\boxed{z(t) = \pi b \left[-i\,\mathrm{Ai}'(t) + \mathrm{Bi}'(t) - \mathrm{Hi}'(t)\right]} \tag{122}$$

To verify that everything is consistent, we match this solution at $t \to -\infty$ to the trailing free WKB density waves studied in Sect. 4.4. This will allow us to find the angular momentum carried by the excited wave. The solution (122) at $t = -\infty$ is

$$z(t) \sim i\pi b\alpha(t) e^{i\phi(t)} - \frac{b}{t^2}; \qquad t \to -\infty \tag{123}$$

Using (89), (93), (94), (120), (121) we can rewrite this as

$$g_1(x) \sim -\frac{i\gamma\psi\pi^{1/2}}{\beta^{1/4}|x|^{1/4}} e^{i\left(\frac{2}{3}|x|^{3/2}\beta^{1/2}+\frac{\pi}{4}\right)} - \frac{\gamma\psi}{|x|^{5/2}\beta}; \qquad x \to -\infty \tag{124}$$

The second term on the RHS corresponds to the non-wave solution (71) where Q is approximated as in Eq. (80), and $K \simeq \beta x$ consistent with (90) at large $|x|$. The first term corresponds to the trailing WKB wave in Eq. (59) where again $K \simeq \beta x$ and $|C_1| = \pi^{1/2}\gamma\psi$ (note that the dependence on x is the same!). This verifies that the solution (122) asymptotically reduces to the sum of the WKB solution and the non-wave solution (71). The angular momentum flux carried by this wave can be calculated using Eq. (70):

$$\boxed{F_w = m\pi^2\gamma^2\psi^2} \tag{125}$$

This is the famous formula for the torque exerted by an external potential at the Lindblad resonance found by [3] (their Eq. 46). Although we calculated it for the case without self-gravity, these authors have shown that it is valid in the case with self-gravity as well.

4.5.3 Excitation of waves near sharp edges

Consider an edge at $R_{\rm edge}$ of width $R_{\rm out} - R_{\rm in} = \Delta R$, where $R_{\rm in}$ and $R_{\rm out} > R_{\rm in}$ are the two extremities of the region over which the edge extends (see Fig. 10). The shape of the edge can be arbitrary. The edge is assumed to be thin but not too thin, otherwise the unperturbed density profile would violate the Rayleigh criterion (21) and become unstable. In practice, considering the Lin-Shu approximation (Eq. 63) this means that the edge should not be thinner than approximately one wavelength, $\lambda = 2\pi/K$.

At $R < R_{\rm in}$ and $R > R_{\rm out}$, the solution of Eq. (56) can be approximated by (73). We assume that waves are excited only near the edge, i.e. at $R_{\rm in} < R < R_{\rm out}$. As in the previous section, we impose radiation boundary conditions. This means that at $R < R_{\rm in}$ we should have only the trailing wave and at $R > R_{\rm out}$ only the leading wave. Therefore we write:

$$g_1(R) = \begin{cases} \frac{C_{\rm in}}{\sqrt{K(R)}} \exp\left[i\int_{R_0}^{R} K(s)\mathrm{d}s\right] + \frac{Q}{K^2}, & \text{for } R < R_{\rm in}\,, \\ \frac{C_{\rm out}}{\sqrt{K(R)}} \exp\left[-i\int_{R_0}^{R} K(s)\mathrm{d}s\right] + \frac{Q}{K^2}, & \text{for } R > R_{\rm out}\,. \end{cases} \tag{126}$$

The constants $C_{\rm in}$ and $C_{\rm out}$ will be determined by solving Eq. (56) near the edge and matching the two solutions.

Near the edge, the forcing term Q varies rapidly, violating condition (72), and the equilibrium solution (71) fails (dashed line in the middle panel of Fig. 5).

To solve Eq. (56) near the edge, we proceed as follows. We assume that K is approximately constant across the edge, i.e. in the range $R_{\rm in} < R < R_{\rm out}$. This is justified by our assumption that the edge is relatively sharp (see also the black dashed line in the top panel of Fig. 5). Under this assumption, Eq. (56) can be solved using the method of variation of parameters. The general solution is:

$$g_1(R) = A_1 e^{iKR} + A_2 e^{-iKR} + \\ - \frac{ie^{iKR}}{2K}\int_{R_0}^{R} Q(s)e^{-iKs}\mathrm{d}s + \frac{ie^{-iKR}}{2K}\int_{R_0}^{R} Q(s)e^{iKs}\mathrm{d}s \ , \tag{127}$$

where $K \simeq K(R_{\rm in}) \simeq K(R_{\rm out})$. The constants A_1 and A_2 are determined by the condition that the solution contains only waves travelling inwards at radii $R < R_{\rm in}$, and waves travelling outwards at radii $R_{\rm out} < R$. These calculations are reported in Appendix D.1. We find:

$$A_1 = \frac{Q_{\rm out}}{2K^2}e^{-iKR_{\rm out}} + \frac{i}{2K}\int_{R_0}^{R_{\rm out}} Q(s)e^{-iKs}\mathrm{d}s \tag{128}$$

$$A_2 = \frac{Q_{\rm in}}{2K^2}e^{iKR_{\rm in}} - \frac{i}{2K}\int_{R_0}^{R_{\rm in}} Q(s)e^{iKs}\mathrm{d}s \ , \tag{129}$$

where $Q_{\rm out} = Q(R_{\rm out})$ and $Q_{\rm in} = Q(R_{\rm in})$.

Both Eq. (126) and Eq. (127) are valid solutions of Eq. (56) in a neighbourhood of $R_{\rm in}$ and in a neighbourhood of $R_{\rm out}$ (shaded regions in Fig. 10). Matching these two solutions gives (see Appendix D.1):

$$C_{\rm in} = -C^*_{\rm out} = \frac{ie^{iKR_0}}{2K^{1/2}}\int_{R_{\rm in}}^{R_{\rm out}} Q(s)e^{-iKs}\mathrm{d}s + \\ + \frac{Q_{\rm out}}{2K^{3/2}}e^{iK(R_0-R_{\rm out})} - \frac{Q_{\rm in}}{2K^{3/2}}e^{iK(R_0-R_{\rm in})} \ . \tag{130}$$

The coefficients $C_{\rm in}$ and $C_{\rm out}$ give the amplitude of density waves excited at the edge. It can be shown that the absolute values $|C_{\rm in}|$ and $|C_{\rm out}|$ are independent of R_0, as they should since the angular momentum flux at the edge should be independent of this arbitrary radius.

The rapid variation of Q near the edge is what generates density waves. The coupling between the forcing term Q and the density waves is expected to be maximum when the scale-length over which Q varies is comparable to the wavelength of the waves $\lambda = 2\pi/K$ (similarly to the toy problem in Appendix C), i.e. when $\Delta R \simeq \lambda$.

We can use Eqs. (70) and (130) to calculate the angular momentum flux carried by the waves excited at the edge. To obtain a closed formula it is necessary to

make some further assumptions on the edge. If the edge of the disc is marginally stable to the Rayleigh criterion (21), one has $|R_{\rm out} - R_{\rm in}| \sim c_{\rm s}/\Omega \sim 1/K$. This is the sharpest edge that can be constructed without making the unperturbed density distribution unstable. Then the exponential $\exp(-iKs)$ in the integral of Eq. (130) is nearly constant. As shown in Appendix D.2, in this case Eq. (130) reduces to:

$$|C_{\rm in}| \simeq |C_{\rm out}| \simeq \left[\left(\frac{R\rho_0}{K\,|D|} \right)^{1/2} \left| \frac{{\rm d}\Phi_1}{{\rm d}R} + \frac{2\Omega}{\Omega - \Omega_{\rm p}} \frac{\Phi_1}{R} \right| \right]_{R=R_{\rm edge}} . \tag{131}$$

Note that this is essentially the impulse approximation, i.e. we have assumed that the force Q gives an instantaneous "kick" at $R = R_{\rm edge}$. Using Eq. (70), the flux of angular momentum of waves excited at a sharp edge is then

$$\boxed{F_w \simeq m\pi \left[\left(\frac{R\rho_0}{K\,|D|} \right) \left(\frac{{\rm d}\Phi_1}{{\rm d}R} + \frac{2\Omega}{\Omega - \Omega_{\rm p}} \frac{\Phi_1}{R} \right)^2 \right]_{R=R_{\rm edge}}} \tag{132}$$

Eq. (132) is correct when the distance of the edge from the inner Lindblad resonance is larger than approximately one wavelength, i.e. $K|R_{\rm edge} - R_{\rm ILR}| \gg 1$. At $|R_{\rm edge} - R_{\rm ILR}| = \lambda/(2\pi^2) = 1/(\pi K)$ and approximating $D(R) \simeq (R - R_{\rm ILR})({\rm d}D/{\rm d}R)$, as appropriate near the ILR where D vanishes, Eq. (132) becomes identical to Eq. (125) which gives the flux of angular momentum of waves excited at the resonance.

5 The formation of nuclear rings

We are finally ready to put everything together and describe our picture of the formation of nuclear rings. For simplicity, we imagine to start with a uniform density distribution ρ_0 extending everywhere. The formation of the ring can be schematically divided into two stages, which depend on the distance of the edge of the gas disc from the ILR. The regions corresponding to the two stages are schematically marked in Fig. 10.

We stress that this picture is developed using the linear regime, but the actual process is very non-linear. Thus, the quantitative results are most certainly *not* accurate. The reason we do it this way is to gain a deeper understanding of the dynamical mechanisms at play.

5.1 First stage ($|R_{\rm edge} - R_{\rm ILR}| \lesssim \lambda$)

In the first stage, a trailing spiral wave is excited near the ILR by the external bar potential. This is the regime analysed in Sect. 4.5.2. The wave travels inwards

but for realistic strengths of the bar potential it very quickly becomes non-linear and develops into a shock. The wave then dissipates, depositing its (negative) angular momentum into the gas disc (i.e., removing angular momentum from the gas disc).[10] This reduces the angular momentum of the disc, causing the gas to move inward. A gap opens around the ILR. The gas accumulating at the inner edge of this gap starts forming the nuclear ring.

The width of the gap opened in the first stage is the range of validity of the calculations of Sect. 4.5.2, which is approximately one wavelength, i.e. $|R_{\rm edge} - R_{\rm ILR}| \sim \lambda$. Using the Lin-Shu approximation[11] (Eq. 63) and approximating $D \simeq (R - R_{\rm ILR})({\rm d}D/{\rm d}R)$ (recall that $D = 0$ at the resonance) we have $\lambda \sim c_{\rm s}/|D|^{1/2} \sim c_{\rm s}/|(R_{\rm edge} - R_{\rm ILR})({\rm d}D/{\rm d}R)|^{1/2}$ and therefore $|R_{\rm edge} - R_{\rm ILR}| \sim |c_{\rm s}^2/({\rm d}D/{\rm d}R)|^{1/3} \sim (c_{\rm s}/v_0)^{2/3} R_{\rm ILR}$. The size of the gap is therefore much smaller than the radius of the resonance, and increases for increasing sound speed.

The velocity at which the edge of the gap moves in this phase can be estimated by dividing the flux of angular momentum of the waves, F_w, by the amount of angular momentum per unit radius in the unperturbed disc, $2\pi\rho_0 R^3\Omega$:

$$\frac{{\rm d}R_{\rm edge}}{{\rm d}t} = -\left[\frac{F_w}{2\pi\rho_0 R^3\Omega}\right]_{R=R_{\rm edge}} , \tag{133}$$

where F_w can be estimated using Eq. (125), bearing in mind that these calculations are valid in the linear approximation and should not be expected to be accurate for the highly non-linear waves excited by a realistically strong bar potential. Taking into account that $m = 2$, we find

$$\boxed{\frac{{\rm d}R_{\rm edge}}{{\rm d}t} = -\left[\left(\frac{\pi}{R^2\Omega({\rm d}D/{\rm d}R)}\right)\left(\frac{{\rm d}\Phi_1}{{\rm d}R} + \frac{2\Omega}{\Omega - \Omega_{\rm p}}\frac{\Phi_1}{R}\right)^2\right]_{R=R_{\rm edge}}} \tag{134}$$

Inserting the numbers of our gravitational potential (Sect. 3) into Eq. (134) we obtain ${\rm d}R_{\rm edge}/{\rm d}t \simeq 6\,{\rm km\,s^{-1}}$. The duration of the first stage can be estimated by dividing the size of the gap by the velocity of the edge. Using $|R_{\rm edge} - R_{\rm ILR}| \sim (c_{\rm s}/v_0)^{2/3} R_{\rm ILR}$, $v_0 = 220\,{\rm km\,s^{-1}}$, $R_{\rm ILR} = 1.6\,{\rm kpc}$ and the value of ${\rm d}R_{\rm edge}/{\rm d}t$ found above we obtain:

$$T_1 \simeq \left(\frac{c_{\rm s}}{10\,{\rm km\,s^{-1}}}\right)^{2/3} 30\,{\rm Myr}\,. \tag{135}$$

This time is relatively short, just a few orbital times for a typical nuclear ring. Actually, the evolution is likely to be even faster than Eq. (135) suggests because

[10] Linear waves of small amplitude travel to the centre without affecting the unperturbed density of the disc. It is only when they become non-linear that they can dump their angular momentum in the unperturbed disc.

[11] More precisely, we should calculate λ from the first oscillation of the solution (122).

of the non-linearity of the process. The evolution of the gap after the first stage and its final size are determined by the second stage.

5.2 Second stage ($|R_{\rm edge} - R_{\rm ILR}| \gtrsim \lambda$)

At the beginning of the second stage there is a gap around the ILR, and the distance between the inner edge of the gap and the ILR is approximately one wavelength. Since the width of the edge at this point can be at most one wavelength (because the edge tail cannot extend beyond the ILR), the edge is "sharp" by definition and strong waves will be excited at its location according to the analysis in Sect. 4.5.3. Similarly to the waves excited near the ILR in the first stage, the waves excited near the edge will become quickly non linear and dissipate, removing the angular momentum from the gas disc and causing the edge to move inwards. The gas accumulating at the edge as it sweeps inward is the nuclear ring.

The speed at which the edge moves during the second stage can be estimated using Eq. (133), where we use Eq. (132) to estimate the flux of angular momentum F_A of waves excited at sharp edges. Using the Lin-Shu approximation (Eq. 63) to write $K \simeq |D|^{1/2}/c_{\rm s}$, and $m = 2$, we find:

$$\boxed{\frac{{\rm d}R_{\rm edge}}{{\rm d}t} = -\left[\left(\frac{2c_{\rm s}}{R^2\Omega\,|D|^{3/2}}\right)\left(\frac{{\rm d}\Phi_1}{{\rm d}R} + \frac{2\Omega}{\Omega-\Omega_{\rm p}}\frac{\Phi_1}{R}\right)^2\right]_{R=R_{\rm edge}}}\,. \tag{136}$$

where the factor of 2 takes into account that the outward-travelling leading wave excited at the edge will be reflected at $R = R_\star$ into an inward-travelling trailing wave. Note that Eq. (134) and (136) only differ for the factor in the first round parentheses on the right-hand-sides, and this factor coincides in the two equations at a distance of approximately one wavelength from the ILR. This is the point where we transition from the analysis of Sect. 4.5.2 to the analysis in Sect. 4.5.3, and from the first to the second stage.

When does the edge stop moving? The process above goes on until waves continue to be effectively excited at the edge, which happens when both of the following conditions are satisfied: (i) the edge is "sharp", i.e. the edge width is smaller than a few times the wavelength of density waves $\lambda = 2\pi/K$; (ii) the gravitational potential Φ_1 is sufficiently strong. The distance between the edge and the ILR poses an upper limit to the width of the edge since the edge cannot cross the ILR, $\Delta R < |R_{\rm ILR} - R_{\rm edge}|$.[12] Therefore, when the edge is not sufficiently far from the ILR, it *must* be sharp. In particular, we can expect the edge to keep

[12] Recall also that as discussed in Sect. 4.2 the edge cannot be too thin, otherwise the system becomes Rayleigh-unstable. Thus, we expect the edge width to remain of order λ during the shrinking process.

moving until it is located a few wavelengths away from the ILR. Since λ increases roughly linearly with c_s (Eq. 63), we expect the edge to move farther at larger sound speed, proportionally to the sound speed. Thus, the gap is wider for larger sound speed, and the radius of the nuclear ring is smaller.

Predicting exactly where the edge will stop, and therefore the final radius of the ring, is a difficult task. The process is highly non-linear, and the unperturbed density profile changes in a way that cannot be calculated in the linear approximation. Numerical experiments [4] show that for our assumed gravitational potential the ring stops when one can fit approximately 7 wavelengths λ between $R_{\rm edge}$ and $R_{\rm ILR}$. For weaker barred potential, the edge might stop sooner if Φ_1 is too small to generate sufficient flux of angular momentum at the edge. More details, including hydrodynamical simulations that show the non-linear development of the scenario outlined above, can be found in [4] and references therein.

5.3 Analogies with the formation of gaps in planetary rings

Goldreich & Tremaine in 1978 [3] developed a picture for the formation of the Cassini division in Saturn's ring that has many similarities with our picture for the formation of nuclear rings described in Sect. 5. Indeed, they were the first to derive the solution described in Sect. 4.5.2 for the waves excited near the Lindblad resonance. In both pictures: (i) a gap opens near the Lindblad resonance due to waves excited at the resonance; (ii) subsequent excitation of waves at the edge of the gap continues to widen the gap. The main differences are: (i) The bar potential is a non-axisymmetric perturbation many orders of magnitude stronger than the one from Saturn's satellite Mimas; (ii) the sound speed is negligibly small in Saturn's problem, while the effects of finite sound speed are important in our problem, since the size of the gap (and therefore the radius of the ring) depend strongly on the sound speed [4]; (iii) self-gravity is negligible for our case, but it is not negligible in Saturn's problem. In particular, gravity is the main means of transport of angular momentum in Saturn's problem, while advective transport through pressure is the main mechanism for transport in our problem.

6 Conclusions

We have seen how the formation of nuclear rings can be explained as an accumulation of gas at the inner edge of a gap that forms around the inner Lindblad resonance of a barred potential. The gap opens because waves excited by the bar potential remove angular momentum from the gas, moving it inwards. The gap widens in time because the bar potential continues to excite trailing waves at the inner edge of the gap. The widening stops when the edge is at a distance of

several wavelengths from the ILR. The gas accumulating at the inner edge of the gap forms the nuclear ring. The linear approximation allowed us to elucidate the basic dynamical process at play. For a realistic strength of the bar potential, the process is highly non-linear, and the process is better quantitatively studied using hydrodynamical simulations. More details can be found at the references provided below, in particular see [4] and references therein.

A The WKB method

The WKB method is a method for finding approximate solutions to linear differential equations with spatially varying coefficients. We illustrate how this method works for an equation of the type (56) with $Q = 0$. A more comprehensive exposition of the WKB method can be found for example in Chapter 10 of [5].

A.1 First order WKB

Consider the following differential equation:

$$\ddot{x} + \omega(t)^2 x = 0 \tag{137}$$

where the dot denotes derivative with respect to t, ω is real and x is complex. If $\omega =$ constant, this is the equation of a simple harmonic oscillator. The general solution is:

$$x(t) = A \exp[i\omega t] + B \exp[-i\omega t]\,, \tag{138}$$

where A, B are arbitrary complex constants. The period of the oscillation is $T = 2\pi/\omega$. When $\omega(t)$ is not constant, Equation (137) has in general no analytic solution. However, when $\omega(t)$ is 'slowly varying', we can find solutions using the WKB method. By 'slowly varying', we mean that the change in $\omega(t)$ during an oscillation is small, so that ω changes appreciably only after many oscillations. The typical timescale over which $\omega(t)$ changes is $T_\omega = \omega/\dot{\omega}$. The condition that ω changes slowly is $T_\omega \gg T$, or equivalently:

$$\boxed{\frac{2\pi\dot{\omega}}{\omega^2} \ll 1} \tag{139}$$

This is the condition under which we can apply the WKB approximation. Physically, Equation (137) corresponds to a mass m connected to a spring with a spring constant $k(t) = \omega^2 m$ that slowly changes over time. We expect such a system to instantaneously oscillate with a frequency that is given by the (instantaneous)

ω, while the amplitude of the oscillation $A(t)$ slowly changes over time. Thus we guess a solution of the following form:

$$x(t) = A(t) \exp\left[\pm i \int_{t_0}^{t} \omega(s) \mathrm{d}s\right] \tag{140}$$

where A and B are complex. Note that in the argument of the exponential we have the integral $\int \omega(s)\mathrm{d}s$ (and not the product ωt). Intuitively, we can think of $\int \omega(s)\mathrm{d}s$ as the phase of the oscillation, i.e. a number that quantifies how many oscillations occurred since the beginning of the motion. In the case $\omega =$ constant, the integral reduces to ωt and we recover the simple harmonic oscillator. Taking the derivatives of (140) we find:

$$\dot{x}(t) = \left(\dot{A} \pm iA\omega\right) \exp\left[\pm i \int_{t_0}^{t} \omega(s) \mathrm{d}s\right] \tag{141}$$

$$\ddot{x}(t) = \left(\ddot{A} \pm 2i\dot{A}\omega \pm iA\dot{\omega} - A\omega^2\right) \exp\left[\pm i \int_{t_0}^{t} \omega(s) \mathrm{d}s\right] \tag{142}$$

Substituting (142) into (137) and simplifying the exponential we obtain:

$$\left(\ddot{A} \pm 2i\dot{A}\omega \pm iA\dot{\omega}\right) = 0 \tag{143}$$

Until now everything has been exact. Equation (143) is completely equivalent to (137), and Equation (140) can be simply seen as a change of variable in which we replace x with A. Now comes the WKB approximation. The essence of this approximation is that every time you take a derivative of $A(t)$ or $\omega(t)$, you get something smaller by a factor $\epsilon \sim 2\pi\dot{\omega}/\omega^2$ (see Equation 139). In other words, you can estimate the magnitudes of time derivatives by replacing $\mathrm{d}/\mathrm{d}t \sim \epsilon\omega$. Thus $\dot{A} \sim \epsilon\omega A$, $\dot{\omega} \sim \epsilon\omega$. For the second derivatives $\ddot{A} \sim \epsilon\omega\dot{A} \sim \epsilon^2\omega^2 A$. Using these relations, we see that the term $\ddot{A}$ in Equation (143) can be neglected compared to the others. This gives:

$$2\dot{A}\omega + A\dot{\omega} = 0\,. \tag{144}$$

The solution to this equation is

$$A(t) = \frac{C}{\sqrt{\omega(t)}} \tag{145}$$

$$\tag{146}$$

where C is an arbitrary complex constant. Plugging this into (140) we see that the general solution of equation (137) in the WKB approximation is:

$$\boxed{x(t) = \frac{C_1}{\sqrt{\omega(t)}} \exp\left[i \int_{t_0}^{t} \omega(s) \mathrm{d}s\right] + \frac{C_2}{\sqrt{\omega(t)}} \exp\left[-i \int_{t_0}^{t} \omega(s) \mathrm{d}s\right]} \tag{147}$$

where C_1 and C_2 are arbitrary complex constants. We can also write an approximate expression for the derivative using (141) and neglecting the small terms $\dot{A}$ and $\dot{B}$ as:

$$\boxed{\dot{x}(t) = iC_1\sqrt{\omega(t)}\exp\left[i\int_{t_0}^{t}\omega(s)\mathrm{d}s\right] - iC_2\sqrt{\omega(t)}\exp\left[-i\int_{t_0}^{t}\omega(s)\mathrm{d}s\right]} \tag{148}$$

Figure 11 shows a comparison between an exact solution of Equation (137) and the WKB approximated solution obtained using (147).

Remark: the total energy of a simple harmonic oscillator is

$$E = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}m\omega^2 x^2\,. \tag{149}$$

This is in general not conserved when $\omega(t)$ is not constant. However, if we calculate E using the approximate solution (147) and (148) we obtain $E = \text{constant} \times \omega$. In other words, we obtain that the following quantity is constant:

$$J = \frac{E}{\omega}\,. \tag{150}$$

This is rather remarkable, because it means that the amplitude of the oscillation becomes a function of ω. If we increase ω slowly then we slowly decrease it to its original value, at the end of the process the amplitude will be the same as it was at the start. It is easy to see that this is violated if $\omega(t)$ does not change slowly (think for example of abruptly changing ω when the system passes through $x = 0$: in this case the energy does not change, but ω does). In fact, even if we vary $\omega(t)$ arbitrarily slowly but choose the phase of oscillation under at which the changes happen, we can make the amplitude increase even without a net increase in ω ('parametric resonance'). Thus, there must other conditions of 'randomness' on how ω is changed in order for J to be constant.

The quantity J is an example of an **adiabatic invariant**. Given an Hamiltonian $H = H(p, q; \lambda)$ that depends on some external parameter $\lambda(t)$, an adiabatic invariant is a quantity which remains constant if λ changes sufficiently slowly with time. As mentioned above, in addition to varying slowly, there must also be other conditions on how ω changes in order to avoid parametric resonances. The most general condition under which this is true are difficult to define rigorously. However, it can be proven that a sufficient condition is that $H(p, q; \lambda)$ is a twice continuously differentiable function of λ. The interested reader can consult for example the book of Arnold [6], in particular Section 52E, and Sections 49, 50 and 51 of [7].

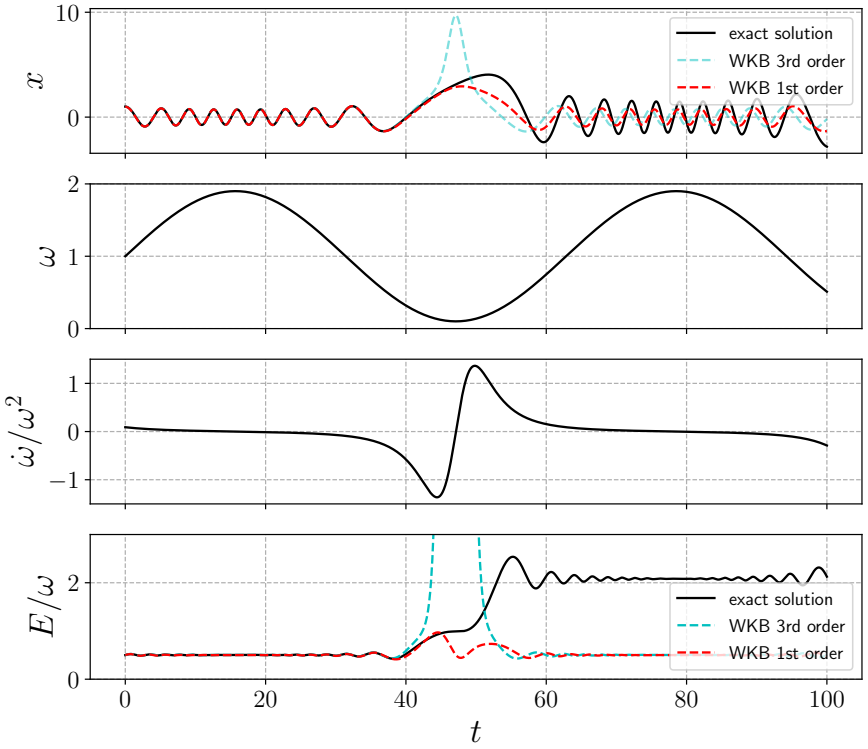

Figure 11: Comparison between the exact and the WKB approximated solution of Equation (137). The black full line is the exact solution (calculated numerically). The red dashed line is the 1st order WKB solution (Equation 147 or equivalently Equation 152 keeping only the S_0 and S_1 terms). The cyan dashed line is the 3rd order WKB solution (Equation 152 keeping terms up to S_3). The WKB solutions approximate the exact solution very well when the condition (139) is satisfied (third row in the figure), but deviates considerably when this condition is not satisfied. Note that the 3rd order solution does not do a better job than the 1st order solution when the condition (139) is violated. Note also how the WKB solution captures the small oscillations in the quantity $J = E/\omega$ (Equation 150). In this example, $\omega(t) = 1.0 + 0.9\sin(t/10)$.

A.2 Higher order WKB

Let's consider again

$$\ddot{x} + \omega^2(t)x = 0 \tag{151}$$

where $\omega(t)$ varies slowly. We want to generalise the analysis of Section A.1 to higher order in the small parameter $\epsilon = \dot{\omega}/\omega^2 \ll 1$. We look for solutions of the form

$$x(t) = \exp\left[i\int_{t_0}^{t}\sum_{n=0}^{\infty} S_n(s)\mathrm{d}s\right] . \tag{152}$$

We require that the $n+1$-th term is smaller than the n-th term a factor $\sim \epsilon$, and we expect the $S_n(t)$ to be slowly varying, so we can estimate their time derivatives using $\mathrm{d}/\mathrm{d}t \sim \epsilon\omega$. Thus we have

$$S_n \sim \epsilon S_{n-1} \sim \epsilon^n S_0 \,, \tag{153}$$

$$\dot{S}_n \sim \epsilon\omega S_n \sim \epsilon^{n+1} S_0 \,. \tag{154}$$

When only S_0 and S_1 are retained, we recover the analysis of Section A.1. Taking the derivatives of (151) we obtain:

$$\dot{x}(t) = \left(i\sum_{n=0}^{\infty} S_n\right)\exp\left[i\int_{t_0}^{t}\sum_{n=0}^{\infty} S_n(s)\mathrm{d}s\right] , \tag{155}$$

$$\ddot{x}(t) = \left(i\sum_{n=0}^{\infty}\dot{S}_n - \left(\sum_{n=0}^{\infty} S_n\right)^2\right)\exp\left[i\int_{t_0}^{t}\sum_{n=0}^{\infty} S_n(s)\mathrm{d}s\right] . \tag{156}$$

Substituting these into (151) and simplifying the exponential we get:

$$i\sum_{n=0}^{\infty}\dot{S}_n - \left(\sum_{n=0}^{\infty} S_n\right)^2 + \omega^2 = 0 \,. \tag{157}$$

Using the scalings (153)-(154) and equating order by order in powers of ϵ, we obtain the following recursive relations:

$$S_0^2 = \omega^2 \tag{158}$$

$$i\dot{S}_0 - 2S_0S_1 = 0 \tag{159}$$

$$i\dot{S}_{n-1} - \sum_{j=0}^{n} S_j S_{n-j} = 0 \qquad (n \geq 2) \tag{160}$$

The first few terms are

$$S_0 = \pm\omega\,, \tag{161}$$

$$S_1 = \frac{i\dot{\omega}}{2\omega} = \frac{i}{2}\frac{\mathrm{d}\log\omega}{\mathrm{d}t}\,, \tag{162}$$

$$S_2 = \pm\left[\frac{3\dot{\omega}^2}{8\omega^3} - \frac{\ddot{\omega}}{4\omega^2}\right]\,, \tag{163}$$

$$S_3 = i\left[-\frac{\dddot{\omega}}{8\omega^3} + \frac{3\dot{\omega}\ddot{\omega}}{4\omega^4} - \frac{3\dot{\omega}^3}{4\omega^5}\right] = i\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{3\dot{\omega}^2}{16\omega^4} - \frac{\ddot{\omega}}{8\omega^3}\right]\,. \tag{164}$$

It can be verified that if we keep only S_0 and S_1 we recover the result of A.1. Note that for $\omega > 0$, S_n is purely imaginary for n odd and purely real for n even. This is also true in general at all orders, as can be easily verified by induction. Note also that S_1 and S_3 can be written as a total derivative. It can be demonstrated that this is true for all terms with n odd (see Chapter 10 of [5]).

Using (152) and Equations (161)-(164), we can write the general solution of Equation (151) to third order in the WKB approximation as:

$$x(t) = C_1 \exp\left[\log(\omega^{-1/2}) + \frac{3\dot{\omega}}{16\omega^4} - \frac{\ddot{\omega}}{8\omega^3}\right] \exp\left[i\int_{t_0}^{t}\left(\omega + \frac{3\dot{\omega}^2}{8\omega^3} - \frac{\ddot{\omega}}{4\omega^2}\right)\mathrm{d}s\right] \tag{165}$$

$$+ C_2 \exp\left[\log(\omega^{-1/2}) + \frac{3\dot{\omega}}{16\omega^4} - \frac{\ddot{\omega}}{8\omega^3}\right] \exp\left[-i\int_{t_0}^{t}\left(\omega + \frac{3\dot{\omega}^2}{8\omega^3} - \frac{\ddot{\omega}}{4\omega^2}\right)\mathrm{d}s\right] \tag{166}$$

where C_1 and C_2 are arbitrary complex constants.

Figure 11 compares the first order solution obtained by keeping only the S_0 and S_1 terms with the third order solution obtained keeping terms up to S_3 included. In this example, the third order solution gives a better approximation than the first order solution at $t \lesssim 40$ (this cannot be seen from the figure, but it can seen if one zooms-in). However, it can be seen the third order solution gives a much worse approximation at $40 \lesssim t \lesssim 60$, where the parameter $\dot{\omega}/\omega^2$ is not small.

This is indicative of a general problem regarding the convergence of the WKB expansion (152). It turns out that for a given $\omega(t)$, the series $\sum_{n=0}^{\infty} S_n$ usually diverges and grows without bounds. It does not converge to the exact solution if we keep more and more terms. If we truncate the series truncated after more and more terms, one typically obtains result that improve up to some maximal accuracy, and then become worse. For this reason, one should be very careful when considering higher order WKB approximations. They may give results that are worse than the 1st-order approximation. A more detailed discussion of these issues can be found for example in Chapter 10 of [5].

B Angular momentum

An equation that expresses the conservation of angular momentum in a fluid disc can be obtained from Eq. (2). Multiplying the azimuthal component of this equation by R, using standard cylindrical coordinates (R, θ, z) and rearranging gives:

$$\frac{\partial(l_z)}{\partial t} + \nabla \cdot \mathbf{F}_J = -\rho \frac{\partial \Phi}{\partial \theta} \,, \tag{167}$$

where

$$l_z = \rho R v_\theta \,, \tag{168}$$

$$\mathbf{F}_J = R\left(\rho v_\theta \mathbf{v} + P\hat{\mathbf{e}}_\phi\right) \,. \tag{169}$$

The quantity l_z is the angular momentum per unit volume, while $\mathbf{F}_J$ is the flux of angular momentum, which is the sum of contributions due to bulk motions of the gas and pressure forces. The term $\rho\partial\Phi/\partial\theta$ is a source term representing the changes in angular momentum due to torques from the external potential. When $\partial\Phi/\partial\theta = 0$, the total angular momentum of the system is conserved. Indeed, the only agent that can change the total angular momentum in our problem is the external bar potential.

B.1 Transport in discs

Integrating Eq. (167) over the volume V of a cylinder of radius R_0 and using the divergence theorem,[13] we obtain the following equation for the rate of change of the total angular momentum contained within the cylinder:

$$\frac{\partial L_z}{\partial t} = -F_\mathrm{A} - F_\Phi \,, \tag{170}$$

where

$$L_z = \int_V \rho R v_\theta \, \mathrm{d}V \,, \tag{171}$$

is the total z angular momentum contained inside the cylinder, and

$$F_\mathrm{A} = R^2 \int_{-\infty}^{\infty} \mathrm{d}z \int_0^{2\pi} \mathrm{d}\theta \rho v_\theta v_R \tag{172}$$

$$F_\Phi = \int_V \rho \frac{\partial \Phi}{\partial \theta} \, \mathrm{d}V \,, \tag{173}$$

[13] The divergence theorem states that for any vector-valued function $\mathbf{F}(\mathbf{x})$:

$$\int_V \mathrm{d}V \, \nabla \cdot \mathbf{F} = \oint_S \mathrm{d}\mathbf{S} \cdot \mathbf{F}(\mathbf{x}) \,.$$

are the fluxes of angular momentum in and out of the cylinder.

Eq. (170) states that the change in the total angular momentum of the gas contained within the cylinder is the sum of two contributions: F_A, the angular momentum flux due to advection, and F_Φ, the gravitational torques from the external bar potential. $F_\mathrm{A} > 0$ means that material inside the cylinder is losing angular momentum. Notice that even in a steady-state, in which single fluid elements neither gain nor lose angular momentum on average, it is nevertheless possible that $F_\mathrm{A} \neq 0$. This can happen if fluid elements carry more angular momentum on their outward journey (as they are exiting the cylinder) than on their return. This type of transport has been named *lorry transport* by Lynden-Bell & Kalnajs 1972, who explained how fluid elements can "transport angular momentum just as a system of lorries can transport coal without accumulating a growing store on the lorries themselves".

B.2 Angular momentum of waves in perturbed axisymmetric disc

Calculating the angular momentum of waves requires some care. In linear theory, we keep only first order terms *in the equations of motion* and neglect second-order terms, but when dealing with *angular momentum fluxes*, the first order terms are generally absent and we need to care about second order terms. Consider a general perturbation around an axisymmetric steady-state gas disc:

$$\rho(R,\theta,t) = \rho_0(R) + \Delta\rho(R,\theta,t) \tag{174}$$
$$v_\theta(R,\theta,t) = v_{\theta 0}(R) + \Delta v_\theta(R,\theta,t) \tag{175}$$
$$v_R(R,\theta,t) = \Delta v_R(R,\theta,t) \tag{176}$$

where the steady-state quantities contain *all orders* (not just the first-order perturbation, i.e. $\Delta\rho = \rho_1 + \rho_2 + \dots$). Consider the azimuthal average of Eq. (167):

$$\frac{\partial\langle l_z\rangle}{\partial t} + \frac{1}{R}\frac{\partial\langle RF_{JR}\rangle}{\partial R} = -\langle\rho\frac{\partial\Phi}{\partial\theta}\rangle\,, \tag{177}$$

where $\langle\cdot\rangle = (2\pi)^{-1}\int_0^{2\pi}\cdot\,\mathrm{d}\theta$ denotes the azimuthal average. Now consider the following decomposition, which is exact to *all orders*:

$$l_z \equiv l_m + l_w = \rho R v_{\theta 0} + \rho R\Delta v_\theta\,, \tag{178}$$
$$F_{JR} \equiv F_{Jm} + F_{Jw} = R\rho v_{\theta 0}\Delta v_R + R\rho\Delta v_\theta\Delta v_R\,. \tag{179}$$

Using the continuity equation (9) it is easy to see that

$$\frac{\partial\langle l_m\rangle}{\partial t} + \frac{1}{R}\frac{\partial\langle RF_{Jm}\rangle}{\partial R} = 0\,. \tag{180}$$

and therefore subtracting this last equation from (177):

$$\frac{\partial\langle l_w\rangle}{\partial t}+\frac{1}{R}\frac{\partial\langle RF_{Jw}\rangle}{\partial R}=-\langle\rho\frac{\partial\Phi}{\partial\theta}\rangle \tag{181}$$

Thus, l_m and l_w separately satisfies conservation equations, correct to all orders. We identify l_w with the 'wave' angular momentum. Eq. (181) says that this is where the angular momentum deposited or subtracted by the external potential goes. The wave angular momentum is therefore the relevant quantity if we are interested in calculating net gain/loss of angular momentum in the disc. The equation with l_m can be thought as a separate conservation equation, related to changes in angular momentum related to a net flux of mass (note that F_{Jm} is $Rv_{\theta 0}$ times the radial flux of mass).

B.3 Transport in perturbed discs

Using the same expansion of the previous section (174)-(176), the advection flux Eq. (172) can be rewritten, correct *to all orders*, as:

$$F_A=R^2\int_{-\infty}^{\infty}\mathrm{d}z\int_0^{2\pi}\mathrm{d}\theta\rho\left(\Delta v_\theta\Delta v_R\right)+R^2\int_{-\infty}^{\infty}\mathrm{d}z\int_0^{2\pi}\mathrm{d}\theta\rho\left(v_{\theta 0}\Delta v_R\right) \tag{182}$$

$$\equiv F_w+F_m \tag{183}$$

where F_m is $Rv_{\theta 0}$ times the radial mass flux at radius R:

$$F_m=Rv_{\theta 0}\left[R\int_{-\infty}^{\infty}\mathrm{d}z\int_0^{2\pi}\mathrm{d}\theta\rho\Delta v_R\right], \tag{184}$$

The radial flux of angular momentum can therefore be divided in two parts. F_w is a "wave" part, associated with the conservation equation (181). To calculate this term up to second-order, we can replace ρ with ρ_0 and use the first-order solutions for Δv_θ and Δv_R. The F_m is a non-wave part associated with (180) that is non-zero only if there is a net radial mass flux. Although we will not prove it, calculations retaining the second-order terms show that the mass flux is generally zero, because the fluid elements just slosh back and forwards during each cycle.

C An exactly solvable toy model: excitation of 1D waves by an oscillating Gaussian potential

Here we describe a toy model that shares various similarities with the actual problem studied in the main text but has the advantage that it can be solved fully

analytically. This toy model is helpful for understanding why the coupling of the external potential and the fluid, and therefore the amplitude of the waves excited by an external potential, can depend very strongly on the gas sound speed.

We can draw the following correspondences between this toy problem and the problem studied in the main text (i) 1D sound waves correspond to spiral density waves; (ii) The 1D Gaussian potential corresponds to the bar potential; (iii) The linear momentum of plane waves plays a similar role to the angular momentum of the spiral density waves; (iv) Eq. (191) is the analog of Eq. (56).

C.1 Statement of the problem

Consider a 1D isothermal fluid at rest with uniform density ρ_0. Our goal is to study the waves excited in this medium by a "small" time-varying external potential $\Phi(x,t)$.

The equations of motion of this system are the same as Eqs. (1), (2) where the gradient is replaced by $\mathrm{d}/\mathrm{d}x$ since the problem is one dimensional. We assume an isothermal equation of state, $P = c_\mathrm{s}^2\rho$, where c_s is a constant. We linearise these equations around the background state by writing $\rho(x,t) = \rho_0 + \rho_1(x,t)$ and $v(x,t) = v_1(x,t)$ and keeping only the first-order terms in the quantities with subscript 1. We obtain:

$$\partial_t\rho_1 + \rho_0(\partial_x v_1) = 0\,, \tag{185}$$

$$\partial_t v_1 = -c_\mathrm{s}^2\frac{\partial_x\rho_1}{\rho_0} - \partial_x\Phi\,. \tag{186}$$

Without loss of generality, we can write all variables as

$$F(x,t) = \tilde{F}(x)\exp(-i\omega t)\,. \tag{187}$$

Where $\tilde{F}$ are complex quantities. We use complex notation for mathematical convenience, but it is understood that the physical quantities are given by the real part. Substituting all perturbation variables in the form (187) into (185) and (186), and omitting the symbol $\tilde{}$ hereafter for simplicity of notation, we obtain

$$-i\omega\rho_1 + \rho_0(\partial_x v_1) = 0 \tag{188}$$

$$-i\omega v_1 = -c_\mathrm{s}^2\frac{\partial_x\rho_1}{\rho_0} - \partial_x\Phi(x) \tag{189}$$

Isolating v_1 from (189) and introducing the variable $s_1 = \rho_1/\rho_0$ we have

$$v_1 = \frac{c_\mathrm{s}^2(\partial_x s_1) + (\partial_x\Phi)}{i\omega} \tag{190}$$

Substituting (190) into (188) we obtain an ODE for s_1:

$$\boxed{\omega^2 s_1 + c_{\rm s}^2 \frac{{\rm d}^2}{{\rm d}x^2} s_1 = F} \tag{191}$$

where

$$F(x) = -\partial_x^2 \Phi \,. \tag{192}$$

Eq. (191) is the equation of a forced harmonic oscillator. Now consider an oscillating Gaussian potential of the form:

$$\Phi(x,t) = \Phi_1 \exp\left[-\left(\frac{x}{x_0}\right)^2\right] \exp(-i\omega_0 t) \tag{193}$$

where Φ_1 is the strength of the potential, x_0 is the width of the Gaussian perturbation, ω_0 is the oscillation frequency. Since in the linear approximation there is no coupling between modes at different frequencies, only modes with frequency $\omega = \omega_0$ will be excited by this potential. Hence we assume $\omega = \omega_0$ hereafter. Introducing the dimensionless coordinate $\xi = x/x_0$ and using (193), Eq. (191) becomes:

$$\boxed{\frac{{\rm d}^2 s_1}{{\rm d}\xi^2} + a^2 s_1 = b a^2 K(\xi)} \tag{194}$$

where

$$K(\xi) = (1 - 2\xi^2) \exp\left[-\xi^2\right] \,, \tag{195}$$

and we have introduced the following dimensionless parameters:

$$a = \frac{\omega_0 x_0}{c_{\rm s}} \,, \tag{196}$$

$$b = \frac{2\Phi_1}{\omega_0^2 x_0^2} \,. \tag{197}$$

The parameter a is the inverse of the sound speed, normalised with the typical scale-length and frequency of the problem. The parameter b is the normalised strength of the external potential.

C.2 Analytical solution

The general solution of Eq. (194) is

$$s_1(\xi) = C_1 \exp(ia\xi) + C_2 \exp(-ia\xi) + bW(\xi, a) \,, \tag{198}$$

where C_1 and C_2 are arbitrary constants and

$$W(\xi, a) = -\frac{1}{2}a^2 e^{-\xi^2} + X(\xi, a)\,, \tag{199}$$

$$X(\xi, a) = -i\alpha\left[e^{ia\xi}\ \mathrm{erf}\left(\xi + i\frac{a}{2}\right) - e^{-ia\xi}\ \mathrm{erf}\left(\xi - i\frac{a}{2}\right)\right]\,, \tag{200}$$

$$\alpha = \frac{\sqrt{\pi}}{8}a^3 e^{-a^2/4}\,. \tag{201}$$

Here, $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,\mathrm{d}t$ is the error function, which is defined for complex argument z (to evaluate the integral, you can choose any integration path in the complex plane that leads to z). Note however that the functions X and W are real because the erf function has the properties $\mathrm{erf}(\overline{z}) = \overline{\mathrm{erf}(z)}$ and $\mathrm{erf}(-z) = \mathrm{erf}(z)$, where the bar denotes the complex conjugate.

Figure 12 plots the function $W(\xi, a)$ for various values of a. In the limit $\xi \to \pm\infty$ we have that $\mathrm{erf}(\xi + ic/2) \to \pm 1$ for any fixed c, so $W(\xi, a) \to \mp i\alpha[e^{ia\xi} - e^{-ia\xi}]$. Therefore $W(\xi, a)$ becomes a plane wave when $\xi \to \pm\infty$ (as one would expect). In the limit $a \to \infty$ the W tends to the forcing term K on the right-hand-side of Eq. 12.

What is the amplitude of the waves that are excited by the external potential (193)? In order to answer this question we have to determine the constants C_1 and C_2 in Eq. (198) by imposing appropriate boundary conditions. Causality requires that for large $|x|$ the waves propagate "away" from the potential (this is the same boundary condition that is used to derive retarded potential in electrodynamics). A solution in which the waves come from infinity towards the potential would instead require a source at infinity, which is unphysical. Therefore, we impose that the solution propagates towards positive ξ as $\xi \to +\infty$ and towards negative ξ as $\xi \to -\infty$. To see in which direction the solution (198) is travelling, we look at its time-dependence by reattaching the factor $\exp(-i\omega_0 t)$ to it:

$$s_1(\xi, t) = \left[C_1 e^{ia\xi} + C_2 e^{-ia\xi} + bW(\xi, a)\right] e^{-i\omega_0 t}\,. \tag{202}$$

A plane wave of the form $e^{ia\xi - i\omega_0 t}$ ($e^{-ia\xi - i\omega_0 t}$) travels towards positive (negative) ξ. The solution that satisfies our "radiation" boundary conditions is then:

$$s_1(\xi, t) = b\left[-i\alpha e^{ia\xi} - i\alpha e^{-ia\xi} + W(\xi, a)\right] e^{-i\omega_0 t}\,. \tag{203}$$

This solution tends to $s_1(\xi, t) \to -2i\alpha b e^{\pm ia\xi} e^{-i\omega_0 t}$ for $x \to \pm\infty$. Thus, the potential excites waves with an amplitude of

$$\boxed{A = 2b\alpha = b\frac{\sqrt{\pi}}{4}a^3 e^{-a^2/4}.} \tag{204}$$

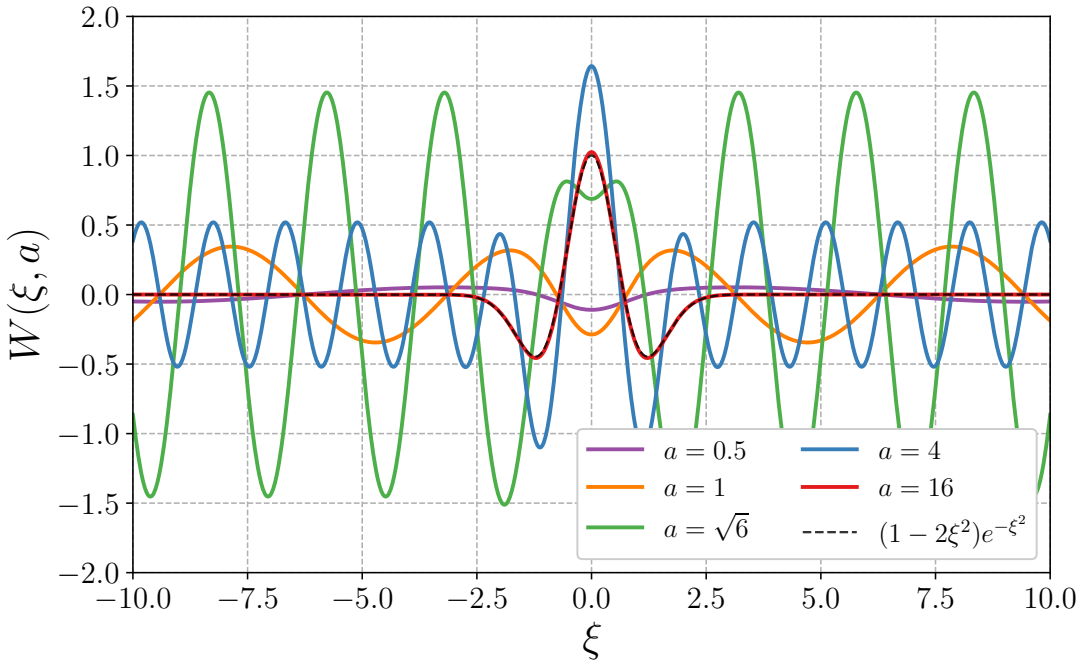

Figure 12: The function $W(\xi, a)$ defined by Equation (199) for various values of a.

The key point here is that the amplitude A of the excited waves has an extremely strong dependence on the sound speed $c_s \propto 1/a$. The amplitude A tends to zero *very* quickly both for $a \rightarrow 0$ ($c_s \rightarrow \infty$) and $a \rightarrow \infty$ ($c_s \rightarrow 0$), and (for fixed b) has a maximum in between at $a = \sqrt{6}$. This has a simple physical interpretation. The coupling between the external potential and sound waves in a uniform medium is strongest when the wavelength of free sound waves at the frequency of the external potential is comparable to the scale-length of the potential. This is indeed what happens, as can be seen as follows. The dispersion relation of free sound waves travelling in a uniform medium is $\omega = c_s k$, where $k = 2\pi/\lambda$ is the wavenumber and λ is the wavelength. Therefore, the wavelength of free sound waves travelling in a uniform medium at frequency ω_0 is $\lambda_0 = 2\pi c_s/\Omega_0$. The parameter $a = 2\pi x_0/\lambda_0$ is, apart from a numerical constant, the ratio between the scale-length of the potential and the wavelength of free sound waves at that frequency. Thus, we expect the potential to be most effective in driving waves when a is of order unity.

D Details on the calculations of waves excited at a sharp edge

D.1 Derivation of Eqs. (128)-(130)

In a neighbourhood of the point $R = R_{\text{out}}$ we can approximate Eq. (127) as

$$\begin{aligned} g_1(R) = e^{iKR}\left[A_1 - \frac{Q_{\text{out}}}{2K^2}e^{-iKR_{\text{out}}} - \frac{i}{2K}\int_{R_0}^{R_{\text{out}}} Q(s)e^{-iKs}\mathrm{d}s\right] + \\ + e^{-iKR}\left[A_2 - \frac{Q_{\text{out}}}{2K^2}e^{iKR_{\text{out}}} + \frac{i}{2K}\int_{R_0}^{R_{\text{out}}} Q(s)e^{iKs}\mathrm{d}s\right] + \\ + \frac{Q_{\text{out}}}{K^2} \qquad (\text{neighbourhood of } R_{\text{out}}) \,, \end{aligned} \tag{205}$$

where $Q_{\text{out}} = Q(R_{\text{out}})$. Since the waves are travelling outwards at $R = R_{\text{out}}$, the term proportional to e^{iKR} should vanish. This condition gives Eq. (128).

Similarly, in a neighbourhood of the point $R = R_{\text{in}}$ we can approximate Eq. (127) as

$$\begin{aligned} g_1(R) = e^{iKR}\left[A_1 - \frac{Q_{\text{in}}}{2K^2}e^{-iKR_{\text{in}}} - \frac{i}{2K}\int_{R_0}^{R_{\text{in}}} Q(s)e^{-iKs}\mathrm{d}s\right] + \\ + e^{-iKR}\left[A_2 - \frac{Q_{\text{in}}}{2K^2}e^{iKR_{\text{in}}} + \frac{i}{2K}\int_{R_0}^{R_{\text{in}}} Q(s)e^{iKs}\mathrm{d}s\right] + \\ + \frac{Q_{\text{in}}}{K^2} \qquad (\text{neighbourhood of } R_{\text{in}}) \,, \end{aligned} \tag{206}$$

where $Q_{\text{in}} = Q(R_{\text{in}})$. Since the waves are travelling inwards at $R = R_{\text{in}}$, the term proportional to e^{-iKR} should vanish. This condition gives Eq. (129).

Substituting Eqs. (128) and (129) into Eq. (205) and Eq. (206) respectively we find

$$\begin{aligned} g_1(R) = e^{-iKR}\left[\frac{i}{2K}\int_{R_{\text{in}}}^{R_{\text{out}}} Q(s)e^{iKs}\mathrm{d}s\right] + \\ + e^{-iKR}\left[\frac{Q_{\text{in}}}{2K^2}e^{iKR_{\text{in}}} - \frac{Q_{\text{out}}}{2K^2}e^{iKR_{\text{out}}}\right] + \\ + \frac{Q_{\text{out}}}{K^2} \qquad (\text{neighbourhood of } R_{\text{out}}) \,, \end{aligned} \tag{207}$$

and

$$g_1(R) = e^{iKR}\left[\frac{i}{2K}\int_{R_{\rm in}}^{R_{\rm out}} Q(s)e^{-iKs}\mathrm{d}s\right] + \\ + e^{iKR}\left[\frac{Q_{\rm out}}{2K^2}e^{-iKR_{\rm out}} - \frac{Q_{\rm in}}{2K^2}e^{-iKR_{\rm in}}\right] + \\ + \frac{Q_{\rm in}}{K^2} \qquad \text{(neighbourhood of } R_{\rm in}) \ . \tag{208}$$

Matching Eqs. (207) and (208) with Eq. (126), one obtains Eq. (130).

D.2 Derivation of Eq. (131)

We approximate Eq. (130) as follows. First, we neglect the terms proportional to $Q_{\rm in}$ and $Q_{\rm out}$ because Q varies rapidly at radii $R_{\rm in} < R < R_{\rm out}$. We obtain

$$|C_{\rm in}| \simeq |C_{\rm out}| \simeq \frac{1}{2K^{1/2}}\left|\int_{R_{\rm in}}^{R_{\rm out}} Q(s)e^{-iKs}\mathrm{d}s\right| \ . \tag{209}$$

Second, the exponential $\exp(-iKs)$ is nearly constant as we have assumed $|R_{\rm out} - R_{\rm in}| \sim \lambda \sim 1/K$, so we can write

$$|C_{\rm in}| \simeq |C_{\rm out}| \simeq \frac{1}{2K^{1/2}}\left|\int_{R_{\rm in}}^{R_{\rm out}} Q(s)\mathrm{d}s\right| \ . \tag{210}$$

We have

$$\int_{R_{\rm in}}^{R_{\rm out}} Q(s)\mathrm{d}s = I_1 + I_2 + I_3 + I_4 \ , \tag{211}$$

where

$$I_1 = -\int_{R_{\rm in}}^{R_{\rm out}} \mathrm{d}s \left(\frac{s\rho_0}{|D|}\right)^{1/2} \frac{\mathrm{d}^2\Phi_1}{\mathrm{d}s^2} \tag{212}$$

$$I_2 = -\int_{R_{\rm in}}^{R_{\rm out}} \mathrm{d}s \left(\frac{s\rho_0}{|D|}\right)^{1/2} \frac{\mathrm{d}}{\mathrm{d}s}\left[\log\left(\frac{s\rho_0}{D}\right)\right]\frac{\mathrm{d}\Phi_1}{\mathrm{d}s} \tag{213}$$

$$I_3 = -\int_{R_{\rm in}}^{R_{\rm out}} \mathrm{d}s \left(\frac{s\rho_0}{|D|}\right)^{1/2} \frac{2\Omega}{s\left(\Omega - \Omega_{\rm p}\right)}\frac{\mathrm{d}}{\mathrm{d}s}\left[\log\left(\frac{\rho_0\Omega}{D}\right)\right]\Phi_1 \tag{214}$$

$$I_4 = \int_{R_{\rm in}}^{R_{\rm out}} \mathrm{d}s \left(\frac{s\rho_0}{|D|}\right)^{1/2} \frac{m^2\Phi_1}{s^2} \ . \tag{215}$$

Since $|R_{\rm out}/R_{\rm in} - 1| \ll 1$ and far from the ILR the integrand is bounded, we have $I_1 \simeq I_4 \simeq 0$.

We calculate I_2 and I_3 below. The idea is to integrate by parts in order to isolate the integral of a bounded function. We have

$$\begin{aligned}
I_2 &= -2\int_{R_{\rm in}}^{R_{\rm out}} {\rm d}s \frac{\rm d}{{\rm d}s}\left(\frac{s\rho_0}{|D|}\right)^{1/2}\frac{{\rm d}\Phi_1}{{\rm d}s} = \\
&= -2\left[\left(\frac{R\rho_0}{|D|}\right)^{1/2}\frac{{\rm d}\Phi_1}{{\rm d}R}\right]_{R=R_{\rm out}} + 2\left[\left(\frac{R\rho_0}{|D|}\right)^{1/2}\frac{{\rm d}\Phi_1}{{\rm d}R}\right]_{R=R_{\rm in}} + \\
&+ 2\int_{R_{\rm in}}^{R_{\rm out}} {\rm d}s \left(\frac{s\rho_0}{|D|}\right)^{1/2}\frac{{\rm d}^2\Phi_1}{{\rm d}s^2} = \\
&= 2\left[\left(\frac{R\rho_0}{|D|}\right)^{1/2}\frac{{\rm d}\Phi_1}{{\rm d}R}\right]_{R=R_{\rm in}} ,
\end{aligned} \tag{216}$$

and

$$\begin{aligned}
I_3 &= -4\int_{R_{\rm in}}^{R_{\rm out}} {\rm d}s \left[\frac{\Omega^{1/2}\Phi_1}{s^{1/2}\left(\Omega-\Omega_{\rm p}\right)}\right]\frac{\rm d}{{\rm d}s}\left(\frac{\rho_0\Omega}{|D|}\right)^{1/2} = \\
&= -\left[\frac{4\Omega}{\Omega-\Omega_{\rm p}}\left(\frac{R\rho_0}{|D|}\right)^{1/2}\frac{\Phi_1}{R}\right]_{R=R_{\rm out}} + \\
&+ \left[\frac{4\Omega}{\Omega-\Omega_{\rm p}}\left(\frac{R\rho_0}{|D|}\right)^{1/2}\frac{\Phi_1}{R}\right]_{R=R_{\rm in}} + \\
&+ 4\int_{R_{\rm in}}^{R_{\rm out}} {\rm d}s \left(\frac{\rho_0\Omega}{|D|}\right)^{1/2}\frac{\rm d}{{\rm d}s}\left[\frac{\Omega^{1/2}\Phi_1}{s^{1/2}\left(\Omega-\Omega_{\rm p}\right)}\right] = \\
&= \left[\frac{4\Omega}{\Omega-\Omega_{\rm p}}\left(\frac{R\rho_0}{|D|}\right)^{1/2}\frac{\Phi_1}{R}\right]_{R=R_{\rm in}} ,
\end{aligned} \tag{217}$$

where we have used the fact that $\rho_0(R_{\rm in}) \gg \rho_0(R_{\rm out})$. Substituting Eqs. (216) and (217) into Eq. (211), we find

$$\int_{R_{\rm in}}^{R_{\rm out}} Q(s){\rm d}s = 2\left[\left(\frac{R\rho_0}{|D|}\right)^{1/2}\left(\frac{{\rm d}\Phi_1}{{\rm d}R} + \frac{2\Omega}{\Omega-\Omega_{\rm p}}\frac{\Phi_1}{R}\right)\right]_{R=R_{\rm in}} . \tag{218}$$

Substituting Eq. (218) into Eq. (210), we obtain Eq. (131).

Further reading

[1] Binney, J.; Tremaine, S., *Galactic Dynamics*, Princeton Series in Astrophysics, 2008

[2] Goldreich, P.; Tremaine, S., *The formation of the Cassini division in Saturn's rings*, Icarus, Vol. 34, p. 240-253, 1978

[3] Goldreich, P.; Tremaine, S., *The excitation of density waves at the Lindblad and corotation resonances by an external potential*, Astrophysical Journal, Vol. 233, p. 857-871, 1979

[4] Sormani, M.C.; Sobacchi, E.; Sanders, J.L., *Nuclear rings are the inner edge of a gap around the Lindblad Resonance*, Monthly Notices of the Royal Astronomical Society, Volume 528, Issue 4, pp.5742-5762, 2024

[5] Bender, C.M.; Orszag, S.A., *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory*, Springer, 1999

[6] Arnold, V. I. *Mathematical Methods of Classical Mechanics*, (Springer-Verlag New York, 1989)

[7] Landau, L.D., Lifshitz, E.M. *Course of theoretical physics, Volume* I*: Mechanics*, (Butterworth-Heinemann, 1982)