UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE E TECNOLOGIE

Ph.D. course in Physics, Astrophysics and Applied Physics,
Cycle XXXI

The evolution of binary systems in gaseous environments

Supervisor:
Prof. Giuseppe Lodato

Ph.D. Candidate:
Enrico Ragusa
Cycle XXXI

Referees:
Prof. Zoltán Haiman
Prof. Gordon I. Ogilvie

Examiners:
Prof. Wilhelm Kley
Dr. Chris J. Nixon

Codice PACS: 98.62.-g, 96.35.-j

Anno Accademico 2017-2018

A mio fratello Guido.
...Purchè smetta di dire che
coloro quaderni!

Contents

Disc-satellite interaction in astronomy

"...Ground Control to Major Tom
Take your protein pills and put your helmet on..."

David Bowie, Space Oddity

Systems where a binary, that is two gravitationally bound objects (with mass M_1 and M_2, respectively) orbiting their centre of mass, interacts with the surrounding gas or dust are extremely common in the Universe and involve a wide variety of different astrophysical objects (star + star, black hole + black hole, star + planet, or planet + moon[1]). Among them, protoplanetary systems and black hole binaries (BHBs) are currently capturing the attention of the scientific community.

These two types of systems are extremely different in their nature and relevant parameters. On the one hand, protoplanetary systems consist of one, or more, planets (mass ratio $M_2/M_1 \sim 10^{-6} - 10^{-2}$), or a secondary protostellar companion ($M_2/M_1 \approx 1$), orbiting a protostar ($M_1 \approx 1\,\mathrm{M}_\odot$) in a binary configuration embedded in a relatively thick accretion disc (with aspect ratio $H/R \approx 0.1$). Such discs form during the process of star formation, due to angular momentum conservation, and they are thought to be the site of planet formation.

On the other hand, black hole binary systems consist of two gravitationally bound black holes surrounded by a relatively thin accretion disc $H/R \approx 10^{-3}$. They can be found both in the supermassive (SMBHBs: $M_1 \sim M_2 \approx 10^8\,\mathrm{M}_\odot$) and in the stellar (SBHBs: $M_1 \sim M_2 \approx 10\,\mathrm{M}_\odot$) mass regime. Supermassive black hole binaries are expected to be hosted in the gas rich central regions of galaxies as a consequence of galaxy mergers (Begelman et al. 1980). Stellar mass black hole binaries are instead the endpoint of the evolution of massive stellar binaries. Also in this case, a significant amount of gas can be still present in the

[1]Such as Saturn, its moons and the system of rings surrounding it, see Fig. 1.

surroundings of such binaries: as a consequence of outflows during the life of the stellar progenitors (de Mink & King 2017; Martin et al. 2018) but also because stellar black hole binaries can form in gas rich discs powering active galactic nuclei (AGN) (Stone et al. 2017), or be dragged into them (Bartos et al. 2017). BHBs are expected to merge as a consequence of the dissipation of their mechanical energy through the emission of gravitational waves.

The heating mechanisms and thus the disc electromagnetic emission (EM) of these systems depend on very different physical phenomena. Young stars are typically surrounded by relatively cold ($T \sim 10-1000\,\mathrm{K}$) passive discs (viscous dissipation does not dominate the heating of the disc). The EM emission comes mainly from optically thick gas emission lines (like rotational and roto-vibrational lines of CO isotopologues and of other molecular species) and from the dust component of the disc: in particular, large dust grains (mm-sized) are heated by the central protostar and provide a thermal emission in the radio band, while small grains (μm-sized) scatter the infrared light from the protostar. Black hole binaries are instead surrounded by very hot (up to $T \sim 10^4 - 10^7\,\mathrm{K}$), ionized, active discs (viscous dissipation is the heating source of the disc). The emission mechanism associated with it is thermal optically thick emission (black body) from the gas component, implying that the accretion dynamics modulates the EM luminosity of these sources.

Despite their very different nature and EM appearance, the dynamics of both protoplanetary discs and BHBs systems is very similar and can be described in one unique theoretical framework: the disc-satellite interaction theory.

This project is meant to deepen our knowledge of the theory of circumbinary discs, approaching it in a multidisciplinary way from both the protoplanetary and the BHBs perspective.

Motivation: protoplanetary discs and BH binaries, different manifestations of the same dynamics

We are currently living a golden age for protoplanetary physics and gravitational wave astronomy. The imaging of protoplanetary discs with unprecedented detail, allowed by the advent of the Atacama Large Millimeter Array (ALMA) and of the Spectro-Polarimetric High-contrast Exoplanet REsearch (SPHERE) instrument mounted on the Very Large Telescope (VLT), revealed a large sample of very peculiar structures: such as gaps, cavities, shadows, spirals and other non-axisymmetric features). This represents one of the biggest achievements in observational astronomy of the last decades, offering the opportunity to compare the theoretical models with multi-wavelength observations, and thus sampling different components of the surrounding material: small dust, large dust, gas.

At the opposite side of the mass spectrum, stellar and supermassive black hole binaries instead are still very elusive from an observational point of view. Apart from a number of candidates inferred from periodic luminosity variations (Catalina Real-time Transient Survey, CRTS, Drake et al. 2009; Graham et al. 2015 and Palomar Transient Factory PTF, Rau et al. 2009; Charisi et al. 2016) or kinematical signatures in AGNs (Bogdanović 2015; D'Orazio et al. 2015; Wang et al. 2017),

no robust electromagnetic (EM) detection has occurred so far. Nevertheless, the advent of gravitational wave (GW) astronomy has provided a direct confirmation of the existence of stellar black hole binaries (Abbott et al. 2016c). The gas ejected as an outflow during the lifetime of the stellar progenitors of the black holes hopefully provides a possible EM signature of the merger (de Mink & King 2017). The recent first detection of a GW signal (Abbott et al. 2017c) associated with an EM counterpart (Savchenko et al. 2017b) constitutes a milestone, showing for the first time that we have the technology to associate GW events with their EM counterparts. The presence of gas surrounding the BHB is more established in the supermassive case, for which the detection of GW signals is expected to occur with the launch of the LISA mission (Laser Interferometer Space Antenna) in 2034.

In all such contexts, the disc-satellite interaction on the one hand determines the disc appearance: detectable in the morphology of protostellar discs and in the EM counterpart to GW emission, respectively; while on the other hand it can change the orbital properties of the binary: in terms of migration, eccentricity evolution, spin alignment; properties that can be detected both for planets, via planet detection surveys, and BHs based on the shape of the GW signal.

While the thermal properties and emission features may differ for protoplanetary and BHB systems, the dynamics is well captured within the disc-satellite interaction theory in both cases (Lin & Papaloizou 1979; Goldreich & Tremaine 1980). A secondary object (M_2) orbiting a primary (M_1) produces a non-vanishing time-dependent quadrupole moment that perturbs the axial symmetry of the gravitational potential that would be produced by the primary alone. This produces a torque, exerted by the satellite on the disc, that promotes the exchange of both energy and angular momentum between the binary and the disc.

The disc-satellite interaction occurs at specific disc locations, called "resonances". This mechanism is referred to as "tidal torque" and mediates the disc-satellite "resonant interaction" (Goldreich & Tremaine 1980). The tidal torque acts against the viscous torque, that induces disc spreading and accretion on to the primary/binary. As a consequence, the satellite might carve a gap or a cavity in the disc, but it also produces spiral structures, it promotes the development of eccentric fluid trajectories and the formation of non-axisymmetric features in the disc[2]. Furthermore, as a consequence of Newton's third law, the disc exerts a back reaction torque on the satellite that causes it to migrate (evolution of the semi-major axis) and to modify its orbital eccentricity. The spectacular image in Fig. 1 shows an example of this mechanism in action in Saturn's rings.

Besides the resonant mechanism, but less discussed and studied, secular interaction (Teyssandier & Ogilvie 2016) affects the evolution of the system at very long timescales (10^3–10^5 orbits). It involves the interaction of the satellite with the disc as a whole (contrary to resonant interaction that takes place only at resonant locations), providing an exchange of angular momentum, but not of energy, between the disc and the satellite. This produces periodic oscillations of the eccentricity and the precession of the pericentres of both binary and disc orbits (Miranda et al.

[2]We refer to Fig. 1.3 in Chap. 1 for some examples of non-axisymmetric structures detected by ALMA and SPHERE

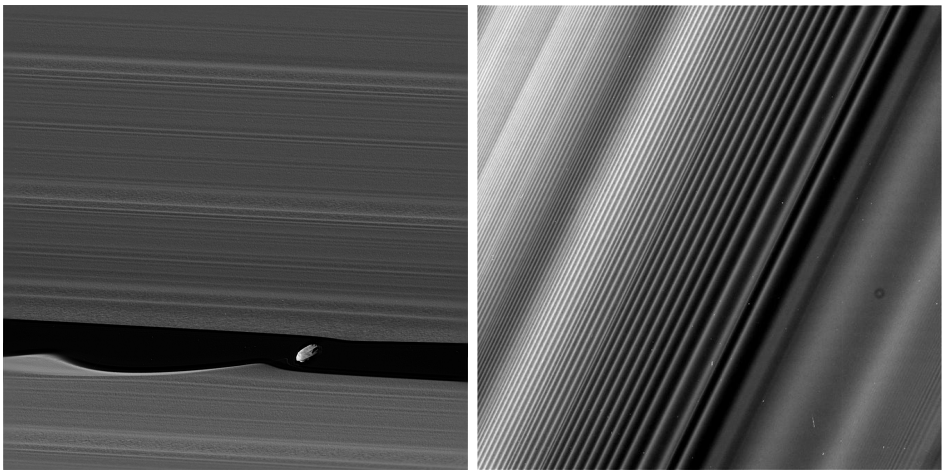


Figure 1: Two pictures of Saturn's rings taken by the Cassini probe. Left Panel: The moon Daphni carving a gap structure at the outer edges of Saturn's rings (image credit:NASA/JPL-Caltech/Space Science Institute). Right Panel: density perturbations in Saturn's rings density known as "Janus 2:1 spiral density wave" produced by the moon Janus in Saturn's B ring (image credit:NASA/JPL-Caltech/Space Science Institute).

2017; Thun et al. 2017; Ragusa et al. 2018). The basic principle of its action is analogous to that producing periodic eccentricity oscillations in multiple-planet systems in the context of celestial mechanics. This type of interaction received less attention than it deserves, but it appears to have a strong impact on the long timescale eccentricity evolution of both satellites and discs.

The entire analytical treatment in the disc-satellite interaction theory relies mainly on the perturbation analysis of the effects produced by the addition of a small binary perturber in a Keplerian potential. As soon as this assumption is not satisfied, the analytical theory fails to give quantitative predictions. For this reason, still much work is required in order to better constrain how the system parameters affect the dynamics using all the approaches at our disposal: analytical, numerical and observational.

Aim and outline of the thesis

In this thesis I aim to investigate and discuss three relevant problems in the context of disc-satellite interaction, namely: the suppression of accretion rate in binary systems, the mutual evolution of satellite/disc eccentricity and the formation of disc non-axisymmetric structures. As mentioned before, these specific problems are relevant in very different contexts of astronomy, from black hole binaries to planetary systems, but they all share the same physical processes in action.

With this in mind, all these subjects will be contextualized and discussed with explicit reference to the problems they are relevant for. Due to its interdisciplinarity, the thesis is divided in 3 parts as follows:

- **Part I** provides a broad introduction about relevant concepts that will be useful throughout the thesis, regarding both black holes and planetary systems. Chap. 1 introduces the two body problem and the process of formation of both black holes and planets. Chap. 2 introduces the formalism of accretion disc physics. Chap. 3 provides a thorough discussion of disc-satellite interaction theory leading to the derivation of the analytical expression of the tidal torque exerted by the satellite on the disc (and vice versa). Finally, Chap. 4 provides a brief review of the effects of the tidal torque on the disc structure (gap formation), satellite migration and evolution of its eccentricity.

- **Part II** is dedicated to the disc-satellite interaction in black hole binary systems. The final goal of this part is to discuss the accretion rate in these systems: the presence of the satellite produces dynamical effects that alter the accretion mechanism with respect to the predictions of classic disc accretion theory. In particular, depending on the disc thermal and viscous properties, the tidal torque produced by the satellite acts as a dam against the inward motion of the gas, possibly producing a reduction of the accretion rate on to the binary. A number of numerical simulations have shown that the accretion rate does not change in binary systems. However, mainly for computational reasons, the dam effect of the tidal torque in thin discs have never been investigated.

  This problem is particularly relevant for black hole binary systems: since they are surrounded by very thin discs, the accretion rate on to them might be strongly reduced, with important consequences for their detectability and dynamical evolution.

  To this end, Chap. 5 introduces black hole binaries, their formation models, the relevance of accretion rate in determining the electromagnetic output of black hole binaries, and the dynamical evolution of their characteristic parameters. Chap. 6 presents the results we published in the paper by Ragusa et al. (2016) obtained in this context. In particular, we find that in thin discs the secondary acts as a dam for the material that is not allowed to accrete on to the black holes. This implies that the electromagnetic luminosity of these systems is likely to be strongly suppressed.

- **Part III** is dedicated to the disc-satellite interaction in young stellar objects. This part focuses on the mutual evolution of the satellite and disc orbital properties. The torque exerted by a satellite on to the disc has been observed to trigger the formation of non-axisymmetric structures into it. At the same time, as a consequence of Newton's third law, the back reaction torque causes the evolution of the eccentricity and semi-major axis (migration) of the satellite. These aspects of the dynamics of disc-satellite systems are particularly relevant in young stellar objects. On the one hand, these studies might help us to better understand the mechanisms originating the peculiar features, such as horseshoes, spirals, gaps that have been broadly observed in young protoplanetary discs after the advent of ALMA and SPHERE. On the other, the migration and eccentricity evolution of planets, caused by

their interaction with the accretion disc where they formed, might be one of physical processes shaping the architecture of solar systems.

In Chap. 7 we discuss the results we obtained in Ragusa et al. (2018) about the mutual evolution of satellite and disc eccentricity, emphasizing the role of secular disc-satellite interaction and contextualizing the results in the field of planet formation and evolution. In particular, we investigate the possibility that disc-planet interaction constitutes the mechanism able to excite the observed high orbital eccentricity of some young hot-Jupiters, i.e. massive giant planets orbiting their host star on short period orbits (~ 10 d). We find a counterintuitive relation between the evolution of the planet eccentricity and the disc mass. In particular, less massive discs appear to be able to excite higher level of eccentricity of the planet at late times, in contrast with previous results in the literature.

Chap. 8 discusses observations and theoretical interpretations of transition discs, i.e. systems where large dust cavities in the central region of the disc are present, where (sub)mm observations have revealed the presence of non-axysimmetric overdense lumps of material. Chap. 9 presents the results we published in Ragusa et al. (2017) about this subject. In that paper, we found that non-axisymmetric features might form as consequence of the interaction of the disc with an object with a mass greater than $\gtrsim 50\,\mathrm{M_J}$, i.e. a brown dwarf. Less massive objects are more consistent with ring structures rather than asymmetries with our choice of disc parameters. This constitutes an alternative scenario to the most frequently invoked Rossby wave instability model (RWI), that involves the formation of a vortex induced by a Jovian planet. The massive companion scenario discussed in Ragusa et al. (2017) allows us to obtain analogous results relaxing some assumptions on the disc "viscosity" (that depends on how turbulent the disc is), about which still much debate is ongoing.

Part I

Introduction

Chapter

1

Basic concepts

"Hey oh, let's go!"

Ramones, Blitzkrieg bop

1.1 Binary systems in celestial mechanics: relevant quantities for the two-body problem

The two body problem is the most fundamental problem of celestial mechanics. It consists in the study of the motion of two point-like bodies with masses M_1 and M_2 mutually exerting on each other an attractive force $\boldsymbol{F}$ that follows the universal law of gravitation, i.e.:

$$|\boldsymbol{F}| = \frac{GM_1M_2}{R^2}, \tag{1.1}$$

where $\boldsymbol{R} = \boldsymbol{R}_2 - \boldsymbol{R}_1$ and $R = |\boldsymbol{R}|$ represents the distance between M_1 and M_2 ($\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ represent the vector position of the two masses), and $G = 6.67259 \times 10^{-8}\ \mathrm{g}^{-1}\,\mathrm{cm}^3\,\mathrm{s}^{-2}$ is the gravitation constant. As a consequence, the equations of motion for the two masses read:

$$\boldsymbol{F}_1 = \frac{GM_1M_2}{R^3}\boldsymbol{R} = m\ddot{\boldsymbol{R}}_1, \tag{1.2}$$

$$\boldsymbol{F}_2 = -\boldsymbol{F}_1 = -\frac{GM_1M_2}{R^3}\boldsymbol{R} = m\ddot{\boldsymbol{R}}_2, \tag{1.3}$$

where the double dot notation $\ddot{\boldsymbol{R}}$ represent the second time derivative of the position vector $\boldsymbol{R}$.

It can be shown that since no external forces are acting on the system, and since the gravitational force is conservative, the total energy is a constant of motion and is equal to the sum of the kinetic and potential energy:

$$E_{\rm tot} = \frac{1}{2}M_1 v_1^2 + \frac{1}{2}M_2 v_2^2 - \frac{GM_1M_2}{R}, \tag{1.4}$$

where $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are the velocity vectors with respect to the centre of mass of the system of M_1 and M_2, respectively. Analogously, since no external torques are acting on the system, also the angular momentum is conserved; in particular, it reads:

$$L_{\rm tot} = \boldsymbol{R}_1 \times M_1\boldsymbol{v}_1 + \boldsymbol{R}_1 \times M_2\boldsymbol{v}_2. \tag{1.5}$$

Eq. (1.2) and (1.3) can be manipulated in order to obtain two new equations (see chapter 2 in the textbook Murray & Dermott 1999): one describing the motion of the centre of mass of the system (since no external forces are present it moves with constant velocity), the other describing the relative motion of the two bodies, using $\boldsymbol{R} = \boldsymbol{R}_2 - \boldsymbol{R}_1$ and $\boldsymbol{v} = \boldsymbol{v}_2 - \boldsymbol{v}_1$. It can be shown that the two body problem in the centre of mass frame reduces to a new problem (the "reduced two body problem") of a particle of mass $\mu = M_1M_2/(M_1 + M_2)$, called reduced mass, with velocity $\boldsymbol{v}$, moving in the gravitational potential $V_{\rm red} = -G(M_1 + M_2)/R$ around a fixed "focus" located at $\boldsymbol{R} = 0$. Using this notation $E_{\rm tot}$ reads[1]

$$E_{\rm tot} = \frac{1}{2}\mu v^2 - \frac{G\mu(M_1 + M_2)}{R}, \tag{1.6}$$

while $L_{\rm tot}$ becomes

$$\boldsymbol{L}_{\rm tot} = \boldsymbol{R} \times \mu\boldsymbol{v}. \tag{1.7}$$

We introduce here for future reference also the angular momentum per unit mass in the reduced problem h that satisfies

$$|\boldsymbol{L}_{\rm tot}| = \mu h. \tag{1.8}$$

The orbit of such a particle lies on the plane perpendicular to $\boldsymbol{L}_{\rm tot}$, its shape can be obtained by noticing that a third constant of motion can be identified, the Lenz vector $\boldsymbol{\mathcal{A}}$:

$$\boldsymbol{\mathcal{A}} = \mu\boldsymbol{v} \times \boldsymbol{L}_{\rm tot} - \mu\frac{GM_1M_2}{R}\boldsymbol{R}. \tag{1.9}$$

One can easily verify that $\boldsymbol{L}_{\rm tot} \cdot \boldsymbol{\mathcal{A}} = 0$, so that $\boldsymbol{\mathcal{A}}$ lies in the orbital plane; furthermore, $d\boldsymbol{\mathcal{A}}/dt = 0$, so that $\boldsymbol{\mathcal{A}}$ is in fact constant. Finally, defining

$$\boldsymbol{e} = \frac{\boldsymbol{\mathcal{A}}}{G\mu(M_1M_2)}, \tag{1.10}$$

and taking $\boldsymbol{e} \cdot \boldsymbol{R}$, one gets the following relationship between R and the angle θ between the directions of $\boldsymbol{e}$ and $\boldsymbol{R}$:

$$R = \frac{L_{\rm tot}^2}{1 + |\boldsymbol{e}|\cos\theta}[G\mu(M_1M_2)]^{-1}. \tag{1.11}$$

[1]Note that $\mu(M_1 + M_2) = M_1M_2$ and that in the centre of mass frame $M_1\boldsymbol{v}_1 + M_2\boldsymbol{v}_2 = 0$.

Eq. (1.11) represent a conic equation: the centre of mass is fixed in one of the foci, the type of the conic section depends on whether the value of $|\boldsymbol{e}| \gtreqless 1$.

Further manipulating Eq. (1.11), it can be shown that the type of orbit depends on the value of the total energy $E_{\rm tot}$, in particular:

1. $|\boldsymbol{e}| < 1$ elliptical orbits (closed bound orbits $E_{\rm tot} < 0$).
2. $|\boldsymbol{e}| = 1$ parabolic orbits (marginally bound orbits $E_{\rm tot} = 0$).
3. $|\boldsymbol{e}| > 1$ hyperbolic orbits (open unbound orbits $E_{\rm tot} > 0$).

In this thesis we will deal exclusively with elliptical closed orbits that characterize any kind of binary system (star + star, BH + BH, planet + moon, etc.). We report here few other properties of these orbits that can be obtained by further developing the equations presented throughout this section, but whose derivation goes beyond the scope of the thesis.

- The total energy of an orbit intrinsically sets its semi-major axis a as follows:

$$a = -\frac{GM_1M_2}{2E_{\rm tot}}. \tag{1.12}$$

- The direction of $\boldsymbol{e}$ indicates the direction of the pericentre of the orbit. Since the Lenz vector is a constant of motion in the two body problem, its direction does not change throughout the evolution of the system.

- For bound orbits the quantity $L^2_{\rm tot}[G\mu(M_1M_2)]^{-1}$ in Eq. (1.11) can be written as

$$\frac{L^2_{\rm tot}}{G\mu(M_1M_2)} = a(1-e^2). \tag{1.13}$$

  Substituting Eq. (1.8) in Eq. (1.13), the eccentricity (i.e. $e = |\boldsymbol{e}|$) of the orbit can be computed as

$$e^2 = 1 - \frac{h^2}{G(M_1+M_2)a}. \tag{1.14}$$

  Note that $h = \sqrt{G(M_1+M_2)a}$ is the total angular momentum per unit of reduced mass of circular orbits (see Eq. 1.8), since from its substitution in Eq. (1.14) one gets $e = 0$. From Eq. (1.14), it directly follows that the total angular momentum of the system for a given eccentricity e and semi-major axis a reads

$$L_{\rm tot} = \mu\sqrt{G(M_1+M_2)a(1-e^2)}. \tag{1.15}$$

- Finally, the orbital frequency Ω, i.e. the mean angular velocity along the orbit, depends exclusively on the masses and semi-major axis of the binary, it reads:

$$\Omega_{\rm bin} = \sqrt{\frac{G(M_1+M_2)}{a^3}}. \tag{1.16}$$

We note in conclusion that when $M_2 \ll M_1$ one gets $\mu \to M_2$, so that, for small satellites, the approximation where the first object is fixed in the focus of the ellipse is generally reliable. Under this assumption, the orbital frequency of M_2 can be approximated as

$$\Omega_{\mathrm{K}} = \sqrt{\frac{GM_1}{a^3}}, \tag{1.17}$$

we will refer to Ω_{K} as the Keplerian frequency.

1.2 Accretion as a source of energy

Let assume that a particle of mass m is on a circular orbit around central body of mass $M_\star$. The energy and angular momentum of the system are

$$E = -\frac{1}{2}\frac{GM_\star m}{a}, \tag{1.18}$$

$$\boldsymbol{L} = m\sqrt{GM_\star a}\,\hat{\boldsymbol{n}}, \tag{1.19}$$

where a is the semi-major axis and $\hat{\boldsymbol{n}}$ is the unit vector normal to the orbital plane. As we discussed in Sec. 1.1, these two quantities are conserved along the orbit if no other forces are acting. Suppose that we have some forces acting on the mass m which remove the angular momentum and dissipate the energy accordingly in order to maintain it on a circular orbit. The particle moves inward at a rate

$$\dot{a} = \frac{2\dot{E}a^2}{GMm}, \tag{1.20}$$

where $\dot{E}$ is the power of the force acting on the mass m. In order to bring the mass m to the central object from infinity, the total energy that needs to be dissipated is

$$\Delta E_{\mathrm{diss}} = -\frac{1}{2}\frac{GM_\star m}{R_{\mathrm{in}}} \tag{1.21}$$

where R_{in} is the radius of the innermost orbit where the centrifugal balance is established.

In an accretion disc, the material accretes on to the central object at a rate $\dot{M}$, so that the power of the dissipative forces in order to enable such an accretion rate has to be

$$\Delta E_{\mathrm{diss}} = \frac{1}{2}\frac{GM_\star \dot{M}}{R_{\mathrm{in}}}. \tag{1.22}$$

For protoplanetary discs, R_{in} represents the location of the boundary layer between the disc and the star[2]. For accretion discs surrounding black holes, when the material reaches the innermost stable circular orbit (ISCO) $R_{\mathrm{in}} = R_{\mathrm{ISCO}}$ centrifugal balance cannot be sustained, and the material spirals into the black hole.

[2] In this case, the material will dissipate its remaining kinetic energy in the boundary layer before being accreted (see Sec. 2.2.4).

1.3 What are black holes?

In the context of general relativity, matter and energy bend the space-time, characterizing the trajectories of the bodies in the surroundings. The bending of the space time acts on bodies as a gravitational field in the context of Newtonian gravity, with the difference that also massless bodies, such as photons, have their motion affected by the distribution of matter. Objects producing a gravitational field or, from the general relativistic perspective, a bending of the space time, so strong that even photons remain trapped in closed orbits around them are called black holes.

Even though black holes were predicted for the first time as singularities of the solutions of the Einstein field equation for a spherical distribution of mass in the mid 1910s (Schwarzschild 1916, a translation from German of the original paper can be found in Schwarzschild 1999), they were not referred with this name until the late 1960s[3] when the extremely intense extra-galactic radio-sources from star-like objects (QSOs, **Q**uasi **S**tellar **O**bjects detected in the early 1960s by Schmidt 1963), were proposed to be powered (see Sec. 1.2) by the accretion of material on to extremely massive black holes hosted in the nuclei[4] of high redshift galaxies (Lynden-Bell 1969).

Even though their name refers specifically to the fact that black holes are not expected to be sources of electromagnetic radiation, as will be better discussed in Sec. 1.2, the accretion of material on to black holes is second only to the direct annihilation of matter-antimatter in terms of efficiency to convert mass into energy, making accreting black holes among the most powerful sources of electromagnetic radiation in our Universe.

Nowadays, it is well established that black holes can be divided in two well separated mass regimes: stellar mass black holes (masses $M_\bullet \sim 3$–$10^2\,\mathrm{M}_\odot$) and supermassive black holes ($10^6\,M_\odot \lesssim M_\bullet \lesssim 10^{10}\,M_\odot$).

- The first have been associated with the powerful radiation (bolometric luminosity $L \sim 10^{38}\,\mathrm{erg\,s^{-1}} \approx 10^5\,\mathrm{L}_\odot$, where $\mathrm{L}_\odot = 4 \times 10^{33}\,\mathrm{erg\,s^{-1}}$ is the bolometric luminosity of the Sun) coming from X-ray binaries: interpreted as close binaries composed by a massive star and a compact object (black holes or in some cases also neutron stars), accreting material from the companion star.

- Concerning the second, as previously mentioned, the accretion of material on to a very massive compact object, is the most reliable explanation for the very localized emission of large amount of radiation ($L \sim 10^{45}\,\mathrm{erg\,s^{-1}}$) and the launch of relativistic jets of material (see Fig. 1.1) from the central region of a large number of galaxies (active galactic nuclei, AGN). Further evidence of massive objects at the center of galaxies relies mainly on the direct observation of stars or material in Keplerian motion at sub-parsec

[3]The first usage "black holes" to refer to these mathematical oddities is attributed to John Archibald Wheeler in 1967.

[4]Even though strong electromagnetic emissions from the nuclei of some spiral galaxies had already been observed in the early '40s (Seyfert galaxies, Seyfert 1943).

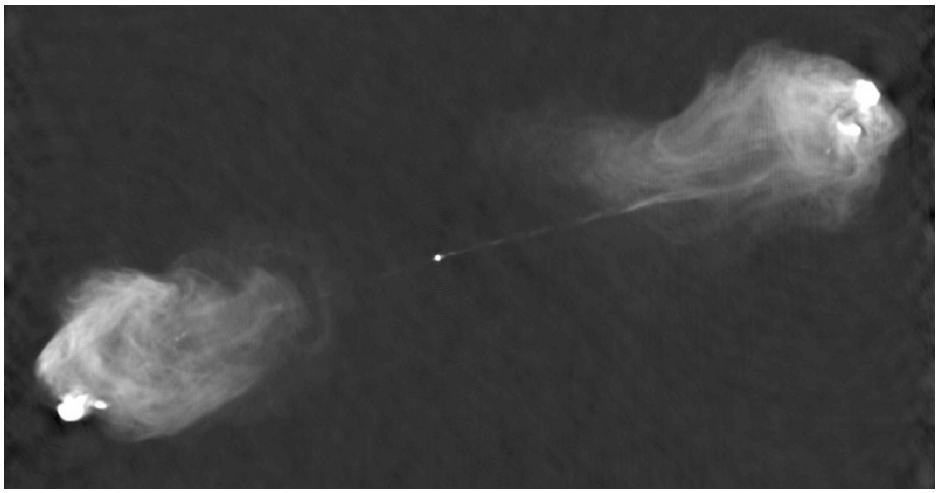


Figure 1.1: The galaxy Cygnus A observed by the Very Large Array (VLA) at radio wavelength $\lambda = 6$ cm. One can clearly distinguish the nucleus of the galaxy (the central dot), where a supermassive black hole of $M_\bullet = 2.5 \pm 0.7 \times 10^9 \mathrm{M}_\odot$ is expected to reside (Graham 2008), from which two jets are emitted toward the plasma radio lobes. Original image from Perley et al. (1984).

scales consistent with very large masses: this is the case for example of the Keplerian motion of material consistent with a central mass of $3.6 \times 10^7\, M_\odot$ in NGC4258 (Miyoshi et al. 1995); or also the tracking of the orbits of a large number of stars in the center of the Milky Way (Gravity Collaboration et al. 2018, for the latest results on the topic), that host a quiescent (SgtA*, $L \sim 10^{34}\,\mathrm{erg\,s^{-1}}$) supermassive black hole having a mass $M_\bullet = 4 \times 10^6 M_\odot$.

In the following sections we will discuss the main parameters characterizing black holes and the main properties of these two categories. However, we note for completeness that the existence of a third intermediate mass regime $10^3 - 10^5 M_\odot$ (Taniguchi et al. 2000), possibly hosted in globular clusters, is currently under debate.

1.3.1 Formalism and relevant quantities

Black holes are fully characterized by two[5] main parameters: their mass $M_\bullet$ and spin $\boldsymbol{S}$ (i.e. the intrinsic angular momentum of the black hole).

Solving the Einstein field equation assuming spherical symmetry, provides the metric of the space-time around non-spinning black-holes (Schwarzschild 1916):

$$ds^2 = -\left(1 - \frac{2GM_\bullet}{c^2 R}\right) dt^2 + \left(1 - \frac{2GM_\bullet}{c^2 R}\right)^{-1} dR^2 + R^2(d\theta^2 + \sin^2\theta d\phi^2), \quad (1.23)$$

where (R, θ, ϕ) are the ordinary spherical coordinates, while c is the speed of light.

[5]The no-hair conjecture (known also as "no hair theorem") for black holes does not prevent black holes to have an electric charge Q, however if that was the case they would be rapidly neutralized attracting opposite charged particles in the surrounding material (Narayan 2005).

The Schwarzschild metric in Eq. 1.23 has two singularities: one for $R = 0$, the other for $R = 2GM/c^2$. The first singularity ($R = 0$) is a singularity of the-space time itself, where the curvature in infinite. The second is a singularity of the metric and can be eliminated providing a change of coordinates (Kruskal 1960), implying that also geodesics crossing this singularity can be studied. This radius defines a spherical surface that characterize the event horizon of a non-rotating black hole:

$$R_{\rm sch} \equiv \frac{2GM}{c^2} \sim 3 \left(\frac{M_\bullet}{M_\odot}\right) \text{ km} \sim 2 \left(\frac{M_\bullet}{10^8 \cdot M_\odot}\right) \text{ au.} \tag{1.24}$$

representing in fact the "size" of the black hole.

Solving Einstein field equation with axial symmetry provides instead the space-time around rotating black holes (Kerr 1963). The event horizon for this type of black holes is located at the Kerr radius(Teukolsky 2015):

$$R_{\rm kerr} \equiv \frac{1}{2}\left(R_{\rm sch} + \sqrt{R_{\rm sch}^2 - \chi_{\rm s}^2 R_{\rm sch}^2}\right). \tag{1.25}$$

where $0 \leq \chi_{\rm s} \leq 1$ is the dimensionless spin of the black hole, which relates to the black hole spin $\boldsymbol{S}$ as

$$\boldsymbol{S} = \chi_{\rm s} \frac{GM_\bullet^2}{c} \boldsymbol{s}. \tag{1.26}$$

Note that the requirement $\chi_{\rm s} \leq 1$ sets a limit for the maximum spin a black hole can have. This enforces the principle of causality at the event horizon of a maximally spinning black hole ($\chi_{\rm s} \leq 1 \rightarrow v_{\rm rot,BH} = |\boldsymbol{S}|/(M_\bullet R_{\rm kerr}) \leq c$).

Finally, we note that the notion of "gravitational radius"

$$R_{\rm g} = \frac{GM_\bullet}{c^2} \tag{1.27}$$

is widely used as a characteristic length scale in the context of black hole physics, regardless of the black hole spin (note however that value of $R_{\rm g}$ is in fact the Kerr radius $R_{\rm kerr} = R_{\rm sch}/2$ of a maximally spinning $\chi_{\rm s} = 1$ black hole).

From the celestial mechanics point of view, as long as the semi-major axis of the orbit is $a \gg R_{\rm g}$, black holes behave as pointlike masses in the context of Newtonian gravity for the material in orbital motion around them. When $a \lesssim 10^2 - 10^3\, R_{\rm g}$ relativistic correction cannot be neglected and processes such as the emission of gravitational waves and frame dragging need to be taken into account in order to properly describe orbital motion of particles around black holes.

An important concept that has to be introduced is that of "innermost stable circular orbit" (ISCO). It represents the ultimate orbit, in terms of separation from the black hole, at which circular orbits are stable. Below that separation no stable orbits are allowed and the material gets "swallowed" by the black hole as soon as it moves from its circular motion. The distance from the black hole at which this orbit is located is generally referred to as $R_{\rm ISCO}$.

It can be shown that if the motion of the material is prograde with respect to the spin of the black hole $R_{\rm g} \lesssim R_{\rm ISCO} \lesssim 6\, R_{\rm g}$; in particular $R_{\rm ISCO} = 6\, R_{\rm g}$ when $\chi_{\rm s} = 0$ and $R_{\rm ISCO} = R_{\rm g}$ when $\chi_{\rm s} = 1$. If the material is on a retrograde orbit with respect to the spin of the black hole $R_{\rm ISCO}$ can grow up to $R_{\rm ISCO} = 9\, R_{\rm g}$.

For practical purposes, it is useful to write $R_{\rm ISCO}$ as

$$R_{\rm ISCO} = k(\chi_{\rm s}) R_{\rm g} \tag{1.28}$$

where $k(\chi_{\rm s})$ is a proportionality factor that depends on the dimensionless spin.

1.3.1.1 Energetics of accreting black holes

Using the dissipation rate for accreting objects we derived in Eq. (1.22), substituting $R_{\rm in} = R_{\rm ISCO}$ we obtain the power released by a black hole accreting at a rate $\dot{M}$

$$\dot{E}_{\rm diss,\bullet} = \frac{1}{2}\frac{GM_\bullet \dot{M}}{k(\chi_{\rm s}) R_{\rm g}} = \frac{1}{2}\frac{\dot{M}c^2}{k(\chi_{\rm s})}. \tag{1.29}$$

It is interesting to note that the process of accretion on to black holes is second only to the direct annihilation matter-antimatter as the most efficient mechanism to convert mass into energy. The maximum efficiency is reached for maximally spinning black holes where $\dot{E} \sim \dot{M}c^2/2$.

In general, the amount of energy dissipated contributes to the heating balance of the disc. However, determining the thermal structure of the disc needs necessarily to account for the optical properties of the material, in order to determine whether the energy is retained in the disc or rapidly radiated away.

An important consequence of this is that since a fraction ϵ of mass is converted in energy during the process of accretion, only a fraction $1-\epsilon$ of the total gas inflow $\dot{M}$ contributes to the mass growth of the black hole, i.e. the mass growth rate of the black hole $\dot{M}_\bullet$ is given by (see also Eq. 1.35)

$$\dot{M}_\bullet = (1-\epsilon)\dot{M}. \tag{1.30}$$

The quantity ϵ is generally referred to as radiation efficiency. Although Eq. (1.29) suggests that $\epsilon = [k(\chi_{\rm s})]^{-1}$, the real value of ϵ is actually only $\epsilon \lesssim [k(\chi_{\rm s})]^{-1}$. Indeed, the full general relativistic treatment of the problem provides some small deviations from this first order estimate, due to the fact that the classical expression of the binding energy of the material (used in the derivation) is not a reliable approximation at such close distances from the black hole (Bardeen 1970, or see the textbook Frank et al. 2002 for a simpler discussion).

1.3.1.2 Limiting the accretion rate: the Eddington luminosity

The large amount of energy produced during the process of accretion of material on to the black holes (see Eq. 1.29) provides a radiation pressure that quenches the infall of the material. This sets a limit on the luminosity of an object, that depends exclusively on its mass, in order to allow the material to be accreted. This limit can be computed balancing the force exerted by a photon on to a nucleus of the infalling gas with the gravitational pull it undergoes.

We compute the luminosity limit for the case of spherical accretion, we will discuss its applicability to the disc case below.

We use the following assumptions: the infall of the material is spherical; the interaction between the radiation and the gas occurs through Thomson scattering;

the material is fully ionized hydrogen. This last assumption is perfectly reasonable since radiation pressure becomes dominant with respect to thermal pressure much above $T > 10^4$ K. However, such an assumption is necessary since the momentum transported by photons needs to be transfered to protons, although photons only interact with the electrons: the overall charge conservation of the plasma of a neutral plasma implies that electrons drag with them also protons.

We balance the force exerted by the radiation on a proton with the gravitational force

$$\frac{L_{\rm Edd}}{c}\frac{\sigma_{\rm T}}{4\pi R^2} = \frac{GM_\bullet}{R^2}m_{\rm p} \tag{1.31}$$

where $L_{\rm Edd}$ is the Eddington luminosity, c is the speed of light, $\sigma_{\rm T}$ is the Thomson scattering cross section and $m_{\rm p} = 1.6726217 \times 10^{-24}\,{\rm g}$ is the proton mass. We recast this last equation in order to obtain

$$L_{\rm Edd} \equiv \frac{4\pi G M_\bullet m_{\rm p} c}{\sigma_{\rm T}} \simeq 1.2 \cdot 10^{38}\left(\frac{M_\bullet}{M_\odot}\right)\ {\rm erg}\cdot{\rm s}^{-1}. \tag{1.32}$$

Assuming that the luminosity produced is given by $L_{\rm Edd} = \epsilon \dot{M}_{\rm Edd} c^2$ we get

$$\dot{M}_{\rm Edd} \equiv \frac{L_{\rm Edd}}{\epsilon c^2} \sim 2.1 \times 10^{-4}\left(\frac{M_\bullet}{M_\odot}\right)\left(\frac{\epsilon}{0.1}\right)^{-1}\ {\rm M}_\odot\,{\rm yr}^{-1}. \tag{1.33}$$

It is important here to remark some important aspects of these results. Firstly, the Eddington limit is on the luminosity: as a consequence the accretion limit is inferred only after the assumption that $L = \epsilon \dot{M} c^2$, i.e. an extremely high efficiency in converting mass into energy, which is obviously not the case for systems not involving black holes. Secondly, the assumption of spherical symmetry is a strong assumption, therefore the real accretion limit for discs might be significantly different from the nominal Eddington accretion rate.

1.3.2 Stellar black holes

Stellar mass black holes are a class of black holes with masses $3{\rm M}_\odot \lesssim M_\bullet \lesssim 10^2{\rm M}_\odot$ and are considered the end point of the life of massive stars. Stars with masses[6] $M_\star \gtrsim 9\,{\rm M}_\odot$ end their life when their cycle of nuclear reactions produces an iron ^{56}Fe core that cannot produce energy by igniting nuclear fusion into heavier elements. The lack of production of energy in the core of the star prevents the maintenance of the hydrostatic equilibrium and the star rapidly collapse. This triggers a complex sequence of physical processes whose outcome is the explosion of the star in a type II supernova event and the formation of a compact object with a radius of a few kilometers: a neutron star or, if originally the star was sufficiently massive, a black hole. Determining what is the initial mass of a star in order to form a black hole depends mainly on two issues: what is the maximum mass for which hydrostatic equilibrium in a neutron star can be achieved, and how much mass the star loses throughout its life.

[6] Stars below this threshold do not produce massive enough cores to collapse, and end their lives as white dwarfs (Heger et al. 2003).

The maximum mass for which pressure balances the self-gravity in a neutron star can be computed using the general relativistic equation for hydrostatic equilibrium (Tolman 1939; Oppenheimer & Volkoff 1939) coupled with a prescription for the equation of state (i.e. a prescription for how pressure depends on the local properties within the neutron star), that however it has not been unambigously determined yet. Currently, different equations of state predicts this limit to be $2\,\mathrm{M}_\odot \lesssim M_{\mathrm{NS,max}} \lesssim 2.5\,\mathrm{M}_\odot$ (Özel & Freire 2016).

Concerning the second issue, the mass loss throughout the life of a star depends on its initial mass $M_\star$ and its metallicity Z.

Finally, it is very important to note that, for stars with initial masses $135\,\mathrm{M}_\odot \lesssim M_\star \lesssim 260\,\mathrm{M}_\odot$ and sufficiently low metallicities Z, no remnants form at all; these stars undergo the so called "pair-instability" (Heger et al. 2003; Woosley et al. 2007). If the star is sufficiently massive, the γ photons produced in the star core are sufficiently energetic to produce couples e^+–e^- when interacting with the nuclei. This lowers the thermal pressure (since part of the energy of the photon is temporarily transformed into electron and positron mass), the core shrinks while its temperature and nuclear density grow, further increasing the efficiency in the production of couples e^+–e^-. This process triggers an uncontrolled increase of the core temperature associated to the ignition of a runaway thermonuclear process that leads to the explosion of the star without leaving remnants. As consequence, a hole in the black hole mass distribution is expected for masses ranging $52\,\mathrm{M}_\odot \lesssim M_\bullet \lesssim 133\,\mathrm{M}_\odot$ (Woosley 2017).

1.3.3 Supermassive black holes

Supermassive black holes are a class of black holes with masses ranging $10^6\mathrm{M}_\odot \lesssim M_\bullet \lesssim 10^{10}\mathrm{M}_\odot$. These compact massive objects have now been long recognized to power the intense electromagnetic emission ($L_{\mathrm{AGN}} \gtrsim 10^{45}\,\mathrm{erg\,s^{-1}} \approx 10^{12}\,\mathrm{L}_\odot$) from the central regions of many galaxies (Active Galactic Nuclei, AGN) (Lynden-Bell 1969).

The origin of supermassive black holes is still a highly debated topic. The main problem with their formation mechanism is that it has to account for the observation of extremely bright sources, that are consistent to be powered by black holes $M_\bullet \gtrsim 10^9\,\mathrm{M}_\odot$, beyond redshift $z = 7$ (e.g., Mortlock et al. 2011; Bañados et al. 2018), i.e. when the Universe was younger than 1 Gyr old (Volonteri 2012). A model able to create such massive compact objects within timescales as short as 1 Gyr, is challenged by the large amount of energy that must be irradiated in order to allow the accretion of material on to the black hole (see Sec. 1.2). Indeed, there exists a limit value for the accretion rate ($\dot{M}_{\mathrm{Edd}} \propto M_\bullet$ Eddington limit for the accretion, see Sec. 1.3.1.2) that represents the maximum accretion rate before the outward moving photons produce a radiation pressure that quenches the infall of the material on to the black hole. Given the linear dependence of the Eddington limit on the black hole mass, the evolution of the black hole mass with time $M_\bullet(t)$ can be written as

$$M_\bullet(t) = M_0 \exp\left(\frac{1-\epsilon}{\epsilon} f_{\mathrm{Edd}} \frac{t}{0.45\,\mathrm{Gyr}}\right) \tag{1.34}$$

where M_0 is the initial mass of the black hole seed, ϵ is the radiation efficiency, i.e. the fraction mass that is radiated away as photons energy (see the end of Sec. 1.2 for further details), $f_{\rm Edd} = \dot{M}_\bullet/\dot{M}_{\rm Edd}$ is the Eddington fraction ($f_{\rm Edd} = 1$ indicates accretion at the Eddington limit, see Sec. 1.3.1.2). This equation can be recast as

$$t = 0.45\,{\rm Gyr} \times \frac{\epsilon}{1-\epsilon} f_{\rm Edd}^{-1} \ln\left(\frac{M_\bullet}{M_0}\right) \tag{1.35}$$

From Eq. (1.35) we see that three parameters rule the mass growth timescale of black holes: the accretion rate $f_{\rm Edd}$, the seed mass M_0 and the radiation efficiency ϵ. In order to reach $M_\bullet \approx 10^9\,{\rm M}_\odot$ in 1 Gyr it is necessary to require high $f_{\rm Edd} \sim 1$, high seed masses and low radiation efficiencies ϵ.

On the one hand, it can be shown that the lower is the spin of the black hole, the lower the radiation efficiency is: in particular, Schwarzschild black holes have $\epsilon = 0.06$ growing to $\epsilon \approx 0.4$ for maximally spinning Kerr black holes. This implies that if the spin of the black hole is low enough, its mass growth rate might be sufficient to form massive black holes in less than 1 Gyr, if the accretion rate is always close to Eddington.

However, even starting with a non-spinning black hole, if the black hole is fed by an accretion disc, when the material is accreted, the spin of the black hole grows by the same amount as the angular momentum of the material at the ISCO. This implies that, as the mass of the black hole increases, also its radiation efficiency does, slowing down its mass growth rate. A possible solution to prevent this process is that the accretion of material on to the black hole occurs through a series of randomly oriented accretion events, this would prevent the black hole spin and radiation efficiency to grow (chaotic accretion, King & Pringle 2006).

In this respect we also note that some works have shown that, although the luminosity cannot be super-Eddington (the validity of this limit has been put into question by Sakurai et al. 2016,), very high accretion rates ($\dot{M} \gtrsim 1000 L_{\rm Edd}/c^2$) can be achieved whenever photons are dragged/advected into the black hole by the inflow of material (Begelman 1979). This mechanism is called "hyperaccretion" and in fact it translates in very low values of ϵ allowing very high accretion rates but maintaining relatively low luminosities (some examples of recent works on the subject are Park & Ricotti 2011, 2012; Inayoshi et al. 2016; Sakurai et al. 2016; Begelman & Volonteri 2017; Jiang et al. 2017).

On the other hand, the production of massive seeds M_0 relies on two main mechanisms (see Sesana 2012; Volonteri 2012 for a review):

- Primordial massive stars: Population III (PopIII) stars (Madau & Rees 2001), are stars formed in the the primordial, metal poor, environment. In such a metal poor environment, cooling and fragmentation are very inefficient, resulting in the formation of massive stars $M_\star \gtrsim 260\,{\rm M}_\odot$.

  PopIII stars were expected to leave at the end of their evolution black holes as massive as $M_\bullet \sim 100\,{\rm M}_\odot$ that can act as seeds M_0 for the further accretion. However, more recent work found that during the process of Pop III stars fragmentation is more efficient than before believed implying that the masses of PopIII remnant black holes are actually much lower (Greif et al. 2011), making in fact this mechanism not effective in order to form massive seeds.

Runaway mergers of PopIII stars have been shown to possibly solve the problem, forming black hole seeds with masses $M_0 \approx 10^3 - 10^5\,\mathrm{M}_\odot$ (Omukai et al. 2008; Devecchi & Volonteri 2009; Davies et al. 2011; Katz et al. 2015).

- Global instabilities: under this category goes a large number of different models able to bring a very large amount of gas toward the centre of primordial dark matter halos (Oh & Haiman 2002; Bromm & Loeb 2003; Begelman et al. 2006; Lodato & Natarajan 2006; Begelman 2010; Latif et al. 2013).

  These models essentially differ among them for the physical processes involved to provide the infall of material: pre-galactic discs that become globally gravitationally unstable might collapse without fragmenting, producing supermassive stars with masses up to $10^5\,\mathrm{M}_\odot$ provided that the accretion rate is $\dot{M} \gtrsim 0.1\,\mathrm{M}_\odot\mathrm{yr}^{-1}$ (Hosokawa et al. 2013; Haemmerlé et al. 2018) that become general relativistic unstable (Chandrasekhar 1964) when their masses reach $M \approx 10^5\,\mathrm{M}_\odot$. In other cases, at the end of the hydrogen burning, super-massive star cores can collapse forming black holes that accrete material from their envelopes at a super Eddington rate (quasistar, stars with a black hole core sustained against collapse by the radiation pressure produced by the accretion on its black hole core, Begelman et al. 2006; Begelman 2010). Analogously, a large amount of material might be driven toward the centre of pre-galactic discs when they become gravitationally unstable: gravitational instability promotes the redistribution of the gas angular momentum allowing fast formation of dense black hole seeds $M_0 \approx 10^5\,\mathrm{M}_\odot$ (Lodato & Natarajan 2006), provided that the disc does not fragment halting the gas inward motion. The key assumption for these mechanisms to be effective is that the cooling mechanisms of the gas are inefficient; this requirement is satisfied in the metal poor young Universe where a strong UV background (Lyman-Wener background), produced by young stars, prevents the formation of molecular hydrogen H_2 (Haiman et al. 1997). Galaxy mergers might promote high inflow rates of gas toward the central regions of merger remnants, possibly triggering the formation of super-massive stars also in galaxies with solar metallicities (Mayer & Bonoli 2019).

The amount of material "feeding" the black hole is balanced by the radiation produced during the process of accretion, that acts as a "feedback" mechanism (see Wise 2018 for a recent review). This implies that the accretion rate on to the first black holes might be significantly reduced with respect to $f_{\mathrm{Edd}} \approx 1$ (e.g., (Pezzulli et al. 2017)). Investigating the feeding-feedback cycle can be done using numerical simulations that include all the physical processes that are relevant to this aim, in order to understand how the mass growth of black holes self-regulates (e.g., Sijacki et al. 2007; Cielo et al. 2018).

Finally, the growth of the supermassive black hole mass has been suggested to stop when its mass reaches $M_\bullet \approx 5 \times 10^{10}$–$2.7 \times 10^{11}\,\mathrm{M}_\odot$ since its event horizon reaches the fragmentation radius of the accretion disc that is feeding it (King 2016; Inayoshi & Haiman 2016). Indeed, beyond this radius, the disc surrounding the black hole is expected to be gravitationally unstable, to clump and form stars, halting the accretion on the black hole.

1.4 What are protoplanetary discs?

Giant molecular clouds, as the name suggests, are large agglomerates (size $\sim 50-100\,\mathrm{pc}$ and mass $10^4-10^6\,\mathrm{M}_\odot$) of gas (mainly hydrogen in molecular form H_2) and dust with a typical dust mass fraction of $M_{\mathrm{dust}} \sim 10^{-2} M_{\mathrm{gas}}$ (dust-to-gas ratio in the interstellar medium, ISM). The hydrostatic equilibrium of a molecular cloud depends on the balance between its pressure and its self-gravity. For a fixed temperature T, assuming a constant density ρ throughout the cloud, perturbations of the gas density are unstable with respect to gravitational collapse when the length-scale of the perturbation exceeds the Jeans length

$$\lambda_{\mathrm{Jeans}} = \frac{2\pi c_{\mathrm{s}}}{\sqrt{4\pi G\rho}}, \tag{1.36}$$

where $c_{\mathrm{s}} = \sqrt{k_{\mathrm{B}}T/\mu m_{\mathrm{p}}}$ is the sound speed in the cloud (assuming an isothermal equation of state), $k_{\mathrm{B}} = 1.3806488 \times 10^{-16}\,\mathrm{erg\,K^{-1}}$ is the Boltzmann constant, m_{p} is the proton mass and μ is the mean molecular weight. Typical clouds have a minimum temperature of $T \sim 10\,\mathrm{K}$, mean molecular density $\mu = 2.3$, providing sound-speeds of the order of $c_{\mathrm{s}} \sim 2 \times 10^4\,\mathrm{cm\,s^{-1}}$, and densities $\rho \sim 10^{-17} - 10^{-19}\,\mathrm{g\,cm^{-3}}$ (Lodato 2008). This implies a typical Jeans length of the order of $\lambda_{\mathrm{Jeans}} \sim 0.1-1\mathrm{pc}$, implying a mass of the unstable core $M_{\mathrm{core}} \approx \rho\lambda^3_{\mathrm{Jeans}} \sim 0.1-1\,\mathrm{M}_\odot$. This qualitatively tells us that a giant cloud will fragment in individual small collapsing cores, each having a mass of $\sim 1\mathrm{M}_\odot$. The fragmentation of a giant molecular clouds, and the collapse of cloud cores constitute the very beginning of the process of star formation.

It is important to remember that despite our previous simplifying assumption of constant density throughout the cloud, giant molecular clouds are far from being homogeneous objects (McKee & Ostriker 2007). In particular, the density and the velocity field span a very broad range of values, showing a characteristic Kolmogorov (1941)-like turbulent energy cascade at different scales. For this reason, cloud cores always have a certain amount of intrinsic rotation $\Omega_{\mathrm{core}} \sim 10^{-14} - 10^{-13}\,\mathrm{s^{-1}}$. This provides an initial angular momentum per unit mass $h \approx \Omega_{\mathrm{core}}\lambda^2_{Jeans} \sim 10^{21} - 10^{22}\,\mathrm{cm^2\,s^{-1}}$. The conservation of the total angular momentum of the cloud core causes the growth of the rotational velocity, until the collapse is stopped as the centrifugal balance is achieved. The geometry of this new stable configuration is a disc-like distribution of gas and dust in Keplerian motion around a central protostar, a circumstellar disc (see Fig. 1.3). Equating the angular momentum per unit mass expected for Keplerian orbits (see the end of Sec. 1.1) to the initial total angular momentum of the cloud core allows us to determine the typical length-scale R_{circ} (called "circularization radius") at which the collapse stops:

$$R_{\mathrm{circ}} = \frac{h^2}{GM_\star} \sim 10^2 - 10^4\,\mathrm{au}, \tag{1.37}$$

where $M_\star$ is the mass of the central mass; the reference values refer to $M_\star = 1\,\mathrm{M}_\odot$ as a mass of the central star. This result is consistent with the observation of gas accretion discs extending up to $10^3\,\mathrm{au}$ surrounding typical stars with masses $\approx 1\mathrm{M}_\odot$ in region of star formation.

Turbulence, sustained by some form of instability (Balbus & Hawley 1998), provides an effective mechanism to transport outward the angular momentum and dissipate the kinetic energy sustaining the centrifugal balance in the disc (see Sec. 2.3), enabling the accretion of material on to the central star.

In general, in these early stage of star formation these systems are often referred in the literature as "young stellar objects" (YSO); However, within the aforementioned discs, the formation of planets is expected to take place, and for this reason circumstellar discs are often referred to as "protoplanetary discs". As a consequence, in order to emphasise the process of planet formation, another common denomination for YSOs presenting discs is "protoplanetary systems".

1.4.1 Classes of young stellar objects

Young stellar objects are divided in four main classes (André et al. 2000) based on their spectral energy distribution (SED), that provides important information about the temperature and mass displacement. Each class is believed represent a different stage of the evolution of young stellar object (from early to late). The parameter used for the classification is the so called IR slope:

$$\alpha_{\rm IR} = \left(\frac{d\lambda F_\lambda}{d\lambda} \right)_{\lambda_{\rm NIR/MIR}} \tag{1.38}$$

that represents the slope of the SED at wavelengths $\lambda_{\rm NIR/MIR} = 2.4 - 25\,\mu$m. The classification is as follows (see also Fig. 1.2, and the textbook Armitage 2010 for a deeper discussion):

- **Class 0**: This class represents very young, collapsing cloud cores enveloping a very faint protostar. These sources do not show any emission at μm wavelengths, but are instead characterized by an optically thick black body emission from the cold outer layers of the envelope at (sub)mm wavelengths. This class can be considered the time $t = 0$ of evolution.

- **Class I**: These sources are characterized by a positive IR slope $\alpha_{\rm IR} \gtrsim 0$. The radiation coming from the central star heats up the dust component of the collapsing cloud core that re-emits it at lower wavelengths. The star is still partially embedded in the collapsing cloud. This phase lasts for the first $t \approx 10^5\,$yr of life of the system.

- **Class II**: When most of the gas + dust envelope has been accreted, the radiation of the protostar becomes visible at optical wavelengths, while at longer wavelengths the electromagnetic emission is largely dominated by the IR and (sub)mm emission from the disc. This provides the so called IR excess that characterizes protostars surrounded by protoplanetary discs. Objects in Class II show a typical $-4/3 \lesssim \alpha_{\rm IR} \lesssim 0$. Class II objects often show an ultraviolet (UV) excess, with respect to the pure emission of the star photosphere, associated with the accretion of material on to the star. T-Tauri stars and protoplanetary systems, are typical objects to be part of this class. This evolutionary stage lasts up to 10 Myr.

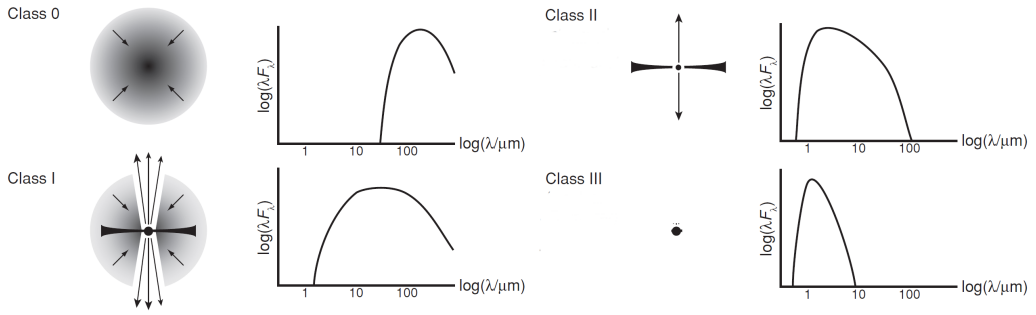


Figure 1.2: Sketch of the geometry of the system associated with the corresponding SED in the four different YSO classes. Adapted from the original picture in Armitage (2010).

- **Class III**: As the time passes the disc is progressively dispersed or accreted. A very small amount of material is left surrounding the protostar. Some systems show in their SEDs the signatures of small dust rings at large distance from the star ("debris discs"). In this class the IR slope steepens to values $\alpha_{\rm IR} \lesssim -4/3$ typical of pre-main sequence stars and the object lies above the main sequence when located in the Hertzprung-Russell color-magnitude diagram.

This SED based classification is not absolute, the geometry or the inclination of the system might in principle alter the optical depth between the observer and the star causing very inclined Class II objects to be mistaken for a Class I or vice versa (McKee & Ostriker 2007).

We finally report that a number of systems have been observed to show the typical Class II excess at long wavelenghts but characterized by a flux reduction at near-mid-infrared wavelengths (NIR/MIR, $\lambda \approx 1-20\,\mu$m) Strom et al. (1989); Skrutskie et al. (1990). These systems have been interpreted to be discs hosting a hole in the distribution of μm sized dust grains and have been suggested to be a population of sources in transition from Class II to Class III (transition discs, see Chap. 8 for a more thorough discussion). More generally, Class II objects have recently attracted a lot of attention from the scientific community thank to the advent of new observational facilities and instruments, such as the Atacama Large Millimetre Array (ALMA) and the Spectro-Polarimetric High-contrast Exoplanet REsearch (SPHERE) instrument mounted on the Very Large Telescope (VLT). The high resolution images these facilities have provided, have shown a large variety of structures (see Fig. 1.3) that have been mainly associated with planet formation, in fact setting the beginning of a new era for astronomy in which the process of planet formation can be directly probed through observations.

1.4.2 Planet formation in protoplanetary discs

In this section we briefly review the current theory of planet formation. Two main mechanisms are generally invoked: gravitational instability and core accretion (see the textbook Armitage 2010 or Armitage 2018 for a recent review). In the first, the dust component of the disc plays just a marginal role while the dynamics of

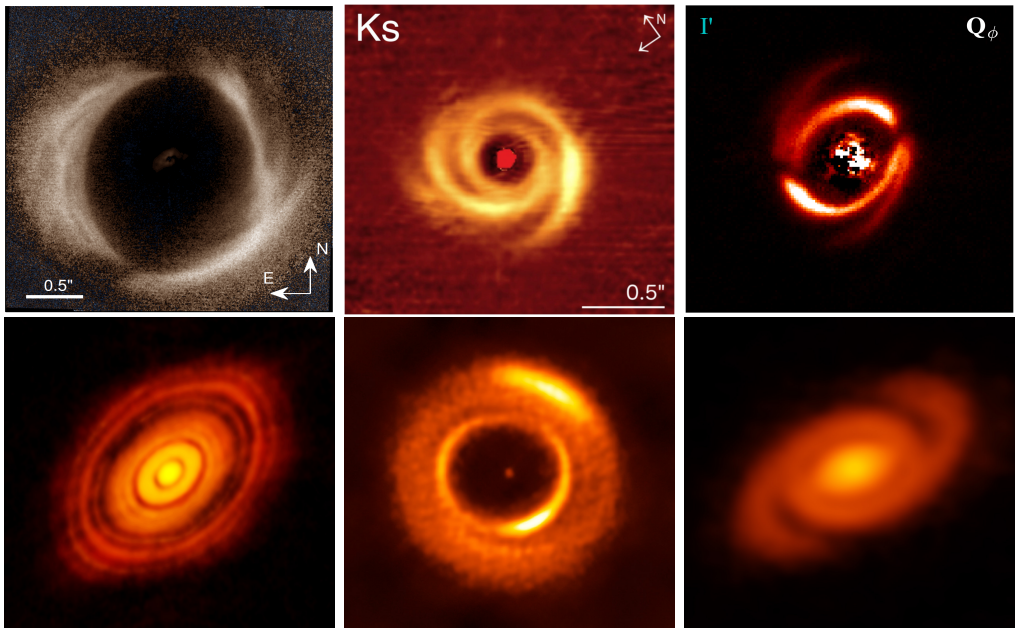


Figure 1.3: Gap, cavities, spirals, shadows, and other non axisymmetric features observed in protoplanetary discs at different wavelengths. Top row, SPHERE NIR resolved images of some protoplanetary discs, obtained using polarimetric differential imaging technique (PDI). From left to right: HD 142527 (SPHERE, Very Broad Band $\sim$ 600-900nm, original image from: Avenhaus et al. 2017); HD 135344B (SPHERE, band Ks, original image from: Garufi et al. 2013); HD 100453 (SPHERE, band I', original image from Benisty et al. 2017). Bottom row, ALMA resolved images in the (sub-)mm waveband of some protoplanetary discs. From left to right: HL Tauri (ALMA, band 6, original image from: ALMA Partnership et al. 2015); MWC 758 (ALMA, band 7, original image from: Dong et al. 2018); Elias 2-27 (ALMA, Band 6, original image: B. Saxton (NRAO/AUI/NSF), Pérez et al. 2016).

gas drives the entire process; in contrast, the second strongly relies on the complex dynamics of dust and its aerodynamical coupling with the gas.

1.4.2.1 Gravitational instability

The gravitational instability scenario for planet formation prescribes that planet formation is a consequence of disc fragmentation when the gaseous disc is massive enough to become gravitationally unstable. The local instability criterion for a disc requires that the Q-parameter is lower than unity[7]:

$$Q = \frac{c_{\rm s}\kappa}{\pi G\Sigma} \lesssim 1, \tag{1.39}$$

where $c_{\rm s}$ is the local sound speed, G is the constant of gravitation, Σ is the local disc surface density and κ is the epicyclic frequency of the gas, that is defined as follows

$$\kappa^2(R) = \frac{1}{R^3}\frac{\partial}{\partial R}[R^2\Omega(R)]^2, \tag{1.40}$$

where $\Omega(R)$ is the rotation frequency profile of the disc[8]. It can be shown that Eq. (1.39) implies that a disc is locally unstable if

$$\frac{M_{\rm d}}{M_\star} \gtrsim \frac{H}{R}, \tag{1.41}$$

where $M_{\rm d}$ is the disc mass, $M_\star$ is the star mass and H/R is the disc thickness-radius aspect-ratio.

Besides the stability criterion, an appropriate description of the gas cooling mechanism needs to be provided in order to study the progression of the gravitational instability: if the gas is not allowed to cool, when it contracts under the effect of self-gravity, the material increases its temperature improving the stability of the system. Furthermore, the onset of the gravitational instability excites spiral density waves that steepens into shocks providing a source of heating and, again, causing the system to stabilize. In this context, the parameter

$$\beta = t_{\rm cool}\Omega(R) \tag{1.42}$$

is introduced with the aim to provide a simple toy model for the cooling, where $t_{\rm cool}$ sets the timescale for one e-fold reduction of the internal energy of the gas (Gammie 2001).

Fragmentation is expected to occur when the criterion in Eq. (1.39) is largely satisfied ($Q << 1$). However, if the cooling timescale is sufficiently long ($\beta \gg 20$), the system self-regulates at values $Q \approx 1$, promoting the formation of a large scale spiral density structure but preventing the disc fragmentation.

A large number of numerical simulations (see Kratter & Lodato 2016 for a recent review) have shown that values $\beta \sim 10-20$ provide the threshold for fragmentation. An estimate of the typical mass of the fragments can be given by noticing

[7]This result can be obtained from Eq. (B.4.1), which describes the dispersion relation of density waves in discs.

[8]Note that for Keplerian discs $\kappa = \Omega_{\rm K}$ where $\Omega_{\rm K}$ is the Keplerian frequency in Eq. (1.17).

that the most unstable wavelength when $Q = 1$ is $\lambda = 2\pi H$, where H is the disc thickness. Then the fragments mass $M_{\rm frag}$ can be readily obtained by

$$M_{\rm frag} = \pi\Sigma\lambda^2 = 4\pi\left(\frac{H}{R}\right)^2 M_{\rm d}. \tag{1.43}$$

which, requiring that the disc is gravitationally unstable, leads to

$$M_{\rm frag} = 4\pi\left(\frac{H}{R}\right)^3 M_\star \approx 10\,{\rm M_J}. \tag{1.44}$$

Even though Eq. (1.44) represents just a crude estimate of the order of magnitude of fragment masses, we can conclude that the gravitational instability can explain mainly the formation of massive planets. However, it should be noted that the core of objects with masses $\gtrsim 13\,{\rm M_J}$ is sufficiently hot to ignite the nuclear fusion of deuterium making them in fact brown dwarfs. Given the relatively high mass of the fragments, it is still debated whether the onset of gravitational instability is more likely to be responsible for the formation of massive planets (Durisen et al. 2007) or to origin low mass star companions of the central proto-star (Zhu et al. 2012a).

1.4.2.2 Core accretion

The mechanism of core accretion models the formation of planets as a four step process:

1. **Dust Growth**: µm-sized dust grains undergo multiple collisions within the disc. Van der Waals or electrostatic forces allow the grains to stick together forming larger grains during this process dust grains grow from $\sim$ µm to mm in size. This process, often referred to as "collisional coagulation", can be indirectly probed by observations: the emission at long wavelengths ((sub)mm-radio) from protoplanetary systems is optically thin; since the opacity is sensitive to the size distribution of dust grains, by measuring the slope of the SED at those wavelengths it is possible to collect information about the dust population.

2. **Formation of "planetesimals"**: The mechanisms leading from millimetric grains to the formation of planetesimals (bodies of ~ 1 km in radius) are still poorly understood (see Youdin 2010, for a review). The progress of dust growth through collisional coagulation from mm to km sized solids is far from being straightforward. The ability of two particles to "stick" together depends on the electrostatic binding energy and on the kinetic energy of the collision. It has been shown that electrostatic forces cannot stick effectively particles with sizes $\approx$ mm for collisions at velocities $\gtrsim 1\,{\rm m\,s^{-1}}$ (Blum & Wurm 2008); considering the gravitational forces, bodies with size ~ 1 m, the binding energy is not sufficient to stick particles even for collision velocities $\approx 1\,{\rm mm\,s^{-1}}$, in contrast with typical collision velocities of $10\,{\rm m\,s^{-1}}$ (Youdin 2010). Furthermore, the migration of meter sized bodies has been shown to occur on timescales of the order of $\sim$ 100 yr (Weidenschilling 1977),

so even in the case of effective collisional coagulation of dust grains the formation of planetesimals must be extremely fast, in contrast with the very long timescales required for the coagulation of dust grains (this issue is often referred to as the "meter size problem").

A possible solution to this issue is the onset of drag instabilities, such as the Goldreich & Ward (1973) or the streaming instability (Youdin & Goodman 2005; Jacquet et al. 2011). These promotes the formation of dust clumps in turbulent flows, possibly leading to the rapid gravitational collapse of clumps into planetesimals.

3. **Formation of rocky cores**: This phase consists in the further growth of the size of planetesimals up to reach the size of terrestrial planets $\sim 1\,\mathrm{M}_\oplus$, or of giant planet cores (where $\mathrm{M}_\oplus = 5.9722 \times 10^{27}\,\mathrm{g}$ is the Earth mass). The growth in this phase is driven by the gravitational attraction and collision of dust grains on to already existing planetesimals. During this phase, smaller bodies that orbit at sub-Keplerian velocities, due to their aerodynamical coupling with the gas, collide with the surface of the planetesimal. The gravity of the planetesimal focuses the trajectories of dust grains, increasing its effective cross section (gravitational focusing). Drag effects on to marginally coupled dust grains, called "pebbles" (size $\sim$ 1 cm), further increase the accretion of material on to the planetesimal since, as they gain velocity toward the planetesimal, the drag force dissipate their kinetic energy increasing the chances they become gravitationally bound to it ("pebble accretion", see Ormel 2017 for a review).

4. **Core accretion**: At the onset of this phase, the body has become a planet; the aerodynamical coupling with the environment becomes negligible and the disc-planet interaction mainly occurs through gravitational forces. In this phase, if the core becomes massive enough from phase 3., its gravitational field becomes sufficiently strong to retain an atmosphere, i.e. a core envelope in hydrostatic equilibrium. As the core mass grows through further collision with smaller planetesimals and a critical mass is reached of $\approx 10\,\mathrm{M}_\oplus$, the planet enters a phase of runaway accretion which is modulated by a circumplanetary disc. The accretion on to the planet is halted only when the planet carves a sufficiently deep gap in the gas (see Chap. 4.1) or as a consequence of the disc dispersal.

Chapter

2

Accretion disc physics

Boards of Canada, Dawn Chorus

Free falling material (mostly hydrogen gas with a small fraction of dust and other molecules or chemical elements) on to a point-like source of gravity, like a star or a black hole, always has an initial amount of angular momentum. During the free fall, the conservation of angular momentum increases the spinning motion of the material. When the centrifugal balance is reached the free fall is halted. When this condition is met, the geometry of the system consists of a central body and a disc of material orbiting around it.

Once such a configuration is reached, no further infall motion of the material would be in principle allowed unless some physical processes are able to dissipate the gravitational energy and transport outward the angular momentum. The turbulent motion of the gas has been suggested to drive the transport of angular momentum toward the outer radii enabling the further infall of the material. For this reason, these systems are called "accretion discs". The disc structure strongly depends on the type of central object around which the disc orbits due to the different nature of the physical processes taking place in it.

In this Chapter we discuss the disc structure under a number of simplifying assumptions, constituting the so called "classic accretion disc theory". A number of textbooks fully cover the topic (Frank et al. 2002; Pringle & King 2007; Armitage

2010; Clarke & Carswell 2014). Here we will limit to present the main results and concepts that will be useful throughout this manuscript.

2.1 Basic equations

We describe the gas as a continuous fluid. As a consequence, governed by fluid dynamics equations[1]

$$\frac{\partial \rho}{\partial t} + (\boldsymbol{v} \cdot \nabla)\rho = -\rho \nabla \cdot \boldsymbol{v}, \tag{2.1}$$

$$\rho \left[\frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v} \cdot \nabla)\boldsymbol{v} \right] = -\nabla p - \rho \nabla \Phi_{\rm tot} + \nabla \cdot \mathbf{S}, \tag{2.2}$$

$$\rho \left[\frac{\partial u}{\partial t} + (\boldsymbol{v} \cdot \nabla)u \right] = p \nabla \cdot \boldsymbol{v} + \rho \mathcal{L}, \tag{2.3}$$

where ρ is the fluid density, $\boldsymbol{v}$ is the fluid velocity, u is the internal energy per unit mass, p is the fluid pressure and $\mathcal{L}$ is the term accounting for heating or cooling of the fluid per unit mass. The term $\nabla \Phi_{\rm tot}$ accounts for the acceleration due to gravity, where $\Phi_{\rm tot}$ is the gravitational potential.

In this and in the following sections $\Phi_{\rm tot} = \Phi_\star + \Phi_{\rm sg}$ has two contributions: point-like objects, such as stars, planets or black holes, $\Phi_\star$, and the self-gravity of the fluid $\Phi_{\rm sg}$; the potential $\Phi_{\rm sg}$ is computed solving the Poisson's equation:

$$\nabla^2 \Phi_{\rm sg} = 4\pi G \rho, \tag{2.4}$$

where G is the universal constant of gravitation. Finally, $\mathbf{S}$ is the viscous stress tensor, it reads:

$$\mathsf{S}^{ij} = \eta \left[\nabla^j v^i + \nabla^i v^j - \frac{2}{3}\delta^{ij} \nabla^k v^k \right] + \zeta \delta^{ij} \nabla^k v^k; \tag{2.5}$$

where η and ζ are the shear and bulk viscosity parameters, respectively. The component S^{ij} represent the flux of the j-th component of the momentum per unit mass in the i-th direction, so that $\nabla^j \mathsf{S}^{ij} = a^i_{\rm visc}$ is the acceleration in the i direction due to the viscous forces.

It is useful for some purposes to define an additional quantity. Assuming a barotropic equation of state (i.e., $p(\rho)$), we define the enthalpy h

$$h \equiv \int \frac{dp}{\rho} \rightarrow \nabla h = \frac{1}{\rho} \nabla p, \tag{2.6}$$

so that Eq. (2.2) can be written as

$$\frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v} \cdot \nabla)\boldsymbol{v} = -\nabla (h + \Phi_{\rm tot}) + \frac{1}{\rho} \nabla \cdot \mathbf{S}. \tag{2.7}$$

[1] Note that for some purposes it is useful to state Eq. (2.2) in tensor notation as follows

$$\frac{\partial v^i}{\partial t} + (v^j \nabla^j) v^i = -\frac{1}{\rho} \nabla^j (p\delta^{ij} - \mathsf{S}^{ij} + \Phi_{\rm tot} \delta^{ij}),$$

where δ^{ij} is the Kroenecker delta.

It is useful to note that a small variation of the enthalpy dh, relates with the density as follows:

$$dh = c_{\rm s}^2 \frac{d\rho}{\rho}, \tag{2.8}$$

where $c_{\rm s}$ is the sound speed of the fluid defined as

$$\frac{dp}{d\rho} = c_{\rm s}^2. \tag{2.9}$$

2.2 Classic accretion disc theory: the thin disc approximation

The equations presented in the previous section can be simplified for our purposes under a few assumptions:

- Axial symmetry: therefore, all the azimuthal gradients vanish in Eq. (2.1), (2.2) and (2.3).

- The disc is thin: i.e., the disc height-to-radius aspect ratio $H/R \ll 1$.

- The disc self-gravity is negligible: the gravitational potential terms satisfy $\Phi_\star \gg \Phi_{\rm sg}$ so that $\Phi_{\rm tot}$ in Eq. 2.2 becomes $\Phi_{\rm tot} \approx \Phi_\star \equiv \Phi$. It can be shown that self-gravity can be neglected if the following relationship (equivalent to require $Q \gg 1$ in Eq. 1.39) between the disc thickness and disc mass is satisfied
$$\frac{M_{\rm disc}}{M_\star} \ll \frac{H}{R}. \tag{2.10}$$

- The temperature profile of the disc is known: we can thus completely neglect Eq. (2.3) prescribing a radial power law profile for the sound speed ($c_{\rm s} \propto R^{-q}$, i.e., $T \propto R^{-2q}$), and setting an equation of state for pressure p.

- Viscosity provides the angular momentum transport required to drive the accretion of material on to the central star: since only shear-forces are required for this purpose (see textbooks Frank et al. 2002; Clarke & Carswell 2014) and since molecular viscosity is negligible, bulk viscosity can be neglected, i.e. $\zeta = 0$ in Eq. (2.5). Under these assumptions, the only non-vanishing term of the stress tensor is $\mathsf{S}^{R\varphi}$. It is important to note that in real discs the angular momentum transport appears to be driven by turbulence (see Sec. 2.3).

Under these assumptions we can recast the Eq. (2.1) and (2.2) in order to study the structure and evolution of the disc. Given the geometry of the system, the best coordinate system to treat the problem is the cylindrical one (radius, azimuthal angle and distance from the midplane, $[R, \varphi, z]$), setting the central body in $(R = 0, z = 0)$.

2.2.1 Dynamics in 2D discs

We can assume that a thin disc behaves dynamically like a 2D system. This assumption is reasonable as long as we can neglect the vertical motion of the gas and the variation of v_R and v_φ along the vertical direction. As a consequence, adding the assumption that the viscous transport of angular momentum is a slow process, the gas velocities satisfy the following scale relations

$$v_z \ll v_R \ll c_\mathrm{s} \ll v_\varphi. \tag{2.11}$$

By vertically integrating the continuity equation (2.1) and Navier-Stokes equation (2.2), and applying the previous considerations, the dynamics of a thin axisymmetric disc is fully described by the following set of simplified equations:

$$\frac{\partial \Sigma}{\partial t} = -\frac{1}{R}\frac{\partial}{\partial R}(R\Sigma v_R), \tag{2.12}$$

$$\Sigma\left[\frac{\partial v_R}{\partial t} - \frac{v_\varphi^2}{R}\right] = -\frac{\partial P}{\partial R} - \Sigma\frac{\partial \Phi}{\partial R}, \tag{2.13}$$

$$\Sigma\left[\frac{\partial v_\varphi}{\partial t} + v_R\frac{\partial v_\varphi}{\partial R} + \frac{v_R v_\varphi}{R}\right] = \frac{1}{R^2}\frac{\partial}{\partial R}R^2\mathsf{T}^{R\varphi}, \tag{2.14}$$

where Σ, P and T^{ij} are the results of the operation of vertical integration of ρ, p and σ^{ij}, respectively[2]. The quantity $\nu = \Sigma^{-1}\int_z \eta\, dz$ is referred to as viscous parameter and it quantifies the intensity of the viscous forces in the disc (see Sec. 2.3.1 for a deeper discussion).

2.2.2 Vertical and radial equilibrium

We discuss the vertical structure by imposing the vertical hydrostatic equilibrium in Eq. (2.2), i.e. vertical pressure and gravity forces balance as follows

$$\frac{1}{\rho}\frac{\partial p}{\partial z} = -\frac{\partial \Phi}{\partial z}. \tag{2.15}$$

Under the assumption of $z \ll R$, and using the definition of sound speed c_s in Eq. (2.9) the solution to the previous equation is

$$\rho(z) = \rho_0 \exp\left(-\frac{1}{2}\frac{z^2}{H^2}\right), \tag{2.16}$$

where $H = c_\mathrm{s}/\Omega_\mathrm{K}$ and ρ_0 is the density value at the midplane. This allows us to define the disc aspect ratio as follows

$$\frac{H}{R} = \frac{c_\mathrm{s}}{v_{\varphi,\mathrm{K}}}, \tag{2.17}$$

[2]We note that in this context it is straightforward to re-define the sound speed as $c_\mathrm{S} = \partial P/\partial\Sigma$ instead of using the definition of the sound speed c_s in Eq. (2.9). Throughout this thesis we will keep referring to c_S as c_s.

where $v_{\varphi,\mathrm{K}} = \Omega_\mathrm{K} R$ is the Keplerian azimuthal velocity. The reader should note that the disc aspect ratio is the inverse of the Mach number of the gas. Therefore, requiring the disc to be be thin $H/R \ll 1$ implies also that the gas azimuthal motion is highly supersonic.

Concerning the radial equilibrium, using Eq. (2.13) we can now study the disc centrifugal balance. We require stationarity by setting the time derivative $\partial_t v_R = 0$. We obtain the following condition

$$\frac{v_\varphi^2}{R} = \frac{1}{\Sigma}\frac{\partial P}{\partial R} + \frac{\partial \Phi}{\partial R}. \tag{2.18}$$

Using a globally isothermal equation of state, the definition of Keplerian velocity $v_{\varphi,\mathrm{K}}$ and Eq. (2.17), Eq. (2.18) can be restated as follows

$$v_\varphi^2 = v_{\varphi,\mathrm{K}}^2 \left[1 + \frac{\partial \ln \Sigma}{\partial \ln R}\left(\frac{H}{R}\right)^2\right]. \tag{2.19}$$

The result in Eq. (2.19) allows us to conclude that in accretion discs pressure gradients alter the centrifugal Keplerian balance that would be expected for pressureless material. In particular, under our assumptions, a negative pressure gradient (i.e. $\partial_R \Sigma < 0$, that is generally the case for unperturbed discs) provides some additional support against the gravitational attraction toward the central object, implying that the material have stable circular orbits with sub-Keplerian velocity. In contrast, a positive pressure gradient requires a super-Keplerian velocity for the material for establishing the centrifugal balance.

2.2.3 Angular momentum conservation and density evolution

In this section we discuss the conservation of angular momentum across the disc. Combining equation (2.14) with (2.12), after a little algebra one gets

$$\frac{\partial}{\partial t}(\Sigma R v_\varphi) + \frac{1}{R}\frac{\partial}{\partial R}(R v_R \Sigma R v_\varphi) = \frac{1}{R}\frac{\partial}{\partial R}(\Sigma \nu R^3 \Omega'). \tag{2.20}$$

This equation expresses the conservation of angular momentum: its time derivative can be easily recognized in the first term in the l.h.s; the second term in the l.h.s. constitutes the radial advection of angular momentum; finally, the term in the r.h.s. is the flux of angular momentum produced by the disc viscosity.

2.2.4 Steady state solution

We now study the structure of a stationary disc. To do so, we neglect time derivatives in Eq. (2.12) and (2.20). We then integrate them over the azimuthal domain (given axial symmetry this is equivalent to multiplying the equation by 2π)

$$\frac{1}{R}\frac{\partial}{\partial R}(2\pi v_R \Sigma R) = 0, \tag{2.21}$$

$$\frac{1}{R}\frac{\partial}{\partial R}(R 2\pi v_R \Sigma R^2 \Omega) = \frac{1}{R}\frac{\partial}{\partial R}(2\pi \nu \Sigma R^3 \Omega'). \tag{2.22}$$

Given the steady nature of the system, these equations naturally prescribe the introduction of two constants: the accretion rate $\dot{M}$ and the angular momentum flux $\dot{J}$. The equations can be closed using the so called "no torque" boundary condition $\dot{J} = 2\pi v_R \Sigma R_{\rm in}^3 \Omega(R_{\rm in})$, where $R = R_{\rm in}$ is the location where Ω' and, as a consequence, the viscous torque vanish. Under these assumptions, it can be easily shown that the profile of $\nu\Sigma$ required in order to have steady accretion reads

$$\nu\Sigma = \frac{\dot{M}}{3\pi}\left(1 - \sqrt{\frac{R_{\rm in}}{R}}\right)^{-1}. \tag{2.23}$$

We note that Eq. (2.23) provides a very useful scale relation that relate the disc properties with the accretion rate. In particular, at large radii ($R \gg R_{\rm in}$), the accretion rate $\dot{M}$ reads

$$\dot{M} = 3\pi\nu\Sigma. \tag{2.24}$$

2.2.5 Density evolution

Substituting Eq. (2.12) into Eq. (2.20), and using $v_\varphi = \Omega R$ we find that v_R reads

$$v_R = \frac{1}{R\Sigma(R^2\Omega)'}\frac{\partial}{\partial R}\left(\Sigma\nu R^3\Omega'\right). \tag{2.25}$$

Substituting back Eq. (2.25) into Eq. (2.12), one gets[3]

$$\frac{\partial\Sigma}{\partial t} = -\frac{1}{R}\frac{\partial}{\partial R}\left[\frac{1}{(R^2\Omega)'}\frac{\partial}{\partial R}(\Sigma\nu R^3\Omega')\right], \tag{2.27}$$

where $(R^2\Omega)' = d(R^2\Omega)/dR$. Two things should be noted in Eq. (2.27). Firstly, the density evolution of a viscous disc satisfies a diffusion equation. Secondly, a necessary condition for the evolution of the disc is that $\nu \neq 0$. This is kind of obvious, since in absence of a mechanism able to transport the angular momentum such as viscosity, each fluid element would remain unperturbed on its initial orbit.

A dimensional analysis of Eq. (2.27), allows us to introduce the typical timescale for the disc viscous evolution

$$t_\nu = \frac{R^2}{\nu}, \tag{2.28}$$

that is generally referred to as "viscous time".

For future reference, we introduce here the Reynolds number Re, that quantifies the intensity of viscous forces (acting on a timescale t_ν) against inertial ones (acting on a timescale $t_{\rm dyn} = \Omega^{-1}$)

$$\mathrm{Re} = \frac{t_\nu}{t_{\rm dyn}} = \frac{\Omega R^2}{\nu}. \tag{2.29}$$

The Reynolds number provides information about the level of turbulence in the fluid: in particular, fluids with high Reynolds numbers $\mathrm{Re} \gtrsim 10^5$ are expected to be turbulent.

[3] It is usual to see Eq. (2.27) recasted under the assumption that the disc is Keplerian, i.e. $\Omega = \Omega_{\rm K}$ (see Eq. 1.17)

$$\frac{\partial\Sigma}{\partial t} = \frac{3}{R}\frac{\partial}{\partial R}\left[R^{1/2}\frac{\partial}{\partial R}(\Sigma\nu R^{1/2})\right]. \tag{2.26}$$

2.3 Turbulent angular momentum transport

In this section we discuss the physical mechanisms responsible for the disc evolution. So far we used a viscous prescription to model the angular momentum transfer, which is necessary for the accretion of material to occur on to the central object. However, accretion discs are not viscous at all, and the concept of viscosity here has nothing in common with that characterizing viscous fluids we are familiar with on Earth.

The gas in accretion discs has a very low density. Computing the Reynolds number (Eq. 2.29) using the viscous parameter $\nu = \nu_{\rm mol}$ of molecular viscosity[4] of the gas we obtain, for typical disc parameters, that it easily exceed $\mathrm{Re} \gtrsim 10^{14}$ (Frank et al. 2002).

Such a large value implies that disc viscous evolution driven by molecular viscosity alone would take place on a timescale much longer than the age of the Universe. Furthermore, from laboratory experiments we also know that fluids with large Reynolds numbers are prone to become turbulent, i.e. characterized by strong fluctuations in the velocity field.

Balbus & Hawley (1998) showed how the turbulent motion of the gas might be able to promote the evolution of accretion discs. We summarize here the main results of their model. We use the following velocity perturbations,

$$v_\varphi = v_{\varphi,0} + u_\varphi, \quad v_R = v_{R,0} + u_R, \tag{2.30}$$

where we assume that the unperturbed velocity field in the disc has $v_{R,0} = 0$ and $v_{\varphi,0} = R\Omega$. Since we have shown that molecular viscosity has a negligible contribution to the angular momentum transport, we assume that the disc is completely inviscid ($\mathsf{T}^{R\varphi} = 0$) and plug the perturbed velocity in Eq. (2.20). It can be shown that the perturbed equation for angular momentum conservation reads

$$\frac{\partial}{\partial t}(\Sigma R v_\varphi) + \frac{1}{R}\frac{\partial}{\partial R}[R u_R \Sigma R v_{\varphi,0}] = -\frac{1}{R}\frac{\partial}{\partial R}[\Sigma R^2 \langle u_\varphi u_R \rangle_\varphi], \tag{2.31}$$

where $\langle\cdot\rangle_\varphi$ represents an azimuthal average. The previous equation clearly shows that even without explicitly including any form of viscous transport of angular momentum, if a positive correlation between u_φ and u_R is present (as a consequence of turbulence), the term

$$\mathsf{W}^{R\varphi}_{\rm Re} = -\Sigma \langle u_\varphi u_R \rangle_\varphi, \tag{2.32}$$

called "Reynolds Stress", acts as the viscous stress tensor $T^{R\varphi}$.

However, having an high Reynolds number is only a necessary condition for the onset of turbulence. A mechanism able to constantly inject energy in the fluid in order to sustain the turbulent motion is required. If this was not the case, turbulence would progressively decay as a consequence of dissipative effects.

Instabilities, such as the magneto rotational instability (MRI) or gravitational instability (GI), the Rayleigh-Taylor (RTI) instability induced in magnetically

[4] $\nu_{\rm mol} = \lambda_{\rm mfp}\bar{v}$, i.e. the product between the average velocity $\bar{v}$ of particles and their mean free path to interact $\lambda_{\rm mfp}$.

arrested discs (MAD, Narayan et al. 2003; Marshall et al. 2018) and other mechanisms[5] able to produce a non vanishing $\langle u_\varphi u_R \rangle_\varphi$ are possible candidates to be the physical mechanisms responsible for the angular momentum transport and evolution of astrophysical discs. In particular, in a magnetized fully ionized disc the magnetic field $\boldsymbol{B}$ leads to the "Maxwell tensor"

$$\mathsf{W}_M^{R\phi} = \Sigma \langle u_\phi^A u_R^A \rangle_{z,\phi}, \quad \boldsymbol{u}^A = \frac{\boldsymbol{B}}{\sqrt{4\pi\rho}}, \tag{2.33}$$

where u^A is called Alfven velocity. Analogously in presence of a perturbation of the gravitational potential $\boldsymbol{g}$, the stress tensor associated becomes

$$\mathsf{W}_g^{R\phi} = -\Sigma \langle u_\phi^g u_R^g \rangle_{z,\phi}, \quad \boldsymbol{u}^g = \frac{\boldsymbol{g}}{\sqrt{4\pi G\rho}}, \tag{2.34}$$

2.3.1 Turbulence as a form of viscosity: the α prescription

The discussion in the previous section necessarily triggers the following question: is the usage of a viscous prescription ($\mathsf{W}^{R\varphi} \to \mathsf{T}^{R\varphi}$) for the angular momentum transport a reliable approximation of the real physical mechanisms behind it?

Since the viscous prescription provides a local energy dissipation, the energy dissipation of the real physical mechanism must occur locally as well. This has been shown to be true for MRI but not for GI, where the formation of spiral structures produce a non-local dissipation mechanism (Balbus & Papaloizou 1999).

However, also for GI non-local effects are sometimes negligible (Cossins et al. 2009): spiral density waves produced in GI become supersonic when $|\Omega - \Omega_\mathrm{p}|R \approx c_\mathrm{s}$. As soon as the waves become supersonic, they shock and dissipate the energy they are carrying. This occurs close to the excitation region if the disc sound speed is small. With simple calculations, it can be shown that the fraction of energy deposited in the disc due to non-local effects is $E_\mathrm{non-loc} \approx (M_\mathrm{d}/M_\star)E_\mathrm{loc}$, in discs that are marginally stable ($Q \approx 1$, see Eq. 1.39).

We conclude that in most cases disc turbulence can be treated as a local mechanism, and for this reason using a viscous stress tensor is a reliable approximation, whose magnitude is dictated by the only free parameter of $\mathsf{T}^{R\varphi}$, i.e. ν.

In order to provide a natural scaling with the disc properties of ν, Shakura & Sunyaev (1973) suggested the following parametrization

$$\nu = \alpha c_\mathrm{s} H, \tag{2.35}$$

where $\alpha < 1$ is required. Thìs parametrization of ν is mostly known as the "Shakura & Sunyaev (1973) prescription". It straightforwardly follows from the modeling of the "virtual" viscosity associated with turbulent flows[6]; in this situation ν depends on the size of the largest eddies (that cannot exceed H) and on the velocity of the fluid in them (that cannot exceed c_s, since supersonic turbulence would be rapidly dissipated by shocks). Thus, $c_\mathrm{s}H$ represents the highest level of

[5] Among them, also the presence of a secondary object orbiting the primary produces velocity field perturbations able to transport angular momentum.

[6] The turbulent motion accelerates the transport processes that would occur by mere diffusion if the flow was laminar, requiring the addition of a "virtual" term.

viscosity that can be sustained in the disc providing a scale relation for ν. Then $\alpha < 1$ is introduced as a free parameter to properly tune its magnitude.

The value of α varies significantly depending on the source. In fully ionized discs α has been estimated to range between $\alpha \sim 0.1$–0.4 (King et al. 2007). In contrast, partially ionized discs, such as those in protoplanetary systems, are more consistent with $\alpha \approx 0.001$–0.01 (Andrews & Williams 2007). These estimates assume a viscous Shakura & Sunyaev (1973) prescription, and infer α from measuring the properties of outbursts and disc sizes comparing them with evolutionary models with reliable disc parameters.

However, although MRI induced turbulence has been considered for more than two decades the main physical mechanism responsible for the transport of angular momentum through the disc, recent studies in the context of protoplanetary discs have revealed that non ideal MHD effects, coupled with the low level of ionization, suppresses the instability in regions 1 – 10 au.

Furthermore, the advent of the Atacama Large Millimetre Array (ALMA) have allowed to directly put upper limits on the level of turbulence in protolanetary discs, using the properties of molecular lines. Recently, using this technique, the systems HD 163296 and TW Hya showed turbulent velocities lower than those expected for MRI induced turbulence (Flaherty et al. 2015, 2018), putting an upper limit of $\alpha \lesssim 0.007$. It must be reported though that direct evidence of non-thermal motion associated with turbulence have instead been detected at radii $R < 1$ au and $R > 40$ au (see Najita & Bergin 2018 and references therein).

Magnetothermal winds have been suggested as a possible alternative mechanism able to remove angular momentum also in laminar discs (see Turner et al. 2014 for a review). This last mechanism would cause the discs not to spread to large radii, as would instead occur in the turbulent transport scenario. However, the disc sizes inferred from observations still do not allow to discern whether one, the other, or a combination of the two mechanisms are in action (Najita & Bergin 2018).

On the other hand, recently Lodato et al. (2017) showed that the correlation between mass accretion rate and disc mass in the Lupus star forming region (Manara et al. 2016) can be reproduced with a simple viscous disc evolution model.

2.4 Active discs: gas temperature profile

As we discussed in Sec. 1.2, the material, in order to reach the central object, dissipates its gravitational energy producing heat. In this section we derive the temperature profile for a steady state disc where the heating mechanism is mainly provided by the dissipative processes that allow the accretion (see the textbook Frank et al. 2002 for a deeper discussion). These discs are said to be “active”, since their heating comes directly from the physical processes taking place inside them.

It can be shown that, assuming that the disc is Keplerian, the rate of dissipation

per unit surface[7] $\mathcal{D}$ is given by

$$\mathcal{D}(R) = \frac{9}{4}\nu\Sigma\Omega^2. \tag{2.36}$$

It is easy to note that integrating Eq. (2.36) across the disc domain, and assuming that the disc is in steady state (i.e. $\nu\Sigma$ is set by Eq. 2.23) one gets

$$L_{\rm disc} = \int_{R_{\rm in}}^{R_{\rm out}} \mathcal{D}(\mathcal{R}) 2\pi R dR = \int_{R_{\rm in}}^{R_{\rm out}} \frac{3}{2}\frac{GM}{R^2}\dot{M}\left(1-\sqrt{\frac{R_{\rm in}}{R}}\right) dR = \frac{1}{2}\frac{GM}{R_{\rm in}}\dot{M}. \tag{2.37}$$

That is perfectly consistent with the considerations about the accretion luminosity discussed in Sec. 1.2

We equate (2.36) to the bolometric emission of a black body, we get the following result:

$$2\sigma_{\rm SB}T_{\rm s}^4 = \frac{3}{4\pi}\frac{GM}{R^3}\dot{M}\left(1-\sqrt{\frac{R_{\rm in}}{R}}\right), \tag{2.38}$$

where $T_{\rm s}$ is the surface temperature of the disc, $\sigma_{\rm SB}$ is the Stefan-Boltzmann constant, and the factor 2 on the l.h.s. comes from the fact that the disc emits radiation from two surfaces (both disc sides). Rearranging Eq. (2.38) one gets

$$T_{\rm s} = \left[\frac{3}{8\pi\sigma_{SB}}\frac{GM}{R^3}\dot{M}\left(1-\sqrt{\frac{R_{\rm in}}{R}}\right)\right]^{1/4}, \tag{2.39}$$

i.e. for $R \gg R_{\rm in}$ the radial dependence of $T_{\rm s}$ reads

$$T_{\rm s} \simeq T_0\left(\frac{R}{R_0}\right)^{-3/4}. \tag{2.40}$$

If the disc is optically thick it is reasonable to assume that the production of energy takes place in the midplane, where the disc is denser, and then it is transported to the surface. It can be shown that the temperature in the midplane relates with the surface one as

$$T_{\rm c}^4 \simeq \frac{3}{4}\tau T_{\rm s}^4, \tag{2.41}$$

where τ is the mean optical depth (i.e. averaged on frequencies) computed using the Rosseland mean opacity.

2.5 Passive discs: dust temperature profile

In some discs the accretion mechanisms produce an amount of energy that is too small to represent the heating source. Such discs are said to be "passive", since

[7] We note that this result can be obtained both from the relation $\dot{E} = T\Omega$, where T is the torque exerted by viscous forces and Ω the gas orbital frequency, both writing the equations ruling the evolution of kinetic energy $1/2\rho v^2$ and noticing that the sink term $\dot{E}_{k,\rm diss} = \sigma^{ik}\nabla_k v_i$ appears.

their heating is mainly provided by external processes like the presence of an external source of electromagnetic radiation (e.g. a star) or cosmic rays. In general, properly estimating the temperature profile of a passive disc is a very complicated task and has to rely on numerical radiative transfer simulations. Further details about the discussion in this section can be found in Kenyon & Hartmann (1987) and Chiang & Goldreich (1997) (we follow here the approach given in the textbook by Armitage 2010).

Protoplanetary discs are typically passive discs. The main source of external heating is the central star: the disc is generally flared, i.e. H/R grows with R, so that the disc surface is directly exposed to the light irradiated by the star.

To first order, we assume that the disc is in LTE. To fulfill this requirement, we assume that the radiation is absorbed by the dust and re-emitted at a different wavelength with the same rate. We also assume that the optical depth is the same for both the absorption and emission, and consider disc regions that are far enough from the central star ($R \gg R_\star$), in order to neglect the effects of the finite size of the star. Under these assuptions, we balance the energy absorbed and re-emitted by a disc element of area ΔA

$$2\left(\frac{L_\star}{4\pi R^2}\right)\alpha\Delta A = 2\sigma_{\rm SB}T_{\rm s}^4\Delta A, \tag{2.42}$$

where the l.h.s. is the disc absorption rate of the star radiation, while the r.h.s. is the luminosity of the radiation re-irradiated by the disc; $T_{\rm s}$ is the surface temperature of the disc, α is the flaring angle of the disc: the light of the star intercepts the disc with an angle $\theta = \pi/2 - \alpha$ with respect to its normal direction, implying that the effective absorbing surface ΔA is reduced by a factor $\cos(\theta) \rightarrow \sin(\alpha) \approx \alpha$, valid if α is small. The factor two on the l.h.s. and r.h.s. of Eq. (2.42) is due to the fact that the disc absorbs and emits radiation on both sides. Assuming that $L_\star = 4\pi R_\star^2\sigma_{\rm SB}T_\star^4$, Eq. 2.42 can be recast as

$$T_{\rm s}^4 = \alpha^{1/4}\left(\frac{R_\star}{R}\right)^{1/2}T_\star. \tag{2.43}$$

However, determining the temperature profile for passive discs is more complicated than this simple analysis. In the derivation we just provided, the problem was strongly simplified by the assumption that the disc is optically thick. In this case the absorption and emission properties of the disc are exactly the same.

In passive protoplanetary discs, for a fixed size of the dust grains, the dust absorbs more efficiently the radiation with frequencies around the peak of the star SED than those at the lower peak frequency of the dust thermal re-emission. This implies that the stellar radiation penetrates a thin surface layer of the disc ($\tau_{\nu_\star} \approx 1$). However, the same layer is optically thinner at the peak frequencies at which the disc thermally re-emit the radiation ($\tau_{\nu,\rm dst} \ll 1$). Therefore, the stellar radiation heats up only the surface layer of the disc, but the emission from the same hot layer is much less efficient. Furthermore, this implies also that the inner layers (i.e. those that are not directly heated up by the star) contribute to the radiation that emerge from the disc surface, further complicating the problem.

We can improve our estimate of the temperature profile, implementing the different emission efficiency between absorption and emission for dust grains. In this

case, by comparing the absorption opacity with the emission one, it can be shown that the temperature profile becomes

$$T_{\rm s} = \left(\frac{R_\star}{2R}\right)^{2/5} T_\star, \tag{2.44}$$

where the factor 2 at the denominator of the r.h.s. comes from the fact that dust grains absorb the radiation over an effective surface $\pi a_{\rm dst}^2$ and emit from a surface $4\pi a_{\rm dst}^2$. The reader should keep in mind that this approximation neglects the emission from the inner layers. An analytical estimate considering also that contribution can be found in Chiang & Goldreich (1997) but it is beyond the scope this thesis.

2.6 Multicolor black body spectrum

In the regime of optically thick emission, we can derive the electromagnetic spectrum of any source for any temperature profile of the disc assuming that the total flux is composed by the summation of weighted black bodies at a certain temperature. To do so, we integrate the black body brightness function across the disc domain

$$F_\nu = \int_\Omega B_\nu(T) d\Omega, \tag{2.45}$$

where Ω is the solid angle domain subtended by the source.

In most sources, it is reliable to assume that the disc temperature profile is a radial power law

$$T_{\rm s} \simeq T_0 \left(\frac{R}{R_0}\right)^{-q}, \tag{2.46}$$

where T_0 is the temperature at the reference radius R_0. The flux F_ν from such a source thus gets the following scaling with the frequency:

$$F_\nu \equiv \int I_\nu \cos\theta \frac{R dR\, d\varphi}{D^2} = 2\pi \cos\theta \frac{2h\nu^3}{c^2 D^2} \int_{R_{\rm in}}^{R_{\rm out}} \frac{1}{e^{h\nu/k_{\rm B}T_{\rm s}} - 1} R dR \propto \nu^{3-2/q}, \tag{2.47}$$

where we used that the solid angle subtended by an annulus of material is $d\Omega = 2\pi \cos\theta R dR/D^2$, where D is the distance of the source and θ the disc inclination ($\theta = 0$ being a face on disc) and Ω is the solid angle. We note here that the spectral energy distribution νF_ν (SED) is usually used instead of the mere flux F_ν, with the aim to have a direct measure of the relative amount of energy carried at each wavelength. It follows straightforwardly that $\nu F_\nu \propto \nu^{4-2/q}$.

2.6.1 Thermal emission from optically thin discs

Suppose for simplicity that the disc is an isothermal slab of material at temperature T. The properties of its surface emission at a given frequency ν will be characterized by the following surface brightness function $I_\nu^{\rm obs}$

$$I_\nu^{\rm obs} = (1 - e^{-\tau_\nu}) B_\nu(T), \tag{2.48}$$

where B_ν is the thermal black body emission function; $\tau_\nu = \kappa_\nu \Sigma$ is the "optical depth", where κ_ν and Σ are the opacity and surface density, respectively.

From Eq. (2.48) one can clearly see that when the disc is optically thick ($\tau_\nu \gg 1$) the emission spectrum can be safely assumed to be that of a black body $I_\nu^{\rm obs} = B_\nu$.

In contrast, when the disc has $\tau_\nu \ll 1$, the emission is said to be optically thin and the surface brightness becomes[8] $I_\nu^{\rm obs} \approx \kappa_\nu \Sigma B_\nu(T)$. This has the important consequence that the opacity κ_ν and Σ shape the intensity of the electromagnetic output at different frequencies. We also note that the requirement $\tau_\nu \ll 1$ causes the brightness from the disc surface to be always smaller than the black body case (optically thick case).

The footprint of the opacity κ_ν and surface density Σ in the emission makes the optically thin emission a very powerful observational probe of the disc properties. For example, making a reasonable assumption about the opacity function, it is possible to infer the dust mass of the source from the intensity of the observed flux. Analogously, from the dependence on ν of the observed flux it is possible to infer κ_ν.

A practical example of the application of this last property can be found in protoplanetary systems, where the dust emission at radio wavelenghts, i.e. in the Rayleigh-Jeans regime of the thermal dust emission ($B_\nu \propto \nu^2$), carries information about the dependence on ν of the opacity ($\kappa_\nu \propto \nu^\beta$). In particular, observing a slope $I_\nu^{\rm obs} \propto \nu^\xi$ provides the following relation with β (Testi et al. 2014)

$$\beta = \xi - 2. \tag{2.49}$$

This is extremely useful, since it can be shown that β is strongly dependent on the size distribution of dust grains, providing information about their growth and population.

2.7 Dust dynamics

As previously mentioned, the material present in the interstellar medium is composed by gas (mainly hydrogen, and other elements or molecules in a small fraction) and dust (mainly composed by silicates, i.e. sand), typically representing the 1% of the total mass of the material. Dust particles can be found in a very large variety of sizes $a_{\rm dst}$ and shapes (see Testi et al. 2014 for a review about dust properties).

Dust particles can be modeled as a pressureless and viscousless fluid. Their dynamics is different from that of the gas, implying that we need to provide an additional set of equations in order to account for the dynamics of a second fluid species.

Gas and dust interact through the aerodynamical drag the gas exerts on dust grains and vice versa, so that their evolution is not independent. The drag force per unit volume reads

$$f_{\rm D} \equiv \rho_{\rm d} \frac{\Delta \boldsymbol{v}}{t_{\rm s}}, \tag{2.50}$$

[8] Some frequencies might be optically thinner, while others optically thicker (e.g. this is the typical situation of absorption/emission lines).

where $\Delta \boldsymbol{v} = \boldsymbol{v}_\mathrm{d} - \boldsymbol{v}_\mathrm{g}$ is the velocity difference between dust grains and the gas, the subscripts g and d refer to gas and dust quantities, respectively; the variable t_s in Eq. (2.50) is the stopping time. This variable quantifies the intensity of aerodynamical coupling between the gas and the dust; it represents the time it takes for a particle moving with a velocity v_d in a still fluid ($v_\mathrm{g} = 0$) to stop under the action of the drag force f_D: if t_s is small the dust-gas coupling is strong.

The equations ruling the dynamical evolution of the dust density ρ_d and velocity $\boldsymbol{v}_\mathrm{d}$ are

$$\frac{\partial \rho_\mathrm{d}}{\partial t} + (\boldsymbol{v}_\mathrm{d} \cdot \nabla)\rho_\mathrm{d} = -\rho_\mathrm{d} \nabla \boldsymbol{v}_\mathrm{d}, \tag{2.51}$$

$$\rho_\mathrm{d} \left[\frac{\partial v_\mathrm{d}}{\partial t} + (\boldsymbol{v}_\mathrm{d} \cdot \nabla)\boldsymbol{v}_\mathrm{d} \right] = f_\mathrm{D} - \nabla \Phi_\mathrm{tot}, \tag{2.52}$$

Since the aerodynamical drag depends on the relative motion of the gas with respect to dust grains, one needs to add the term f_D changed in sign in the gas momentum conservation equation (2.2). Then Eq. (2.1), (2.2), (2.51) and (2.52) have to be solved simultaneously as a system of equations.

Given the strong disproportion between the mass of the dust and that of the gas, it is usually assumed that, to first order, the dynamics of the dust is altered by the gas but not vice versa[9]. As a consequence, a question spontaneously arises: if the dust is not expected to affect much the dynamics of the gas, which represents the most abundant disc component, why are we interested in studying its dynamics?

For accretion discs surrounding black holes, in fact, the region that produces the electromagnetic emission is so hot that dust grains are not allowed to remain solid. As a consequence, the dynamics and electromagnetic emission of the system can be fully described by one single gaseous component.

However, in protoplanetary systems, although the regions closer to the star are hot enough to cause the dust sublimation ($R_\mathrm{subl} \lesssim 0.2$ au, e.g. Eisner et al. 2005), at larger radii, the dust thermal emission of (sub)mm grains at radio wavelengths and the forward scattering on μm grains of IR photons from the central star are the main targets of the latest observational campaigns of close star forming regions. The appropriate modeling of the dust dynamics is thus of paramount importance for properly understanding the physical processes that are taking place at this stage of the evolution of the system.

Below, we rapidly review the main results in the context of dust dynamics which are relevant for this thesis.

2.7.1 Stopping time

Let us now briefly discuss the stopping time t_s appearing in Eq. (2.50). Two main regimes of drag can be identified depending on the the properties of dust particles and of the environment in which they move: the Epstein and Stokes regimes.

[9] However, the reader should bear in mind that properly accounting for the backreaction of the dust on the gas naturally leads to a large variety of phenomena, such as: the streaming instability (Youdin & Goodman 2005; Jacquet et al. 2011), the destruction of vortices (Fu et al. 2014b) and in general alters the dynamics of the gas with respect to its pure viscous evolution (Kanagawa et al. 2017; Dipierro et al. 2018b)

The Epstein regime (Epstein 1924) is effective when the mean free path $\bar{\lambda}_{\rm mfp}$ of gas molecules is larger than ~ 0.5 times the size of the particle ($\bar{\lambda}_{\rm mfp} > 4/9 a_{\rm dst}$). In this regime the dust particles move in a very sparse medium with respect to their size. The collisions with gas molecules or atoms occur from all the directions with a net transfer of kinetic energy that depends on the temperature of the gas.

In contrast, the Stokes regime (Stepinski & Valageas 1996) is effective when $\bar{\lambda}_{\rm mfp} < 4/9 a_{\rm dst}$. In this case, gas particles behave as a fluid and the transfer of kinetic energy to a dust particle occurs trough viscous friction rather than through random individual encounters as in the Epstein regime. Within this second regime, three additional sub-regimes can be found: Stokes I, Stokes II and Stokes III depending on the Reynolds number of the flow around the grains $\mathrm{Re_d}$

$$\mathrm{Re_d} = \frac{2 a_{\rm dst} |\Delta \boldsymbol{v}|}{\nu_{\rm m}}, \tag{2.53}$$

where $\nu_{\rm m}$ is the molecular viscosity coefficient. The stopping time in the four regimes for one grain of mass $m_{\rm gr}$ and density $\rho_{\rm gr}$ reads:

$$t_{\rm s} \equiv \frac{\rho_{\rm gr} |\Delta \boldsymbol{v}|}{f_{\rm D}} = \begin{cases} \dfrac{a_{\rm dst} \rho_{\rm gr}}{c_{\rm s} \rho_{\rm g}} \sqrt{\dfrac{\pi \gamma}{8}} & \text{Epstein} \\ \dfrac{1}{9} \dfrac{\rho_{\rm gr} a_{\rm dst}}{\rho_{\rm g} |\Delta \boldsymbol{v}|} \mathrm{Re_d} & \text{Stokes I} \quad \mathrm{Re_d} < 1 \\ \dfrac{1}{9} \dfrac{\rho_{\rm gr} a_{\rm dst}}{\rho_{\rm g} |\Delta \boldsymbol{v}|} \mathrm{Re_d}^{0.6} & \text{Stokes II} \quad 1 \lesssim \mathrm{Re_d} \lesssim 800 \\ 2.3 \times \dfrac{8}{3} \dfrac{\rho_{\rm gr} a_{\rm dst}}{\rho_{\rm g} |\Delta \boldsymbol{v}|} & \text{Stokes II} \quad \mathrm{Re_d} > 800 \end{cases}, \tag{2.54}$$

where γ is the adiabatic index.

The reader should note that the typical value of $\bar{\lambda}_{\rm mfp}$ in protoplanetary systems is $\bar{\lambda} \approx 10\,\mathrm{m} \times (10^{-13} \mathrm{g\,cm^{-3}}/\rho_{\rm g})$ (Laibe et al. 2012). This implies that in most situation the drag force is effective in the Epstein regime.

A useful dimensionless parameter for quantifying the intensity of the drag force with respect to the dynamical forces in action is the so called "Stokes" number

$$\mathrm{St} = t_{\rm s} \Omega, \tag{2.55}$$

where Ω is the Keplerian frequency. Stokes numbers $\mathrm{St} \ll 1$ imply that the dust is strongly coupled to the gas; in this case, the dynamics of dust particles is indistinguishable from that of the gas. In contrast, in a dust-gas mixture with $\mathrm{St} \gg 1$, the coupling between gas and the dust is extremely weak, implying that the two fluids evolve separately, without influencing each other. The regime of $\mathrm{St} \approx 1$ is the so called "marginally coupled" regime, where the effects of the coupling between dust and gas are expected to alter most significantly the dynamics of the two fluids.

In the Epstein regime, using the relation $c_{\rm s} \rho_{\rm g} = \Sigma_{\rm g} \Omega$ in Eq. (2.54) we can easily notice that for typical protoplanetary discs with $\Sigma_{\rm g} = 1\,\mathrm{g\,cm^{-2}}$, grains with sizes $a_{\rm dst} \sim 1$ mm have $\mathrm{St} \sim 1$.

2.7.2 Dust trapping in pressure maxima and radial drift

It can be shown that the evolution of the difference $\Delta\boldsymbol{v}$ between the velocity of dust $\boldsymbol{v}_\mathrm{d}$ and the gas $\boldsymbol{v}_\mathrm{g}$ can be effectively rewritten as follows (Youdin & Goodman 2005; Laibe & Price 2014a)

$$\frac{\partial\Delta\boldsymbol{v}}{\partial t} + (\boldsymbol{v}\cdot\nabla)\Delta\boldsymbol{v} = -\frac{\Delta\boldsymbol{v}}{t_\mathrm{s}} + \frac{\nabla p}{\rho_\mathrm{g}} - (\Delta\boldsymbol{v}\cdot\nabla)\boldsymbol{v} + \frac{1}{2}\nabla\left(\frac{\rho_\mathrm{g}-\rho_\mathrm{d}}{\rho_\mathrm{g}+\rho_\mathrm{d}}\Delta\boldsymbol{v}^2\right). \tag{2.56}$$

where $\boldsymbol{v}$ here is the so called "barycentric" velocity

$$\boldsymbol{v} = \frac{\rho_\mathrm{g}\boldsymbol{v}_\mathrm{g} + \rho_\mathrm{d}\boldsymbol{v}_\mathrm{d}}{\rho_\mathrm{g}+\rho_\mathrm{d}}. \tag{2.57}$$

It can be shown (e.g., see Laibe & Price 2014a) that assuming $\mathrm{St} \ll 1$, i.e. dust and gas are strongly coupled, Eq. (2.56) reduces to

$$\Delta\boldsymbol{v} = \frac{\nabla p}{\rho_\mathrm{g}} t_\mathrm{s}. \tag{2.58}$$

This equation goes under the name of "terminal velocity limit". It represents a sort of stationary condition where the net acceleration $\boldsymbol{a}_\mathrm{d} - \boldsymbol{a}_\mathrm{g}$ acting to increase the velocity difference between dust and gas particles is rapidly balanced by the drag ($d\Delta\boldsymbol{v}/dt = 0$); in this situation, $\Delta\boldsymbol{v} = t_\mathrm{s}(\boldsymbol{a}_\mathrm{d} - \boldsymbol{a}_\mathrm{g})$ does not change in time and dust particles move at constant velocity with respect to the gas, directed as the pressure gradient in the gas. This has an extremely important consequence: taking the divergence of Eq. 2.58, provides the following result (Youdin 2010)

$$\nabla\cdot\Delta\boldsymbol{v} = \nabla\cdot\left(\frac{\nabla p}{\rho_\mathrm{g}} t_\mathrm{s}\right), \tag{2.59}$$

Assuming $\rho_\mathrm{g} = \mathrm{const}$ (i.e. also $t_\mathrm{s} = \mathrm{const}$) and $\nabla\cdot\boldsymbol{v}_\mathrm{g} = 0$, one gets

$$\nabla\cdot\boldsymbol{v}_\mathrm{d} = t_\mathrm{s}\frac{\nabla^2 p}{\rho_\mathrm{g}}. \tag{2.60}$$

This tells us that even when the gas density is constant and the gas velocity field solenoidal, dust particles tend to accumulate in pressure maxima ($\nabla^2 p < 0$), while they are repelled away from pressure minima. This phenomenon goes under the name of "dust trapping".

Furthermore, Eq. (2.58) also shows that, in general, dust particles drift in the direction of growing pressure. In protoplanetary discs, where a negative pressure gradient in the radial direction is generally present, this causes the dust to drift inward. The drift speed depends on the intensity of the aerodynamical coupling. Relaxing the strong coupling assumption we made for Eq. (2.58), it can be shown that for $\rho_\mathrm{d}/\rho_\mathrm{g} \ll 1$ the following general result holds

$$v_{R,\mathrm{d}} = \frac{|\Delta\boldsymbol{v}|}{\mathrm{St}+\mathrm{St}^{-1}} + \frac{v_\nu}{1+\mathrm{St}^2}, \tag{2.61}$$

where v_ν represent the viscous radial velocity of the unperturbed gas (Eq. 2.25). One can clearly see that when $\mathrm{St} \approx 1$ the velocity of the dust radial drift is

maximum. The rapid inward drift of marginally coupled dust grains constitutes a problem in the context of planet formation: if the dust drifts inward too fast, there is no time to form planets. Possible solutions to the problem invoke the presence of pressure traps in the disc, or a peculiar density and temperature structure able to slow down or even to halt the radial drift.

Chapter

3

Tidal torque produced by a satellite

"...I've been around for a long, long year
Stole many a man's soul to waste..."

The Rolling Stones, Sympathy for the Devil

In this chapter we discuss the formalism developed starting from the late 1970s to account for the presence of a secondary point-like mass (i.e. a satellite) embedded in the accretion disc orbiting a primary central object. This theory was originally developed to model the tidal interaction taking place between Saturn's moons and its rings. It later found many applications both in the field of protoplanetary discs and in that of supermassive black hole binaries.

A satellite embedded in the disc perturbs the gravitational potential of the primary. This induces a "tidal torque" that the satellite exerts on the accretion disc, and vice versa. On the one hand, this torque promotes the transfer of angular momentum from the satellite to the disc, inducing perturbations in the disc density structure. On the other hand, as a consequence of Newton's third law, the disc exerts a back reaction torque on the satellite, causing the evolution of its orbital parameters (semimajor axis $a_{\rm s}$ and eccentricity e of its orbit).

This chapter is dedicated to the formalism that leads to the expression of the formulae ruling the intensity of this tidal torque. Satellite migration will be discussed in the following chapters.

We start providing a simplified derivation of the tidal torque, in the so called "impulse approximation" (Sec. 3.1, Lin & Papaloizou 1979). We will then discuss a more thorough derivation of the tidal torque (Goldreich & Tremaine 1979, 1980; Ward 1986; Meyer-Vernet & Sicardy 1987;Artymowicz 1993a; Tanaka et al. 2002; Baruteau & Masset 2008), stating the equations ruling the evolution of linear perturbations (Sec. 3.2) and further developing them for the special case of

bar-like perturbing gravitational potential (Sec. 3.3). This special case is particularly relevant: the gravitational potential of any satellite can be decomposed in a summation of bar-like perturbing potentials (Sec. 3.4); given the linear nature of the perturbations, the solution for the more general case where the perturbing potential is that of a satellite can be written by summing the contributions of individual bar-like modes.

The equations ruling the evolution of the perturbations are singular at some specific locations in the disc, called "resonances". Resonances can be found where the disc orbital period is an integer multiple or fraction of that of the satellite (Sec. 3.5). The perturbations are excited at resonant locations and they propagate throughout the disc as waves, carrying angular momentum that is progressively deposited in the disc as they are damped (by viscosity or by shock steepening).

We derive the expression of the tidal torque for bar-like potentials (Sec. 3.6) and finally discuss how the individual bar-like perturbations can be assembled to provide the expression of the tidal torque produced by a satellite on a circular orbit (Sec. 3.7). In Appendix B, some further mathematical details of the derivation presented in this chapter can be found.

Useful reviews of the topic can be found in Papaloizou et al. (2007) Masset (2008), Hahn (2009), Kley & Nelson (2012) Baruteau & Masset (2013), Baruteau et al. (2014), Paardekooper & Johansen (2018).

3.1 Impulse approximation

The analytical expression of the tidal torque exerted by a satellite on a gaseous disc can be obtained using the "impulse approximation" model (Lin & Papaloizou 1979; Armitage 2010). This model describes the exchange of angular momentum between the satellite and one individual "fluid" element during a close encounter, as sketched in Fig. 3.1. The total angular momentum exchange with the entire disc is then inferred starting from the results for the fluid element.

In this section we will firstly introduce the assumptions behind this approximation; then we will discuss the problem more quantitatively, deriving an expression for the tidal torque.

In the frame of reference comoving with the satellite, a fluid element approaches the satellite on a straight trajectory and velocity Δv. The satellite has a mass M_s and orbits at a distance a_s a central primary object with mass $M_\star$. The trajectory of the fluid element is characterized by an impact parameter $b \ll a_\mathrm{s}$ with respect to the satellite (as sketched in Fig. 3.1). After the encounter with the satellite, the fluid element gains a perpendicular velocity $\delta v_\perp$; then, assuming the conservation of energy during the interaction, we can compute the variation of the parallel velocity $\delta v_{||}$ with the associated angular momentum variation. Assuming that this change in the angular momentum of the fluid element occurs once every complete orbit allows us to derive the expression of the torque exerted by the satellite on the fluid element during the interaction.

This model relies on five main assumptions:

1. The viscosity is able to dissipate the perpendicular component of the velocity before the next encounter. In fact, this requirement is equivalent to assume

that the fluid element moves on circular orbits before and after the encounter.

2. The motion and the interaction of the fluid element with the perturbing satellite neglect all hydrodynamical effects. The only fluid behaviour that is considered here are the viscous effects mentioned in assumption 1.

3. The mass of the satellite is much smaller than the mass of the central object $M_\mathrm{s} \ll M_\star$.

4. The interaction is impulsive: the effects of the interaction modify the trajectory of the fluid element all at once when its distance from the satellite is the closest[1]; i.e. an imaginary line divides the space in two distinct regions: before and after interaction.

5. During the interaction the fluid element does not change significantly the direction of its initial trajectory ($\delta v_\perp \ll \Delta v$).

Here below we discuss more quantitatively the impulsive approach we just described.

During the encounter with the satellite, the fluid element moves from $x = -\infty$ to $x = \infty$ gaining a perpendicular velocity $\delta v_\perp$ given by

$$\delta v_\perp = \int_{-\infty}^{\infty} \frac{F_\perp}{m} \frac{dx}{\Delta v} = \frac{2GM_\mathrm{s}}{b\Delta v}. \tag{3.1}$$

The fluid element conserves its initial energy after the encounter, so that

$$\Delta v^2 = \delta v_\perp + (\Delta v - \delta v_{||})^2. \tag{3.2}$$

Using assumption 5., we obtain

$$\delta v_{||} \approx \frac{1}{2\Delta v} \left(\frac{2GM_\mathrm{s}}{b\Delta v} \right)^2. \tag{3.3}$$

We note that the variation in the parallel velocity occurs always in the opposite direction with respect to that of Δv. This implies that during the interaction a fluid element on an inner orbit loses angular momentum while a fluid element on an outer orbit gains it. In order to obtain the correct sign in the angular momentum change it is thus sufficient to choose $\Delta v \to -\Delta v$ consistently. This change in the parallel velocity produces a variation of the angular momentum per unit mass Δj of the fluid element given by

$$\Delta j = \delta v_{||}(a_\mathrm{s} + b) \approx -\frac{2G^2 M_\mathrm{s}^2 a_\mathrm{s}}{b^2 \Delta v^3}. \tag{3.4}$$

The angular momentum change of the fluid element occurs once per encounter, thus, assuming the Keplerian motion of the satellite and of the fluid element, we compute the time lag between two encounters as follows:

$$\Delta t = \frac{2\pi}{|\Omega - \Omega_\mathrm{s}|} \approx \frac{4\pi}{3} \frac{a_\mathrm{s}}{|b|} \Omega_\mathrm{s}^{-1}, \tag{3.5}$$

[1] This assumption justifies considering a straight trajectory for the fluid element, since we are here considering a short section of a long circular orbit

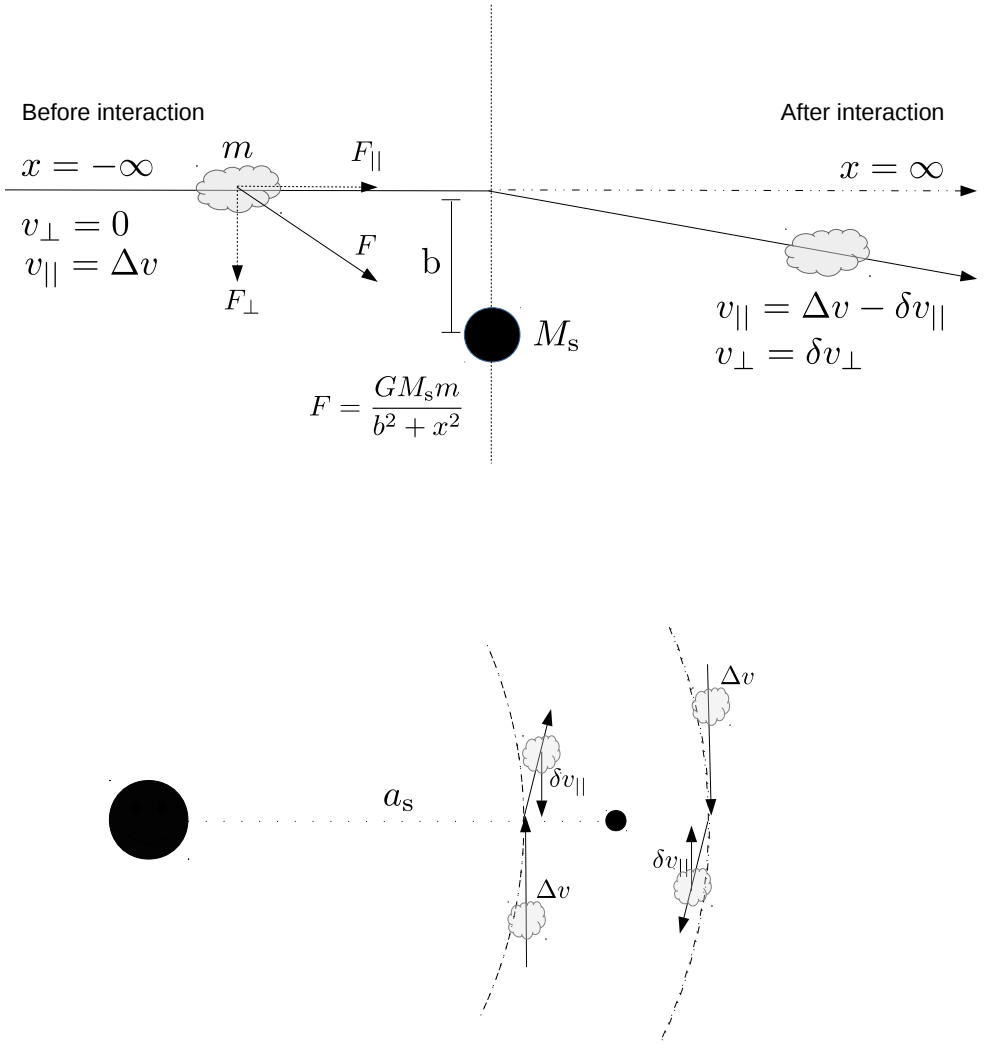


Figure 3.1: Top panel: Sketch of the physical interpretation of the impulse approximation. $\delta v_{||}$ is the velocity component parallel to the initial velocity Δv, that is the velocity difference between the fluid element and the satellite, $v_\perp$ is the perpendicular component of the velocity that the fluid element develops after the close encounter. b is the impact parameter. Bottom panel: Sketch showing the direction of the variation of the parallel velocity $\delta v_{||}$ for fluid elements on inner and outer orbits with respect to that of the satellite. Both top and bottom panel refer to the frame co-rotating with the satellite anticlockwise.

where $\Omega(R) = \sqrt{GM_\star/R^3}$ and $\Omega_\mathrm{s} = \Omega(a_\mathrm{s})$ are the angular frequencies of the fluid element and of the satellite, respectively. In our notation, R is the radial coordinate of the fluid element, $R = a_\mathrm{s} + b$; we note that our sign convention prescribes $b > 0$ for fluid elements on outer orbits, $b < 0$ for inner orbits. Using again the assumption of Keplerian motion, we compute Δv as

$$\Delta v = [\Omega(R) - \Omega_\mathrm{s}](a_\mathrm{s} + b) \approx -\frac{3}{2}\Omega_\mathrm{s}\frac{b}{a_\mathrm{s}}R. \tag{3.6}$$

Substituting Eq. (3.6) in Eq. (3.4), dividing it by Eq. 3.5 we obtain the torque exerted by the satellite (torque per unit mass) on the fluid element:

$$\frac{dT_\mathrm{s}}{dm} = \mathrm{sgn}(R - a_\mathrm{s})\frac{4}{9\pi}q^2\Omega^2(R)a_\mathrm{s}^2\frac{a_\mathrm{s}^4}{b^4} \tag{3.7}$$

where $q = M_\mathrm{s}/M_\star$ and $\mathrm{sgn}(b)$ is the sign function of the impact parameter, that is positive for the outer orbits and negative for the inner ones.

Equation (3.7) gives the torque per unit mass exerted on a fluid element. We can now derive the torque density exerted by the satellite on the whole disc. The mass of the fluid element is given by:

$$dm = \Sigma\, RdR\, d\varphi, \tag{3.8}$$

where Σ is the gas surface density and φ the azimuthal cylindrical coordinate.

Assuming axial symmetry for the gas surface density, the torque density on the disc is

$$\frac{1}{R}\frac{dT_\mathrm{s}}{dR} = \mathrm{sgn}\,(R - a_\mathrm{s})\frac{8}{9}q^2\Omega^2(R)a_\mathrm{s}^2\Sigma\frac{a_\mathrm{s}^4}{(R - a_\mathrm{s})^4}. \tag{3.9}$$

As a consequence of Newton's third law, the satellite is subject to a back reaction torque exerted by the disc that has the same intensity as the one the satellite exerts on the disc but with opposite sign.

We compute the total torque the disc exerts on the satellite changing the sign and integrating Eq. (3.9); we denote this quantity with the letter Γ (also in the following chapters the letter T indicates the torque exerted by the satellite on the disc and Γ vice versa)

$$\Gamma_\mathrm{s} = -\frac{8}{9}q^2a_\mathrm{s}^2\int_{R_0}^{\infty}\mathrm{sgn}\,(R - a_\mathrm{s})\Omega^2(R)\Sigma(R)\frac{a_\mathrm{s}^4}{(R - a_\mathrm{s})^4}\,RdR. \tag{3.10}$$

Assuming a disc density profile $\Sigma(R) = \Sigma_0(R/a_\mathrm{s})^{-1}$, and assuming $R_0 \gtrsim a_\mathrm{s}$ the expression for the torque exerted by an outer disc on the satellite Γ_s is approximately

$$\Gamma_\mathrm{s} \approx q^2\Omega_\mathrm{s}^2a_\mathrm{s}^4\Sigma_0\frac{a_\mathrm{s}^3}{(R_0 - a_\mathrm{s})^3}. \tag{3.11}$$

Despite the simplicity of this analysis, we will see that the results obtained in this section are in very good agreement with the predictions based on a more thorough derivation that we will discuss in the next sections.

In particular, we will see that the intensity of the tidal torque exerted by the satellite on the disc, and vice versa, depends on the square of the mass ratio q

between secondary and primary object. Furthermore, the intensity of this torque decreases steeply with the distance from the satellite. Finally, the overall contribution of the outer disc exerts a negative torque on the satellite, causing the satellite to migrate inward. Conversely, the inner disc exerts a positive torque on the satellite, causing the satellite to migrate outward.

3.2 Disc response to a perturbing potential

In this section we discuss the perturbations induced in a gaseous disc by the presence of a satellite embedded in it. From the mathematical point of view, this consists in studying how the properties of the disc are altered when the second point mass perturbs the gravitational potential $\Phi_\star$ in Eq. (2.2). We assume small perturbations X_1 such that the generic disc quantity X is described by $X = X_0 + X_1$, where X_0 represents the unperturbed disc and satisfies the radial equilibrium conditions discussed in Sec. 2.2. To this aim, we consider only the first order perturbed quantities (linear approximation), and thus assume that the perturbation is small with respect to the unperturbed state $X_1 \ll X_0$.

The first order response of a gaseous disc to the presence of a generic perturbing potential Φ_1 is described by the following set of equations (Goldreich & Tremaine 1979; Meyer-Vernet & Sicardy 1987)

$$\frac{\partial \Sigma_1}{\partial t} + \nabla \cdot (\Sigma_0 \boldsymbol{v}_1) + \nabla \cdot (\Sigma_1 \boldsymbol{v}_0) = 0, \tag{3.12}$$

$$\frac{\partial \boldsymbol{v}_1}{\partial t} + (\boldsymbol{v}_0 \cdot \nabla)\boldsymbol{v}_1 + (\boldsymbol{v}_1 \cdot \nabla)\boldsymbol{v}_0 = -\nabla(\Phi_1 + \Phi_{\Sigma_1} + h_1), \tag{3.13}$$

where the term Φ_{Σ_1} accounts for the additional gravitational force produced by the perturbation of the disc density Σ_1. It satisfies Poisson's equation

$$\nabla^2 \Phi_{\Sigma_1} = \frac{1}{R}\frac{\partial}{\partial R}\left(R\frac{\partial \Phi_{\Sigma_1}}{\partial R}\right) + \frac{1}{R^2}\frac{\partial^2 \Phi_{\Sigma_1}}{\partial \varphi^2} + \frac{\partial^2 \Phi_{\Sigma_1}}{\partial z^2} = 4\pi G \Sigma_1 \delta(z). \tag{3.14}$$

The term Φ_{Σ_1} can be set to zero if the disc self-gravity is assumed to be negligible. The quantity h_1 is the perturbed enthalpy:

$$h_1 = c_{\mathrm{s},0}^2 \left(\frac{\Sigma_1}{\Sigma_0}\right), \tag{3.15}$$

where $dP/d\Sigma = c_{s,0}^2$ is the unperturbed sound speed, using a generic barotropic equation of state of the type

$$P = K\rho^\gamma \tag{3.16}$$

The vectorial form of Eq. (3.13) can be written explicitly for each component[2] in the following way:

$$\left(\frac{\partial}{\partial t} + \Omega(R)\frac{\partial}{\partial \varphi}\right) v_{R,1} - 2\Omega(R) v_{\varphi,1} = -\frac{\partial}{\partial R}(\Phi_1 + \Phi_{\Sigma_1} + h_1), \tag{3.17}$$

$$\left(\frac{\partial}{\partial t} + \Omega(R)\frac{\partial}{\partial \varphi}\right) v_{\varphi,1} + 2B(R) v_{R,1} = -\frac{1}{R}\frac{\partial}{\partial \varphi}(\Phi_1 + \Phi_{\Sigma_1} + h_1). \tag{3.18}$$

[2] $\boldsymbol{v}_1 = v_{R,1}\hat{\boldsymbol{e}}_R + v_{\varphi,1}\hat{\boldsymbol{e}}_\varphi$, where $\boldsymbol{e}_R$ and $\boldsymbol{e}_\varphi$ are the radial and azimuthal unit vectors, respectively.

where $B(R)$ is the Oort parameter

$$B(R) = \Omega(R) + \frac{R}{2}\frac{d}{dR}\Omega(R) \tag{3.19}$$

that relates with the epicyclic frequency κ (see Eq. 1.40) as follows[3]

$$\kappa^2(R) = 4B(R)\Omega(R). \tag{3.20}$$

Finally, the continuity equation (3.12) can be rewritten as

$$\left(\frac{\partial}{\partial t} + \Omega(R)\frac{\partial}{\partial \varphi}\right)\Sigma_1 = -\frac{1}{R}\frac{\partial}{\partial R}(R\Sigma_0 v_{R,1}) - \frac{1}{R}\frac{\partial}{\partial \varphi}(\Sigma_0 v_{\varphi,1}). \tag{3.21}$$

3.3 Equations for a "bar-like" perturbing potential

We study the response of the disc to a "bar-like" gravitational potential. We set $\Phi_1 = \Phi_1^m$, with

$$\Phi_1^m = \psi(R)e^{i(m\varphi - \omega t)}. \tag{3.22}$$

The function $\psi(R)$ is purely real, and the term $e^{i(m\varphi-\omega t)}$ is a complex phase. This choice is motivated by the fact that, as we will see in Sec. (3.4), the satellite perturbing potential can be decomposed in a summation over m of Φ_1^m potentials.

The complex notation is extremely helpful to describe periodic oscillations. However, the physical evolution of any perturbed quantity is given by its real part only, so that the physical potential is actually given by

$$\mathrm{Re}[\Phi_1^m] = \psi(R)\cos(m\varphi - \omega t), \tag{3.23}$$

where $\mathrm{Re}[z]$ indicates the real part of $z \in \mathbb{C}$. Hereafter we will drop the Re notation, unless differently prescribed.

The form of $e^{i(m\varphi-\omega t)}$ can be thought as a m-armed bar wave; the wave propagates in the disc with pattern frequency $\Omega_\mathrm{p} = \omega/m$, that represents the angular velocity at which the arms of the perturbation span the disc. The parameter m relates with the azimuthal wave number of the wave as $k_\varphi = m/R$.

We assume that the response of the disc is stationary, i.e. the satellite has existed for a sufficiently long time to enable the perturbed quantities to have an analogous azimuthal wave-like dependence (Meyer-Vernet & Sicardy 1987) and in particular, for the perturbed density, velocity and self-gravity we define

$$\Sigma_1 = \sigma_1(R)e^{i(m\varphi-\omega t)}, \tag{3.24}$$

$$v_{R,1} = v_{R,1}(R)e^{i(m\varphi-\omega t)}, \tag{3.25}$$

$$v_{\varphi,1} = v_{\varphi,1}(R)e^{i(m\varphi-\omega t)}, \tag{3.26}$$

$$\Phi_{\Sigma_1} = \phi_{\Sigma_1}(R)e^{i(m\varphi-\omega t)}. \tag{3.27}$$

It is important to note that while $\psi(R)$ of the bar potential is a real function, no assumption has been done so far on $\sigma_1(R)$, $v_{R,1}(R)$, $v_{\varphi,1}(R)$ and $\phi_{\Sigma_1}(R)$, so that

[3]Note that since for Keplerian discs $\kappa(R) = \Omega_\mathrm{K}(R)$, it follows that $B(R) = \Omega_\mathrm{K}(R)/4$.

they can be complex functions. As a consequence, an additional complex phase lag can actually "wind" the bar transforming it into a spiral.

Under these assumptions the equations describing the perturbation of velocities are

$$i[m\Omega(R)-\omega]v_{R,1}-2\Omega(R)v_{\varphi,1}=-\frac{\partial}{\partial R}(\Phi_1^m+\Phi_{\Sigma_1}+h_1), \tag{3.28}$$

$$2B(R)v_{R,1}+i[m\Omega(R)-\omega]v_{\varphi,1}=-\frac{im}{R}(\Phi_1^m+\Phi_{\Sigma_1}+h_1), \tag{3.29}$$

while Eq. (3.21) becomes

$$i[m\Omega(R)-\omega]\Sigma_1=-\frac{1}{R}\frac{\partial}{\partial R}(R\Sigma_0 v_{R,1})-i\frac{m}{R}\Sigma_0 v_{\varphi,1}. \tag{3.30}$$

After a little algebra, equations (3.28,3.29) can be rewritten as

$$v_{R,1}=-\frac{i}{D(R)}\left\{[m\Omega(R)-\omega]\frac{\partial}{\partial R}+\frac{2m\Omega(R)}{R}\right\}(\Phi_1^m+\Phi_{\Sigma_1}+h_1), \tag{3.31}$$

$$v_{\varphi,1}=\frac{1}{D(R)}\left\{2B(R)\frac{\partial}{\partial R}+\frac{m}{R}[m\Omega(R)-\omega]\right\}(\Phi_1^m+\Phi_{\Sigma_1}+h_1), \tag{3.32}$$

where $D(R)$ is given by

$$D(R)=\kappa^2(R)-[m\Omega(R)-\omega]^2. \tag{3.33}$$

The equation for Σ_1 finally reads

$$\Sigma_1=\frac{i}{[m\Omega(R)-\omega]}\left[\frac{1}{R}\frac{\partial}{\partial R}(R\Sigma_0 v_{R,1})+i\frac{m}{R}v_{\varphi,1}\Sigma_0\right]. \tag{3.34}$$

Eq.s (3.31), (3.32) and (3.34) present singularities where $D(R)=0$ and when $m\Omega(R)-\omega=0$. Locations where $D(R)=0$ are called Lindblad resonances, while locations where $m\Omega(R)-\omega=0$ is satisfied are called "corotation" resonances; we postpone a more detailed discussion about these locations to Sec. 3.5.

Finally, the dependence of Φ_{Σ_1} on Σ_1 can be obtained by solving Eq. (3.14). Equations (3.31), (3.32), (3.34) with the addition of Eq.s (3.15) and (3.14) constitute the closed set of equations ruling the evolution of perturbations of the fluid quantities due to the presence of an m-barred potential Goldreich & Tremaine (1979).

In order to improve the readability of this chapter, further details and developments of the equations presented in this section, including the formal solution under some simplifying assumptions to the density wave equation, have been moved to Appendix B.

3.4 Harmonic decomposition of the satellite potential in bar-like potentials

The gravitational potential $\Phi_{\rm s}$ of a point-like mass (i.e. the satellite) located at $\boldsymbol{r}_{\rm s}(t)$, computed at the position $\boldsymbol{r}$ at time t is

$$\Phi_{\rm s}(\boldsymbol{r},t)=-\frac{GM_{\rm s}}{|\boldsymbol{r}_{\rm s}(t)-\boldsymbol{r}|}. \tag{3.35}$$

Assuming that the satellite orbits with Keplerian motion (epicyclic frequency $\kappa_\mathrm{s} = \Omega_\mathrm{s}$) a central star $M_\star$, where we set the centre of our coordinates, with an orbital frequency Ω_s, we can decompose Φ_s in a summation of rotating "bar"-potentials Φ_1^{lm} as follows (Goldreich & Tremaine 1980; Artymowicz & Lubow 1994)

$$\Phi_s(R,\varphi,t) = \sum_l \sum_m \underbrace{\psi_1^{lm}(R)\,\mathrm{Re}\left[e^{i(m\varphi - l\Omega_\mathrm{s}t)}\right]}_{\Phi_1^{lm}(R)} \quad l,m \in \mathbb{N}, \tag{3.36}$$

where $\psi^{lm}(R)$ is a purely real function of the radius.

Eq. (3.36) tells us that we can decompose the potential of a satellite in a summation of "bar"-like potentials Φ_1^m, for which we have studied the disc response in the previous section. This has a very important implication for our analysis: given the linear nature of our perturbation analysis, we can study the response of the disc to the individual "bar" modes of the potential, and then find the final solution summing their individual contributions.

It is easy to recognize here that the frequency ω, i.e. the frequency at which the perturbation span the azimuthal domain, for the different modes of the potential is set by $\omega = l\Omega_\mathrm{s}$, implying a pattern frequency

$$\Omega_\mathrm{p}^{lm} = \frac{l}{m}\Omega_\mathrm{s}. \tag{3.37}$$

The magnitude of $\psi_1^{lm}(R)$ represents the "intensity" of the mode (l,m) of the potential and it scales as $e^{2|l-m|}$, where e is the satellite eccentricity. In Fig. 3.2 the location and relative intensity of a set of resonances are shown (Artymowicz & Lubow 1994). The expression $\psi_1^{lm}(R)$ has the following analytical form (Goldreich & Tremaine 1980)

$$\psi_1^{lm}(R) = \frac{1}{\pi^2}\int_0^{2\pi} d(\Omega_\mathrm{s}t)\int_0^{2\pi} \Phi_\mathrm{s}(R,\varphi,t)\cos(m\varphi - l\Omega_\mathrm{s}t)\,d\varphi \tag{3.38}$$

If the satellite has a circular orbit, i.e. $e = 0$, only the modes $m = l$ contribute to the summation. In particular, for a satellite on circular orbit, ψ^{mm} for $m \geq 2$ reads

$$\psi_1^{mm}(R) = -\frac{GM_\mathrm{s}}{a_\mathrm{s}} b_{1/2}^{(m)}(\alpha)\,, \tag{3.39}$$

where a_s is the semi-major axis of the satellite orbit, $\alpha = R/a_\mathrm{s}$, and $b_{1/2}^{(m)}(\alpha)$ is the m-th Laplace coefficient of order 1/2

$$b_{1/2}^{(m)}(\alpha) = \frac{1}{\pi}\int_0^{2\pi} \frac{\cos(m\theta)}{(1-2\alpha+\alpha^2)^{1/2}} d\theta. \tag{3.40}$$

3.5 Lindblad and corotation resonances for a perturbing satellite

We showed in Sec. 3.3 that if Φ_1 is a generic bar-like perturbation of the type Φ_1^m (Eq. 3.22) with frequency ω and pattern frequency Ω_p, Eq.s (3.31,3.32) and Eq.

(3.34) present singularities when $m\Omega(R) - \omega = 0$ and when $D(R) = 0$. In any shearing disc, these relations are satisfied at specific locations called resonances.

In this section, we require that Φ_1 is produced by a satellite (i.e. $\Phi_1 = \Phi_\mathrm{s}$, Eq. 3.36) orbiting with Keplerian motion the primary object. We discuss the location of resonances for a generic shearing gaseous disc, then we will discuss the Keplerian disc case. The frequency of the perturbations ω is $\omega = m\Omega_\mathrm{p}^{lm}$, with the pattern frequency Ω_p^{lm} defined in Eq. (3.37).

The condition $m\Omega(R) - \omega = 0$ is satisfied at "corotation" resonances (usually abbreviated "CR"), located where the orbital frequency of the disc $\Omega(R)$ equals the pattern frequency of the bar-like perturbation, i.e. at $R = R_c$

$$\Omega(R_c) = \Omega_\mathrm{p}^{lm}. \tag{3.41}$$

The condition $D(R) = 0$ is satisfied instead at "Lindblad" resonances, located where the Doppler shifted frequency of the material in the disc with respect to the pattern frequency Ω_p equals the epicyclic frequency $\kappa(R)$, i.e. those radii $R = R_L$ where

$$\Omega(R_L) - \Omega_\mathrm{p}^{lm} = \epsilon\frac{\kappa(R_L)}{m}, \tag{3.42}$$

where $\epsilon = \mathrm{sgn}[\Omega(R) - \Omega_\mathrm{p}^{lm}]$. In discs where the angular frequency $\Omega(R)$ decreases with the radius (such as Keplerian discs $\Omega(R) \propto R^{-3/2}$), Lindblad resonant locations with the $\epsilon = +1$ on the r.h.s. are called inner Lindblad resonances (ILRs), since they fall inside the corotation radius R_c in the inner part of the disc; while those characterized by the $\epsilon = -1$ sign outer Lindblad resonances (OLRs).

We note that for modes $l = m$, the corotation radius coincide with the semi-major axis of the satellite so that $\Omega_\mathrm{p}^{mm} = \Omega_\mathrm{s}$. In contrast, modes with $l \neq m$, arising when the satellite has an eccentric orbit, provide a series of additional corotation resonances where $\Omega(R)$ matches the condition in Eq. (3.41). Each $l \neq m$ corotation resonance has its own ILRs and OLRs. Some of these may overlap with satellite orbit, since they have a different corotation radius; when this situation occurs, they are referred to as co-orbital Lindblad resonances.

Assuming Keplerian rotation of the disc, i.e. $\kappa(R) = \Omega(R)$ and $\Omega(R) = \Omega_\mathrm{K}$, Lindblad resonances satisfy

$$\Omega(R_L) = \frac{m}{m-\epsilon}\Omega_\mathrm{p}^{lm}, \tag{3.43}$$

This implies that at Lindblad resonances the material in the disc makes $m - \epsilon$ orbits every m orbits of the perturbing potential.

The radii where corotation and Lindblad resonances are located are

$$R_c = \left(\frac{m}{l}\right)^{2/3} a_\mathrm{s}, \tag{3.44}$$

$$R_L = \left(\frac{m-\epsilon}{l}\right)^{2/3} a_\mathrm{s}. \tag{3.45}$$

We also note that, for Keplerian discs, ILRs with $(l, m) = (l, 1)$ fall into the origin for any choice of l; this implies that Keplerian ILRs always have $m \geq 2$.

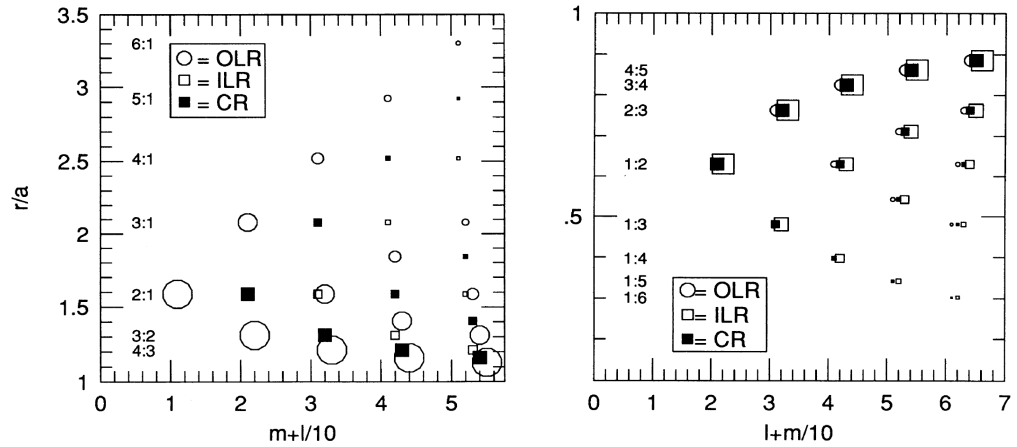


Figure 3.2: Plot of location and intensity of (l, m) resonances, as a function of $m+\frac{l}{10}$ and $l+\frac{m}{10}$ (so that integer part is m while decimal part is l, and vice versa) for outer (left) and and inner (right) discs ($R \gtrless a_{\rm s}$, respectively). The size of the symbols indicates the relative intensity of each potential mode (l, m) for various $|m-l| < 5$, that scales as $\epsilon^{2|m-l|}$. Most intense resonances are of the type (m, m) (largest symbols). It can be clearly seen that, for growing m, the strongest resonances tend to $R = a$. The ILR for the mode $(1, 1)$ gives a resonance in $R = 0$ and for this reason is not present. These plots do not show co-orbital Lindblad resonances. Original images adapted from Artymowicz & Lubow (1994).

Furthermore, it is useful to note that for $m = l$ and $m \gg 1$ Eq. (3.45) can be approximated as

$$R_L = a_{\rm s}\left(1 - \frac{2\epsilon}{3m}\right). \tag{3.46}$$

Finally, in the following sections we will need to approximate the value of $D(R)$ at Lindblad resonant locations. To this aim, we introduce here the quantity $\mathcal{D}$

$$\mathcal{D} \equiv R_L \left|\frac{\partial}{\partial R} D(R_L)\right|, \tag{3.47}$$

so that in the region of Lindblad resonances $R = R_L$ we can write

$$D(R) \equiv \epsilon\mathcal{D}\frac{(R - R_L)}{R_L} = (R - R_L)\frac{\partial}{\partial R} D(R_L) \tag{3.48}$$

3.5.1 Physical interpretation of resonant locations

From the physical point of view (see Masset 2008 for a detailed discussion), it is straightforward to verify that Lindblad resonances are regions where the epicyclic motion is strongly excited in the gas. Indeed, assuming Keplerian motion of the gas (we recall that for Keplerian discs $B(R) = \Omega(R)/4$), Eq.s 3.31 and 3.32 at $R = R_L$ can be rewritten as

$$\lim_{R\to R_L} v_{R,1} = -\frac{i}{D(R)}\left\{\epsilon\Omega(R)\frac{\partial}{\partial R} + \frac{2m\Omega(R)}{R}\right\}(\Phi_1^m + \Phi_{\Sigma_1} + h_1), \tag{3.49}$$

$$\lim_{R\to R_L} v_{\varphi,1} = \frac{1}{D(R)}\left\{\epsilon\frac{\Omega(R)}{2}\frac{\partial}{\partial R} + \frac{m\Omega(R)}{R}\right\}(\Phi_1^m + \Phi_{\Sigma_1} + h_1). \tag{3.50}$$

So that we can write:

$$\lim_{R\to R_L} v_{\varphi,1} = -\frac{i}{2} v_{R,1}. \tag{3.51}$$

Eq. (3.51) implies that $v_{R,1}$ and $v_{\varphi,1}$ have a constant phase difference of $\pi/2$ and amplitude differing by a factor of 2, as one would expect for the epicyclic Keplerian motion. Such a configuration of the perturbed velocities is associated with an advective flux of angular momentum (see the Reynolds stress in Eq. 2.32 and Eq. B.6.2 in the Appendix), i.e. at these locations the exchange of angular momentum between the disc and the satellite is maximum.

Conversely, the interpretation of corotation resonances from a physical point of view is more subtle. In an m-barred potential, the streamlines around the corotation region are divided in two categories: streamlines where the material is in "libration" and those where the material is in "circulation". The left panel of Fig. 3.3 shows the geometry of the streamlines in the corotation region. The type of streamlines depends on the distance from the corotation resonance: in particular, librating streamlines are close to the corotation region, circulating ones can be found farther from it.

The first are characterized by m eye-shaped regions where the gas co-rotate on closed orbits around m driving centres (equally spaced on the azimuthal coordinate) located at the corotation radius; librating material never completes an entire orbit $[0, 2\pi]$ in the co-moving reference frame. The second are open streamlines characterized by radial oscillations that span the entire azimuthal domain in the comoving frame.

In the libration region, each fluid element orbits along its driving centre on a timescale that is much longer than the orbital period of the bar-like potential. For this reason, in an inertial reference frame, fluid elements on librating streamlines move approximately following the unperturbed circular Keplerian flow of the gas but they progressively change the radius of their orbit. Since the disc is Keplerian, in order to slowly modify their orbit to complete a libration cycle, fluid elements have to exchange angular momentum with the perturbing potential. This constant exchange of angular momentum between the perturbing potential and the material makes the corotation resonance a region where the coupling between the disc and the potential is strong.

3.5.2 Pressure shift of resonant locations

From Eq. (3.46) we note that Lindblad resonances accumulate at the corotation resonance for large values of m. However, Artymowicz (1993a) showed that Lindblad resonances with large m values are actually shifted from their nominal radius, due to pressure effects. In particular, the effective location of Lindblad resonances with $m \gg 1$ is given by (see right panel of Fig. 3.3)

$$R_{L,\mathrm{eff}} = R_c - \epsilon \frac{2H}{3\xi}\sqrt{1+\xi^2} \tag{3.52}$$

where H is the disc thickness, and ξ reads

$$\xi = m(H/R)_{R_L}. \tag{3.53}$$

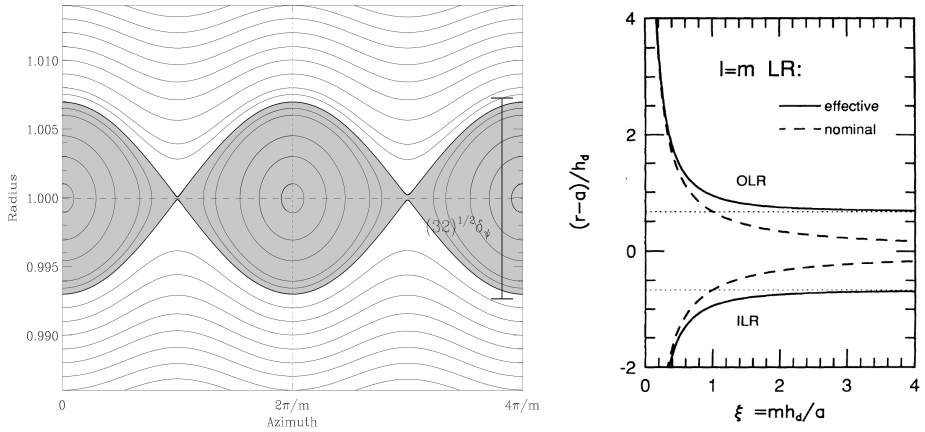


Figure 3.3: Left panel: streamlines in the plane (φ, R) in presence of an m-bar perturbing potential; the gray shaded area represents the libration islands, the white ones the "circulation" regions. Original figure in Masset (2008). Right panel: Plot of effective location of Lindblad resonances $R_{L,\text{eff}}$ (Eq. (3.52, y-axis) vs ξ (Eq. 3.53, x-axis). Where r is the distance from the primary object, a is the semi-major axis of the satellite separation and h_d is the thickness of the disc. Resonances accumulates at $R_{ILRs} = a - (2/3)H$ and $R_{OLRs} = a + (2/3)H$ due to pressure effects. Original figure in the right panel from Artymowicz (1993a)

It is easy to see that for $m \to \infty$ one gets $R_{L,\text{eff}} = R_c - \epsilon 2H/3$. From the physical point of view, this effect can be explained by noting that at a distance $2/3H$ the gas motion in the reference frame corotating at $R = R_c$ becomes supersonic, so that only beyond that location density waves can be excited (Rafikov 2002). This causes a shift of the location where density waves are effectively excited.

3.6 Tidal torque

The total torque exerted by the generic perturbing potential Φ_1 on the disc can be computed applying the standard definition of torque, using the gravitational force it produces $\boldsymbol{F}_{g,1} = -\nabla\Phi_1$:

$$T \equiv -\int_0^{2\pi}\int_0^{\infty} \hat{\boldsymbol{e}}_z \left\{R\hat{\boldsymbol{e}}_R \times \text{Re}[\nabla\Phi_1]\right\} \text{Re}[\Sigma_1]\, R\, dR\, d\varphi \tag{3.54}$$

$$= -\int_0^{2\pi}\int_0^{\infty} \left[\frac{1}{R}\,\text{Re}\left(\frac{\partial\Phi_1}{\partial\varphi}\right)\right] \text{Re}[\Sigma_1]\, R\, dR\, d\varphi. \tag{3.55}$$

It should be noted here that, given the azimuthal symmetry of the potential Φ_1, the contribution from Σ_0 would average to zero when integrated across the azimuthal domain. Furthermore, in order to have a non vanishing torque, the density perturbation needs to be phase shifted with respect to Φ_1, that is the case when solving the perturbed equations to obtain $\sigma_1(R)$. Indeed, if the two perturbations perfectly overlap, no torque can be exerted for reasons of symmetry (Meyer-Vernet & Sicardy 1987). We remind the reader that $\sigma_1(R)$ is the complex function containing the radial dependence of Σ_1 (see Eq. 3.24), it contains the complex phase that makes the bar-like perturbation a spiral (see Appendix B).

In the following subsections we discuss the expression of the torque exerted by a generic m-barred potential with generic pattern frequency $\Omega_{\rm p} = \omega/m$. Then, in Sec. 3.7, we will present how individual m-modes contribute to provide together the total torque exerted by a satellite.

3.6.1 Torque of individual modes at Lindblad resonances

Using $\Phi_1 = \Phi_1^m$, after a little algebra, and integrating Eq. (3.55) over φ, it is straightforward to obtain[4]

$$T^m = -m\pi \int_0^\infty |\Phi_1^m| \, \mathrm{Im}[\sigma_1(R)] \, R \, dR. \tag{3.58}$$

A detailed solution of the previous integral at Lindblad resonances can be found in the appendix of Goldreich & Tremaine (1979) and in Meyer-Vernet & Sicardy (1987). Using the assumption of tight wound perturbations (the length scale of the radial perturbations is much shorter than the azimuthal one, see Appendix B.1), the tidal torque at Lindblad resonances reads

$$T_L^m = -\frac{\pi^2 m \Sigma_0}{\epsilon \mathcal{D}} |\Psi_m(R_L)|^2, \tag{3.59}$$

where Ψ_m is defined as (see also Eq. B.3.2)

$$\Psi_m(R) \equiv R \frac{\partial \Phi_1^m(R)}{\partial R} + \frac{2m\Omega}{m\Omega(R) - \omega} \Phi_1^m(R), \tag{3.60}$$

Eq. (3.59) can be also obtained by explicitly solving the solution for Σ_1 and computing the angular momentum transported by such a density wave (see Appendix B.6).

3.6.2 Pressure corrections

As a consequence of the application of the tight-winding approximation, the expression of the torque in Eq. (3.59) is not valid for $m \gg (H/R)^{-1}$, where H/R is the disc aspect ratio (see Appendix B.1). Relaxing this assumption when deriving the tidal torque for m-barred potentials, Artymowicz (1993a) obtained the following corrections. Firstly, high m resonances do not accumulate at the corotation resonance but at $R_{\rm CR} \pm 2H/3$, as we previously discussed in Sec. (3.5.2).

Secondly, assuming $\psi_1^m(R) = \text{const}$ (constant bar potential), the expression of T_L^m in Eq. 3.59 is corrected as follows

$$T^m_{L,\rm corr} = \frac{\sqrt{1+\xi^2}}{1+4\xi^2} \, T_L^m|_{R_{L,\rm eff}} \, , \tag{3.61}$$

[4]One needs to keep in mind that, since $\psi(R)$ is by definition a real quantity,

$$\mathrm{Re}(\partial_\varphi \Phi_1^m) = im\psi(R) \left[e^{i(m\varphi - \omega t)} - e^{-i(m\varphi - \omega t)} \right] \frac{1}{2}, \tag{3.56}$$

while $\sigma_1(R)$ has a complex part so that

$$\mathrm{Re}(\Sigma_1) = \frac{1}{2} \left[\sigma_1(R) e^{i(m\varphi - \omega t)} + \sigma_1^\star(R) e^{-i(m\varphi - \omega t)} \right]. \tag{3.57}$$

where ξ was defined in Eq. (3.53) and T_L^m has to be computed at the effective resonant location $R_{L,\text{eff}}$ (see Eq. 3.52). The prefactor $\sqrt{1+\xi^2}/(1+4\xi^2)$ introduces what Artymowicz (1993a) defined a "mild" reduction of the torque intensity for high m values, compared with the non corrected formula in Eq. (3.59).

We will show that for the $\psi_1^m(R)$ produced by a satellite the resonance shift causes the term $T_L^m|_{R_{L,\text{eff}}}$ to have an exponential decay for large values of m (torque cut-off, see also Sec. 3.7.1, and Fig. 3.5 for a more detailed discussion).

3.6.3 Further considerations about the Lindblad tidal torque

The torque derived in Eq. (3.59) represents the so called "excitation" torque, which only represents the angular momentum transfered at resonances, and thus the torque that the disc overall exerts on the satellite changed in sign. The sign of the torque is negative at ILRs ($\epsilon = 1$), i.e. it removes angular momentum from the gas, and positive at OLRs ($\epsilon = -1$)

It is interesting to note that the flux transported by a density wave at Lindblad resonances (Eq.s B.6.5 and B.6.7) has the same magnitude as the torque in Eq. (3.59), confirming the expectation that all the angular momentum the satellite exchanges with the disc is transported away by waves.

The angular momentum flux actually does not provide any information about the effective torque the satellite exerts on the disc. Indeed, in a purely non-dissipative disc the angular momentum is transported away by the waves without being effectively deposited in the gas. The angular momentum carried by the wave is effectively transfered to the disc as a consequence of viscous damping or as a consequence of shock dissipation (Goodman & Rafikov 2001; Rafikov 2002; Crida et al. 2006; Duffell 2015). This causes the action of the torque on the disc to be non-local: indeed, the "excitation" torque may differ significantly from the "deposition" torque, with important consequences on how the presence of a perturbing satellite alters the structure of the disc in which it is embedded.

For completeness, we mention that the inclusion in the linear perturbation theory of a toroidal magnetic field (Terquem 2003; Fromang et al. 2005), causes the perturbed equations to be singular also at a new set of locations, introducing a new family of resonances: "magnetic resonances". These resonances fall both in the inner (IMRs) and in the outer disc (OMRs). Such resonances are associated with the excitation of MHD waves that carry angular momentum analogously to what occurs at Lindblad resonances. Also in this case, the excitation of these waves is thus associated with a net transfer of angular momentum from the satellite to the material. In this framework, the satellite exerts at IMRs and OMRs a torque that has the same sign as that exerted ILRs and OLRs, respectively. However, in some circumstances the contribution from the inner magnetic resonances might overcome significantly the contribution from the outer ones.

3.6.4 Tidal torque at corotation resonances

Using the same approach used in the previous section to derive the tidal torque at Lindblad resonances, Goldreich & Tremaine (1979) computed the torque exerted at corotation by a m-bar potential. In a thin isothermal disc the expression for the

corotation torque exerted by the planet on the disc at the corotation resonance reads

$$T_c^m = \frac{m\pi^2}{2}|\Phi_1^m(R)|^2 \left[\left(\frac{d\Omega}{dR}\right)^{-1} \frac{\partial}{\partial R}\left(\frac{\Sigma_0}{B}\right)\right]_{R=R_c} \tag{3.62}$$

The corotation torque depends on the radial gradient of the quantity Σ_0/B, called vortensity. In a Keplerian disc, this implies that the corotation torque vanishes if $\Sigma_0 \propto R^{-3/2}$.

Baruteau & Masset (2008) pointed out that the corotation torque depends also on the radial gradient of specific entropy. In particular in a thin disc with an adiabatic equation of state, i.e. $P = S(R)\Sigma^\gamma$, where $S(R)$ is the radial profile of the gas entropy, corresponding to a radiatively inefficient disc, the contribution to the corotation torque due to the entropy gradient is

$$T_{c,S}^m = 2m\pi^2\Sigma_0\gamma^{-1}|\Phi_1^m(R)|^2 \left[\frac{\Omega}{\kappa}\frac{d\ln S(R)}{d\ln R}\right]. \tag{3.63}$$

In the presence of an entropy gradient, the advection of entropy in the corotation region produces a change of the density that increases the vortensity gradient, resulting in an additional corotation torque (Baruteau & Masset 2008; Baruteau et al. 2014).

The physical interpretation of these two contributions can be qualitatively understood by noticing that the material, during the libration motion described in Sec. 3.5 crosses the corotation radius twice per libration cycle moving through regions characterized by different unperturbed vortensity and/or entropy. The conservation of vortensity and entropy along the streamlines in barotropic and/or adiabatic fluids forces the density to change as the streamlines span a domain of different radii. This creates an asymmetric angular momentum exchange during the two crossings that result in a net torque exerted by the disc on the potential and vice versa (Baruteau & Masset 2008; Baruteau et al. 2014).

It should be noticed that density waves are not excited in the corotation region. As a consequence, the angular momentum transfered to the disc cannot be carried away outside the libration islands. In this way, libration islands are isolated from the rest of the disc. If the advection of entropy and vorticity from outside this region is not fast enough (i.e. if the viscous and thermal timescales are much longer than the libration timescale), the two gradients become progressively shallower (Ogilvie & Lubow 2003). When this condition occurs, the corotation torque is said to be "saturated" and vanishes over a few libration timescales.

We mention here that corotation torques, should not be confused with the co-orbital torque or horseshoe drag (Ward 1991). The term co-orbital torque usually refers to the the torque exerted by a satellite perturber on the material librating in the co-orbital region of the satellite. This torque still depends on the radial gradient of the vortensity, but it arises when a non-linear perturbations are considered. For sufficiently low levels of viscosity (Paardekooper & Papaloizou 2009a), co-orbital torque has been evaluated to be much stronger than the contribution of the co-rotation resonances falling at the orbit of a satellite perturber ($m = l$ harmonics in the expansion in Eq. 3.36).

3.7 Tidal torque exerted by a satellite on a circular orbit

In Sec. 3.4 we have decomposed the potential of a satellite into a summation of (l, m)-barred potentials. Each of these individual potential has its intrinsic intensity set by the value of $\psi_1^{lm}(R)$ and pattern frequency $\Omega_{\rm p}^{lm}$. In the previous sections we have discussed the torque exerted by the generic m-barred mode of the potential. We now have all the ingredients to derive the tidal torque a satellite exerts on the disc. Since the following relation holds[5]

$$\int_0^{2\pi} \mathrm{Re}\left[\sum_{l,m} \frac{\partial \Phi_1^{lm}}{\partial \varphi}\right] \mathrm{Re}\left[\sum_{l,m} \Sigma_1^{lm}\right] d\varphi = -m\pi \sum_{l,m} |\Phi_1^{lm}|\, \mathrm{Im}[\sigma_1^{lm}(R)],$$

the total torque exerted by the satellite on the disc can be computed by considering the individual contributions computed in the previous sections as

$$T_{\rm tot} = \sum_{l,m} T^{lm}. \tag{3.64}$$

In the next sections we discuss the cumulative action of all the resonances under the simplifying assumption that the satellite has a circular orbit (only $l = m$ modes are present).

3.7.1 The contribution of Lindblad resonances

We provide the analytical estimate of the tidal torque exerted on the disc by a circular satellite. This assumption allows us to impose $m = l$, so that we can consider one single pattern frequency $\Omega_{\rm p}^{mm} = \Omega_{\rm s}$. For this reason, here and in the following sections we will refer to quantities X^{mm} as X^m.

Eq.s (3.59) and (3.62) provide the intensity of the torque produced by one single mode m of the perturbing potential Φ_1^m. Each mode exerts a certain amount of torque at a given resonant location. To do so, we start distinguishing between isolated (small m values) and overlapping (large m values) resonances.

Lindblad resonances associated with small m values are far from each other, in this condition resonances are called "isolated". The radial dependence of the tidal torque at those locations is given by Eq.s (3.59) and (3.62) smoothed out across the resonance width. This generates a localized peak in the torque intensity located at the resonance radius.

"Overlapping" resonances (large m values), as the name suggests, are much closer to each other. This causes the tidal torque intensity to grow smoothly as the distance from the satellite decreases. For large m values, it can be shown that $\psi_1^m(R)$ and $\partial_R \psi_1(R)$ entering in the $\Psi_m(R)$ term in Eq. (3.59) can be approximated[6] using the modified Bessel functions K_0 and K_1 of order 0 and 1. In

[5] This result can be derived using the Parseval's theorem.

[6] This is a consequence of the properties for large m values of the term $b_{1/2}^{(m)}(\alpha)$, appearing in the intrinsic strength of the modes discussed in Sec. 3.4.

particular, it can be shown that

$$\psi_1^m(R_L) \approx \frac{2}{\pi}\frac{GM_\mathrm{s}}{a_\mathrm{s}} K_0\left(m\left|\frac{R_L-a_\mathrm{s}}{a_\mathrm{s}}\right|\right), \tag{3.65}$$

$$\frac{d\psi_1^m}{dR}(R_L) \approx \epsilon\frac{2m}{\pi}\frac{GM_\mathrm{s}}{a_\mathrm{s}} K_1\left(m\left|\frac{R_L-a_\mathrm{s}}{a_\mathrm{s}}\right|\right). \tag{3.66}$$

From Eq. 3.46 we know that $|R_L - a_\mathrm{s}|/a_\mathrm{s} = 2/(3m)$. The intensity of the tidal torque at the m-th Lindblad resonance reads (Goldreich & Tremaine 1980; Meyer-Vernet & Sicardy 1987):

$$T_L^m = -\epsilon m^2 \Sigma_0 a_\mathrm{s}^4 \Omega_\mathrm{s}^2 q^2 \frac{4}{3}\left[2K_0\left(\frac{2}{3}\right) + K_1\left(\frac{2}{3}\right)\right]^2. \tag{3.67}$$

Furthermore, using Eq. (3.46), one can define a continuous distribution of m resonances $m(R)$

$$m(R) = \epsilon\frac{2}{3}\frac{a_\mathrm{s}}{R-a_\mathrm{s}}, \tag{3.68}$$

describing how m varies as a function of the radius. Substituting $m(R)$ into Eq. (3.67), we get an expression of the intensity of overlapping resonances that only depends on the radius. Then, the differential torque can be defined as follows (Goldreich & Tremaine 1980, 1982; Meyer-Vernet & Sicardy 1987; Ward 1997):

$$\frac{dT}{dR} \equiv T_L^m\left|\frac{dm}{dR}\right| = -\epsilon\Sigma_0 a_\mathrm{s}^3\Omega_\mathrm{s}^2 q^2\left(\frac{a_\mathrm{s}}{R-a_\mathrm{s}}\right)^4 \frac{32}{81}\left[2K_0\left(\frac{2}{3}\right) + K_1\left(\frac{2}{3}\right)\right]^2. \tag{3.69}$$

The left panel of Fig. 3.4 sketches the position and intensity of individual Lindblad resonances, showing how together they contribute to the differential torque derived in Eq. (3.69).

It is interesting to notice how well the result in Eq. (3.9), obtained in the framework of the impulse approximation (Sec. 3.1), matches the result in Eq. (3.69), obtained instead from the detailed study of the disc response to the perturbing potential. Equations (3.9) and (3.69), share the same radial scaling $(R-a_\mathrm{s})^{-4}$, but they differ by a factor $4/3[2K_0(2/3) + K_1(2/3)] \approx 8.5$ and for the presence of Ω_s instead of $\Omega(R)$ in Eq. (3.9).

In this framework, it is easy to understand the torque cut-off for large m values, mentioned in Sec. 3.6 (Artymowicz 1993a): since the effective location of Lindblad resonances is shifted by pressure effects (Eq. 3.52), for large m values the distance of resonances from the satellite becomes $|R_L - a_\mathrm{s}|/a_\mathrm{s} \approx 2H/(3a_\mathrm{s}) \equiv 2\xi/(3m)$ (see Eq. 3.53 for the definition of ξ). This implies that the terms $K_0(2/3)$ and $K_1(2/3)$ in Eq. (3.67) become $K_0(2\xi/3)$ and $K_1(2\xi/3)$; it can be shown that $K_n(u)$ for large u scales as $K_n(u) \propto ue^{-u}$, so that $K_{0,1}(2/3\xi) \propto \xi e^{-2\xi/3}$. This introduces an exponential decrease of the tidal torque for $2\xi/3 > 1$, i.e. $m > (2H/3R)^{-1}$, as shown in Fig. 3.5 (Papaloizou et al. 2007). Furthermore, the shift of resonances causes the torque to vanish in the region $2H/3 < |R - a_\mathrm{s}|$, since resonances accumulate at $R = a_\mathrm{s}(1 \pm 2H/3)$ (see discussion in Sec. 3.5.2 and right panel of Fig. 3.3).

Although the torque density in Eq. (3.69) can be defined only in regions where overlapping resonances are present, sometimes it can be effectively used as a simple

analytic approximation of the torque density throughout the entire disc. In those cases also a cut-off at the location of the innermost/outermost Lindblad resonance (IMLR/OMLR) has to be applied, since no additional resonances can be found beyond those radii.

Due to the previous considerations, the torque density does not vanish in the radial domain where R satisfies

$$\frac{2H}{3} < |R - a_\mathrm{s}| < |R_\mathrm{cut} - a_\mathrm{s}|, \tag{3.70}$$

where R_cut is the IMLR or OMLR location.

Finally, as we did for Eq. (3.11) we can integrate the differential torque in Eq. (3.69) changing its sign in order to obtain the back reaction torque Γ_L the disc exerts on the satellite due to Lindblad resonances . However, the dependence on the disc parameters cannot be easily captured within the analytical framework. For this reason, the total Lindblad torque has been extensively investigated numerically for both 3D and 2D discs[7](Tanaka et al. 2002; D'Angelo & Lubow 2010; Paardekooper et al. 2010). Assuming a surface density profile $\Sigma \propto R^{-p}$, a temperature profile $T \propto R^{-2q}$, and using an adiabatic equation of state, the Lindblad torque obtained from numerical calculations in a 2D disc reads (Paardekooper et al. 2010)

$$\Gamma_{L,\mathrm{s}} = (-2.5 - 3.4q + 0.1p)\Gamma_0\gamma^{-1}, \tag{3.71}$$

where γ is the adiabatic index, and Γ_0 is defined as

$$\Gamma_0 = q^2\Sigma_{a_\mathrm{s}}a_\mathrm{s}^4\Omega_\mathrm{s}^2\left(\frac{H}{R}\right)^{-2}. \tag{3.72}$$

3.7.2 The contribution of corotation resonances

It is not straightforward to define the differential torque produced by the corotation resonances. Corotation resonances for $m = l$ are all located at $R = a_\mathrm{s}$, making in fact impossible to sum the contributions from individual modes and derive a torque density[8], as we did for Lindblad resonances.

However, the overall contribution of circular corotation resonances to the back reaction torque the disc exerts on the satellite has been computed numerically as a function of the disc parameters by Paardekooper et al. (2010). They provided the following scaling of the total corotation torque

$$\Gamma_{\mathrm{c,s}} = \left[0.7\left(\frac{3}{2} - p\right) + 2.2\frac{\xi}{\gamma}\right]\Gamma_0\gamma^{-1}, \tag{3.73}$$

where $\xi = 2q - (\gamma - 1)p$, assuming again a surface density profile $\Sigma \propto R^{-p}$, a temperature profile $T \propto R^{-2q}$, and using an adiabatic equation of state. The variable ξ represents the power of the radial profile of the entropy changed in sign, under the previous assumptions.

[7]The number of dimensions of the simulation affect the result as well as the gravitational softening length required in 2D simulations (Paardekooper et al. 2010).

[8]However it is possible in principle to do it summing the contributions T_c^{lm} with $m \neq l$ (Goldreich & Tremaine 1980), since for them $R_c \neq a_\mathrm{s}$, see Eq. (3.44).

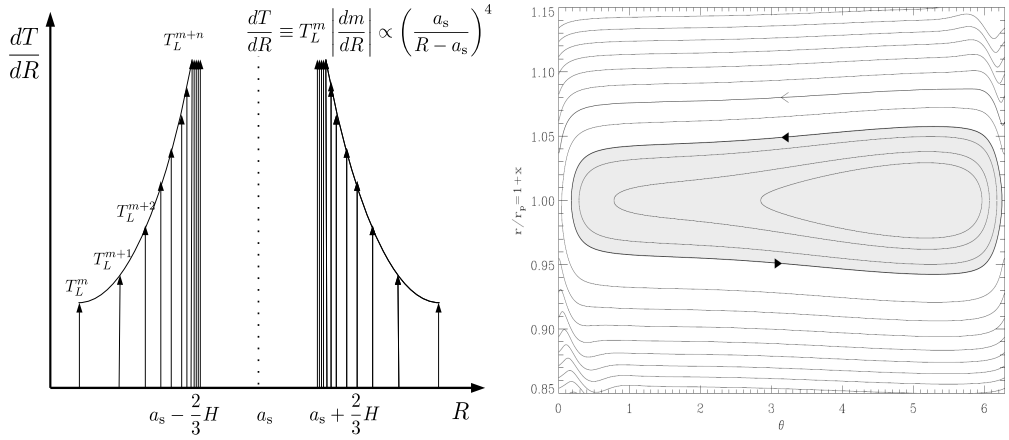


Figure 3.4: Left panel: sketch of the positions and intensity (length of vertical arrows) of the individual Lindblad resonances in the disc, compared with the torque density computed by equation (3.69). Right panel: Horseshoe orbits in presence of the a satellite perturbing potential; the gray shaded area represents the libration region, the white one circulation region. Original figure in Masset (2008).

However, the presence of a satellite in a disc modifies the shape of the gas orbits in the co-orbital region into horseshoe orbits (see right panel in Fig. 3.4, in contrast with the m-eye shaped orbits for individual corotation resonances in Fig. 3.3). The material on these orbits librates in a horseshoe-shaped region (half width $x_{\rm HS} \approx 1.3 a_{\rm s}\sqrt{q(H/R)^{-1}}$, Paardekooper & Papaloizou 2009b) characterized by two U-turns in the immediate vicinity of the satellite. At each U-turn, that occurs every $t_{\rm libr} = 2\pi/|\Omega(R) - \Omega_s|$, the interaction with the satellite moves the material from an inner orbit (w.r.t the satellite location) to an outer one, or vice versa. In order to complete the U-turn, the material needs to exchange angular momentum with the satellite. Analogously to the angular momentum exchange taking place in libration island responsible for the corotation torque (see Sec. 3.6.4), the conservation along the horseshoe streamlines of the fluid vortensity in barotropic fluids, and entropy in an adiabatic ones, causes the exchange of angular momentum to be asymmetric between the two U-turns. This provides a net torque on the satellite called "horseshoe drag" (Ward 1991).

The change in shape of the streamlines from libration islands for barred potential to horseshoes is not predicted within the linear theory we have discussed. This implies in fact that, in order to appropriately describe the torque in the co-orbital region, the co-rotation torque (Sec. 3.6.4) has to be replaced with the horseshoe drag, as soon as horseshoe U-turns appear (i.e. for sufficiently low levels of viscosity Paardekooper & Papaloizou 2009a).

The intensity of the horseshoe drag has been quantified for 2D adiabatic[9] (i.e.

[9]The expression of the horseshoe torque for an isothermal disc has been studied in Casoli & Masset (2009). Discs that are locally isothermal introduce an additional term, called "non-isothermal excess" that depends on the temperature gradient (similarly to the entropy gradient for adiabatic discs). This is due to the fact that vortensity is not anymore conserved along the streamlines since the equation of state is not anymore purely barotropic.

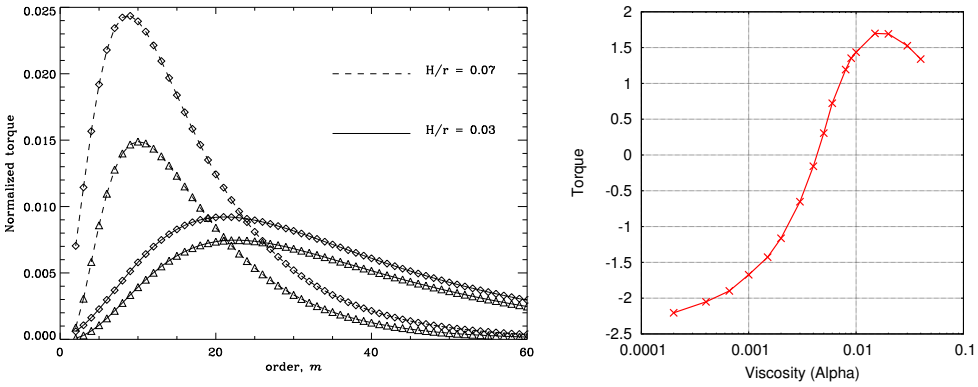


Figure 3.5: Left panel: Normalized intensity of tidal torque exerted by the m-th circular mode of the perturbing potential, showing the cut-off for $m \gtrsim (H/R)^{-1}$. The dashed curve represent $H/R = 0.07$, the solid one $H/R = 0.03$. Triangles represent ILRs, squares OLRs. It is clearly visible that the cut-off occurs for larger m values if H/R is smaller. Original figure in Papaloizou et al. (2007). Right Panel: Plot showing the $\Gamma_{\rm tot} = \Gamma_{L,\rm s} + \Gamma_{\rm HS,s}$ exerted by the disc on the satellite as a function of Shakura & Sunyaev (1973) viscous parameter $\alpha_{\rm ss}$. For low levels of viscosity $\Gamma_{\rm HS,s}$ is saturated and $\Gamma_{\rm tot}$ is dominated by the cumulative effects of Lindblad resonances, i.e. OLRs overcome the effect of ILRs and $\Gamma_{\rm tot} < 0$; for growing viscosity the corotation torque gets desaturated and $\Gamma_{\rm tot} > 0$. Original figure from Kley & Nelson (2012).

radiatively inefficient discs) discs as (Paardekooper et al. 2010)

$$\Gamma_{\rm HS,s} = \left[1.1\left(\frac{3}{2} - p\right) + 7.9\frac{\xi}{\gamma}\right]\Gamma_0\gamma^{-1}. \tag{3.74}$$

However, the horseshoe drag is prone to saturation exactly as the corotation torque (Kley & Nelson 2012). In particular, the level of saturation of the co-orbital torque might even affect the sign of the overall torque exerted by the disc on the satellite, determining whether it will undergo inward or outward migration. The right panel of Fig. 3.5 shows the dependence of the total torque on the turbulent viscous parameter α: high level of viscosity are expected to de-saturate the horseshoe drag, which might even provide the reversal of the total torque exerted on the satellite. In this context, MRI turbulence have been suggested to play a role in desaturating the corotation torque providing a positive torque and outward migration of the satellite (Baruteau et al. 2011). Furthermore, the reader should keep in mind that these results apply for strictly circular orbits of the satellite: even small eccentricities $e \gtrsim 0.03$ might attenuate significantly the value of this torque, leaving active the contribution coming only from the Lindblad resonances (Bitsch & Kley 2010; Fendyke & Nelson 2014).

Chapter

4

Disc-satellite mutual evolution

"...You'd better close your eyes
Ooohhhh bow your head
Wait for the ricochet..."

Deep Purple, Child in Time

As discussed in the previous chapter, the tidal interaction promotes the mutual exchange of energy and angular momentum between the satellite and the gaseous disc.

In this chapter we discuss the role of the satellite in perturbing the density structure of the disc and, conversely, how the disc back-reaction affects the migration and eccentricity evolution of the satellite.

4.1 Gap formation

The ability of a satellite to open a gap in the density distribution depends on two competing effects. On the one hand, since the torque exerted on the disc by the satellite is positive in the outer disc and negative in the inner one (see Eq. 3.69), the satellite pushes away the material from the co-orbital region. On the other hand, the viscous spreading of the disc and pressure effects tend to replenish the depleted region. If the satellite is massive enough, so that the tidal torque is not counterbalanced by viscous or pressure effects, a gap or even a cavity can form in the disc. In Fig. 4.1 we show the results of two numerical simulations where a satellite has carved a gap (left panel) or a cavity (right panel) in the disc.

A number of criteria for gap opening involving pressure, viscosity and the mass of the satellite have been proposed (Lin & Papaloizou 1979, 1993; Bryden et al.

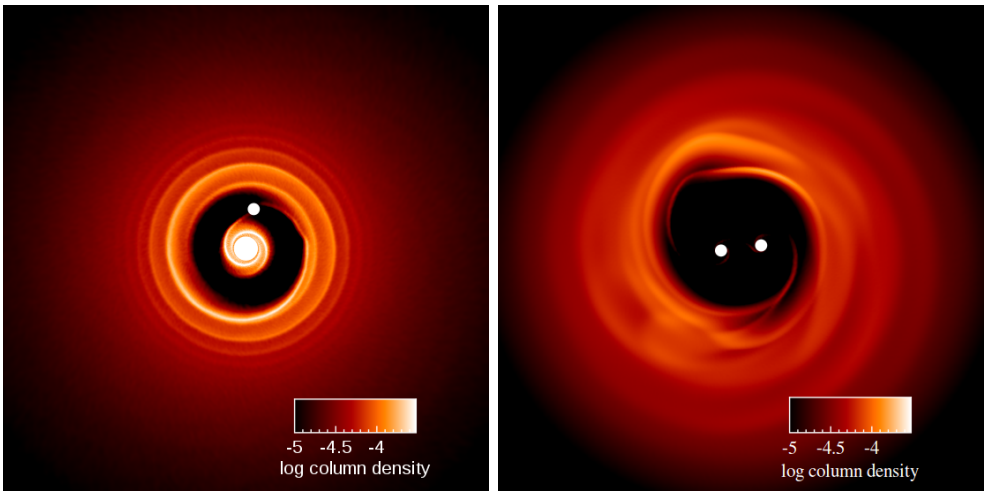


Figure 4.1: Gas density colour plot in the x-y plane for two binary systems with different mass ratios. The left panel shows a gap in the density profile produced by a satellite with a mass ratio $q = 0.01$. The right panel shows a cavity produced by a satellite with mass ratio $q = 1$. The colour-bar is logarithmic in code units. The white dots represent the locations of the primary and secondary objects. Simulations were performed using the 3D SPH code PHANTOM (Price et al. 2018a).

1999; Goodman & Rafikov 2001; Crida et al. 2006; Duffell & MacFadyen 2013; Duffell & Dong 2015), we summarize here the main ones.

In order to find a criterion to open a gap, we can imagine that a gap of width Δ has already been carved and ask under which conditions the gap remains open. To this aim, we compare the tidal torque acting to keep the gap open and the viscous torque closing it.

The tidal torque at the gap edge scales as (see Eq. 3.10, or integrating 3.69 up to the gap edge) (Duffell & MacFadyen 2013):

$$T_{\mathrm{tot},\Delta} \approx q^2 \Omega_{\mathrm{s}}^2 a_{\mathrm{s}}^4 \Sigma_0 \frac{a_{\mathrm{s}}^3}{\Delta^3}. \tag{4.1}$$

The viscous torque, assuming the Shakura & Sunyaev (1973) prescription for viscosity $\nu = \alpha H^2 \Omega$, can be approximated as:

$$T_{\mathrm{visc},\Delta} \approx \nu \Sigma R^2 \Omega(R)|_{a_{\mathrm{s}}} = \alpha H_{\mathrm{s}}^2 \Omega_{\mathrm{s}}^2 \Sigma_0 a_{\mathrm{s}}^2. \tag{4.2}$$

In order to open a gap, one must have $T_{\mathrm{tot},\Delta} \gtrsim T_{\mathrm{visc},\Delta}$. This leads to the following condition on the secondary-to-primary mass ratio q, called "viscous criterion":

$$q^2 \gtrsim q_\nu^2 = \alpha \left(\frac{H}{R}\right)_{a_{\mathrm{s}}}^2 \left(\frac{\Delta}{a_{\mathrm{s}}}\right)^3. \tag{4.3}$$

The gap size Δ cannot be smaller than two threshold characteristic lengthscales: the disc scale height at the satellite location H_{s} and the Hill's radius

$$R_{\mathrm{H}} = \left(\frac{q}{3}\right)^{1/3} a_{\mathrm{s}}, \tag{4.4}$$

that defines a region where the material is gravitationally bound to the satellite. The first threshold, H_s, is set by the fact that pressure gradients are not stable across length scales shorter than the vertical displacement of the disc: a gap width smaller than this value would not be stable and would simply collapse on a sound-crossing timescale. The second, R_H, can be understood by noting that the material in the co-orbital region of the satellite remains trapped within its Hill's sphere, the gap then must be larger than that. However, besides this heuristic explanation, we note that the condition $R_\mathrm{H} > H_\mathrm{s}$ is related to the onset of non-linearity of the perturbations (Korycansky & Papaloizou 1996).

Using $\Delta = \max(H_\mathrm{s}, R_\mathrm{H})$ in Eq. 4.3 provides two "viscous criteria". For small satellite-to-primary mass ratios, the width must satisfy $\Delta > H_\mathrm{s}$. Substituting $\Delta = H_\mathrm{s}$ into Eq. (4.3) the condition for gap opening reads (Armitage 2010; Duffell & MacFadyen 2013):

$$q \gtrsim q_{\nu,1} = \sqrt{\alpha} \left(\frac{H}{R} \right)^{5/2}_{a_\mathrm{s}} , \tag{4.5}$$

where the subscript a_s indicates that the value is computed at the satellite location.

For massive satellites the gap size must satisfy $\Delta > R_\mathrm{H}$ (Lin & Papaloizou 1993). Substituting $\Delta = R_\mathrm{H}$ into Eq. (4.3), we obtain the following condition for gap opening (Lin & Papaloizou 1979, 1993; Bryden et al. 1999)[1]

$$q \gtrsim q_{\nu,2} = \frac{\alpha}{3} \left(\frac{H}{R} \right)^{2}_{a_\mathrm{s}} . \tag{4.6}$$

An additional criterion can be defined starting from the condition $R_\mathrm{H} > H_\mathrm{s}$, from which it follows (Lin & Papaloizou 1993)

$$q \gtrsim q_\mathrm{th} = 3 \left(\frac{H}{R} \right)^{3}_{a_\mathrm{s}} . \tag{4.7}$$

This criterion goes under the name of "thermal criterion" since it sets a condition on the temperature of the disc for the satellite to be able to carve a gap.

It should be noted that this criterion does not depend on viscosity. It has been shown that as the density waves propagate away from Lindblad resonances their amplitude grows until they become non linear and shock at a distance l_th from the excitation region (Goodman & Rafikov 2001). This provides an effective mechanism for the deposition of the angular momentum they carry. When $q > q_\mathrm{th}$, the density waves steepen and shock as soon as they are excited, implying that the deposition of the angular momentum occurs very close to excitation region. Every time the material crosses one of the shocks, it receives a kick of angular momentum that changes its orbit. This mechanism ends up clearing the co-orbital region. For this reason the criterion in Eq. 4.7 is sometimes also referred to as "strong shock" limit.

[1] Note that this criterion is mostly known as $q_{\nu,2} \gtrsim 40\nu_\mathrm{s}/(\Omega_\mathrm{s} a_\mathrm{s}^2)$, where the prefactor 40, comes from different numerical prefactors in Eq. 4.1 and 4.2.

Crida et al. (2006) provided a unified criterion for gap opening involving both the viscous and the thermal criterion, which reads

$$\frac{3}{4}\left(\frac{H}{R}\right)_{a_\mathrm{s}}\left(\frac{q}{3}\right)^{-1/3}+\frac{50\nu_\mathrm{s}}{\Omega_\mathrm{s}a_\mathrm{s}^2}q^{-1}\lesssim 1. \tag{4.8}$$

This criterion represents the condition the parameters have to satisfy so that the satellite carves a gap that reduces the density in the co-orbital region to the 10% of its unperturbed value.

We note that in our derivation of Eq. (4.3) we neglected some numerical coefficients that change the criteria by a numerical factor. The numerical coefficients can be adjusted properly accounting for the constants in the derivation or deducing them from numerical simulations. Typical values of q, satisfying the criterion for typical disc properties in astronomy, range from $q\sim 10^{-4}-10^{-3}$ for protoplanetary discs to $q\sim 10^{-3}-10^{-2}$ for discs surrounding black holes.

Finally, as we will see in the following sections, the back reaction the disc exerts on the satellite might cause it to migrate from its original position. This process is called migration. In order to be able to carve a gap, the satellite migration timescale must be longer than the gap opening one. This provides the so called "inertial" criterion, for fast migrating satellites which reads (Ward & Hourigan 1989)

$$q\gtrsim q_\mathrm{In}=\frac{\Sigma_0 a_\mathrm{s}^2}{M_\star}\left(\frac{H}{R}\right)_{a_\mathrm{s}}^3. \tag{4.9}$$

4.1.1 Gap shape

What is the density profile of the gap? The problem of determining the gap shape, in particular, its width and depth (i.e. what is the minimum density reached into the gap), has been widely investigated in the literature. The typical approach consists in balancing the local expression of the tidal torque exerted by the satellite on the disc with the local viscous torque. The main problem that arises in this procedure is the modeling of the physical mechanism through which waves are damped and, in particular, how much distance they travel before all the angular momentum they transport is transfered to the disc.

The simplest approach consists in assuming that the excitation torque, i.e. the torque exciting the density waves, corresponds to the effective rate of angular momentum transfer from the satellite to the disc (Varnière et al. 2004). This is completely equivalent to assuming that the wave damping is a local process, so that waves transfer all the angular momentum at the same location where they are excited.

Relaxing the local wave damping assumption implies that the excitation torque and the "deposition torque" (also referred to as "pressure torque", Crida et al. 2006) are not the same. In particular, waves are able to travel some distance from their excitation location before being damped. The deposition torque has been computed both numerically (Crida et al. 2006) and analytically (Duffell 2015)[2]

[2]They assumed as deposition mechanism the theory of density waves steepening and shocking, developed by Goodman & Rafikov (2001) and Rafikov (2002).

The full analytical derivation of the gap density profile $\Sigma(R)$ using this approach can be found (Duffell 2015, see their Eq. 12).

More generally, we state here two effective formulae to estimate the gap depth and width. The depth of the gap, expressed as the minimum of the density profile $\Sigma_{\rm min}$, can be effectively estimated as (Kanagawa et al. 2018)

$$\Sigma_{\rm min} = \frac{1}{1+0.04K}\Sigma_0, \tag{4.10}$$

where Σ_0 is the unperturbed surface density, and K is

$$K = q^2\alpha^{-1}\left(\frac{H}{R}\right)^{-5}_{a_{\rm s}}. \tag{4.11}$$

It should be noted that the gap opening criterion in Eq. (4.5) can be restated as $K \gtrsim 1$. A width estimate instead can be obtained from Eq. (4.3) as (Lin & Papaloizou 1979)

$$\Delta \approx \left[\left(\frac{M_{\rm s}}{M_\star}\right)^2 \frac{\Omega_{\rm s} a_{\rm s}^2}{\nu_{\rm s}}\right]^{1/3} a_{\rm s}. \tag{4.12}$$

4.2 Energy and angular momentum exchange during disc-satellite interaction

We consider a satellite with mass $M_{\rm s}$ moving on an elliptic orbit around the mass $M_\star$ that is located in one of the foci[3].

Elliptical orbits can be characterized using two parameters: the semi-major axis of the orbit $a_{\rm s}$ and its eccentricity e. These quantities depend exclusively on the energy $E_{\rm s}$ and the angular momentum $L_{\rm s}$ of the satellite, that are conserved along the entire orbit. The orbital energy of the system is

$$E_{\rm s} = -\frac{GM_\star M_{\rm s}}{2a_{\rm s}} \tag{4.13}$$

the angular momentum is then given by

$$L_{\rm s} = M_{\rm s}\sqrt{GM_\star a_{\rm s}(1-e^2)}. \tag{4.14}$$

To each elliptical orbit characterized by couple $(L_{\rm s}, E_{\rm s})$ corresponds only one couple $(e, a_{\rm s})$. As a consequence, the evolution of the orbital properties of both the satellite and the disc depends on the rate at which the two exchange energy and angular momentum.

In this context, it is very important to understand how the torque arising from different individual (l, m) modes of the potential decomposition (Eq. 3.36) affects the evolution of the semi-major axis and eccentricity of the satellite.

[3]Setting a fixed central mass $M_\star$ in the origin of an inertial reference frame is reliable as long as the mass of the satellite $M_{\rm s} \ll M_\star$, see Sec. 1.1).

By definition, the flux of angular momentum injected in a wavelike disturbance with pattern frequency[4] $\Omega_{\rm p}^{lm}$ gives:

$$\dot{L}_{\rm s} = -T^{lm}, \tag{4.15}$$

where $\dot{L}_{\rm s}$ represent time derivative of the angular momentum of the satellite and T^{lm} is the torque injected in the disc by the satellite. The flux of energy $F_E^{lm} = -\dot{E}_{\rm s}$ transported by the perturbations can be obtained by imposing the conservation of the Jacobi's constant $\dot{J}^{lm} = F_E^{lm} - \Omega_{\rm p}^{lm} T^{lm} = 0$ (since each bar-potential is constant in the frame rotating with the pattern frequency $\Omega_{\rm p}^{lm}$, Goldreich & Sari 2003), it reads[5]:

$$\dot{E}_{\rm s} = -F_E^{lm} = -\Omega_{\rm p}^{lm} T^{lm}. \tag{4.16}$$

4.3 Eccentricity evolution

We assume that the satellite, that orbits the central object with a frequency $\Omega_{\rm s}$, interacts with the disc at a specific resonance, exciting only one perturbation with pattern frequency $\Omega_{\rm p}$ and exerting a torque T^{lm} at that location. We restate $\dot{L}_{\rm s}$ in Eq. (4.15) as function of $\dot{a}_{\rm s}$ and $\dot{e}$

$$\dot{L}_{\rm s} = \frac{M_{\rm s}}{2}\sqrt{\frac{GM_\star(1-e^2)}{a_{\rm s}}}\dot{a}_{\rm s} - \frac{\Omega_{\rm s} M_{\rm s} e a_{\rm s}^2}{\sqrt{1-e^2}}\dot{e} = -T^{lm}. \tag{4.17}$$

The rate of change of the satellite semi-major axis (migration rate) and the energy are related by the following equation

$$\dot{E}_{\rm s} = \frac{GM_\star M_{\rm s}}{2a_{\rm s}^2}\dot{a}_{\rm s} = -\Omega_{\rm p}^{lm} T^{lm}. \tag{4.18}$$

which also provides

$$\dot{a}_{\rm s} = -2\frac{\Omega_{\rm p}^{lm}}{\Omega_{\rm s}}\frac{T^{lm}}{M_{\rm s}\Omega_{\rm s} a_{\rm s}^2} a_{\rm s} \tag{4.19}$$

Substituting Eq. (4.18) in (4.17), after some simple algebra we get (Masset 2008):

$$\frac{M_{\rm s}}{2}\frac{\Omega_{\rm s} a_{\rm s}^2}{\sqrt{1-e^2}}\frac{de^2}{dt} = \left(1 - \frac{\Omega_{\rm p}^{lm}}{\Omega_{\rm s}}\sqrt{1-e^2}\right) T^{lm}. \tag{4.20}$$

The sign of $\dot{e}$ and $\dot{a}_{\rm s}$ thus depends on the sign of the angular momentum injected into the perturbation (T^{lm}) and the ratio between the pattern frequencies $\Omega_{\rm p}$ of the density perturbations and the orbital frequency of the satellite $\Omega_{\rm s}$. Substituting the values for individual resonances, it is possible to collect the following information about the effect of individual resonances on the evolution of $a_{\rm s}$ and e (Goldreich & Sari 2003; Masset 2008), a summary of the following considerations can be found in Tab. 4.1:

[4] We remind that the expression of $\Omega_{\rm p}^{lm}$ as a function of the satellite orbital frequency, can be found in Eq. (3.37).

[5] We remark that T^{lm} is defined as the torque the satellite exerts on the disc. This sets the sign in the r.h.s. in Eq. (4.16): a positive torque will cause the energy of the satellite to decrease.

potential		pattern speed	Torque – T^{lm}		Keplerian position	effects	
		$\Omega_{l,m}$	name	sgn		a_s	e
$\psi^{m,m}$	Circular	Ω_p	$T^{m,m}_\mathrm{OLR}$	$+$	$a_\mathrm{s}(1+2/3m)$	$\Downarrow$	$\uparrow$
			$T^{m,m}_\mathrm{CR}$	?	a_s	?	?
			$T^{m,m}_\mathrm{ILR}$	$-$	$a_\mathrm{s}(1-2/3m)$	$\Uparrow$	$\downarrow$
$\psi^{m-1,m}$	Slow First Order	$\Omega_p - \kappa_p/m$	$T^{m-1,m}_\mathrm{OLR}$	$+$	$a_\mathrm{s}(1+4/3m)$	$\downarrow$	$\Uparrow$
			$T^{m-1,m}_\mathrm{CR}$	$(-)$	$a_\mathrm{s}(1+2/3m)$	$(\uparrow)$	$(\Downarrow)$
			$T^{m-1,m}_\mathrm{ILR}$	$-$	a_s	$\uparrow$	$\downarrow$
$\psi^{m+1,m}$	Fast First Order	$\Omega_p + \kappa_p/m$	$T^{m+1,m}_\mathrm{OLR}$	$+$	a_s	$\downarrow$	$\downarrow$
			$T^{m+1,m}_\mathrm{CR}$	$(+)$	$a_\mathrm{s}(1-2/3m)$	$(\downarrow)$	$(\Downarrow)$
			$T^{m+1,m}_\mathrm{ILR}$	$-$	$a_\mathrm{s}(1-4/3m)$	$\uparrow$	$\Uparrow$

Table 4.1: Resonances for a set of (l, m) pairs. For each one, the pattern frequency, sign and location of Lindblad and corotation torques T^{lm} that excites the perturbation are reported. The last two columns represent the effect of the perturbation on the evolution of a_s and e. Double arrows represent the leading effect under the assumption of a well carved gap. Parentheses indicate that the sign of the torque is computed assuming that the vortensity decreases approaching the planet, situation that typically occurs when a deep gap is carved. (Adapted from the original table in Goldreich & Sari 2003).

1. Circular resonances: the effect of circular resonances ($\Omega_\mathrm{p} = \Omega_\mathrm{s}$, i.e. $m = l$) on the evolution of e and a_s can be obtained expanding Eq. (4.20) for small values of e as follows:
$$\frac{\dot{e}}{e} = \frac{T^{mm}}{2M_\mathrm{s}\Omega_\mathrm{s}a_\mathrm{s}^2} = -\frac{1}{4}\frac{\dot{a}}{a}. \tag{4.21}$$
This equation depends exclusively on the sign of T^{mm} that is the torque exerted satellite on the disc, it is thus straightforward to determine the signs of $\dot{e}$ and $\dot{a}_\mathrm{s}$.

   1.1 Circular Lindblad resonances: OLRs, that are characterized by $T^{mm}_\mathrm{OLR} > 0$, are expected to excite eccentricity ($\dot{e} > 0$) and to provide an inward migration of the satellite ($\dot{a}_\mathrm{s} < 0$); in contrast, ILRs, characterized by $T^{mm}_\mathrm{ILR} < 0$, damp the eccentricity down and provide an outward migration of the satellite.

   1.2 Circular corotation resonances: The sign of $\dot{a}_\mathrm{s}$ and $\dot{e}$ due to the action of circular corotation resonances (co-orbital resonances) cannot be determined a priori since T^{mm}_CR depends on the vortensity gradient in the libration region, as discussed in Sec. 3.6.4; in particular, eccentricity grows only if the density profile $\Sigma \propto R^{-p}$ has $p > 3/2$, it will decrease otherwise.

2. First order resonances: since the intensity of the (m, l) mode of the potential scales as $e^{2|l-m|}$, for eccentric satellites, the strongest contribution in the Fourier expansion in Eq. (3.36) after circular $m = l$ resonances is given by those terms with $l = m \pm 1$. For these resonances the pattern frequency Ω^{lm}_p

(see Eq. 3.37) reads

$$\Omega_{\rm p}^{m\pm1,m} = \left(1 \pm \frac{1}{m}\right)\Omega_{\rm s}. \tag{4.22}$$

Those pattern frequencies characterized by the + sign have their corotation radius located inside the co-orbital radius and are called "fast first order resonances"; in contrast, those with the − sign have their corotation radius located outside the co-orbital radius and are called "slow first order resonances".

2.1 First order Lindblad resonances: for these pattern frequencies, slow ILRs and fast OLRs are located where the condition $\Omega(R) = \Omega_{\rm s}$ is met, i.e. at $R = a_{\rm s}$. For this reason, they are called "co-orbital Lindblad resonances". Using Eq. (4.20), it can be shown that co-orbital Lindblad resonances are expected to damp the eccentricity down. In contrast, following the same approach, it can be shown that fast ILRs and slow OLRs are expected to excite the satellite eccentricity. The sign of $\dot{a}$ and $\dot{e}$ for each first order Lindblad resonance can be found in Tab. 4.1.

2.2 First order corotation resonances: As for circular corotation resonances, the effect of first order corotation resonances on the evolution of e and $a_{\rm s}$ depends on the vortensity gradient, and cannot be determined a priori. Anyway, if the satellite succeed in opening a gap in the disc, a very steep grandient in the vortensity is produced at its edges. The sign of this vortensity gradient is consistent with a steep decrease in density radial profile in the direction approaching the satellite. We can thus safely assume, that the sign of $T_{\rm CR}^{m+1,m} > 0$ for fast first order resonances and that $T_{\rm CR}^{m-1,m} < 0$ for slow ones. The sign of $\dot{a}_{\rm s}$ and $\dot{e}$ for each first order corotation resonance can be found in Tab. 4.1

From these considerations it becomes clear that the overall evolution of the system depends on how the disc and the satellite mutually affect each other: on the one hand, the orbital properties of the satellite determine where are resonant locations and their intrinsic intensity, carving the disc structure; on the other, the the disc structure shapes the relative intensity of resonances that determine the evolution of the orbital properties of the satellite. This mutual evolution of disc and satellite makes hard to predict general results within the framework of the linear theory presented here.

To conclude this section, in Fig. 4.2 we show the outcome of two long timescales simulations (simulations from (Ragusa et al. 2018), see Chap. 7). This figure shows two cases of inward migration in the $(L_{\rm s}, a_{\rm s})$ plane; in one the eccentricity is growing (left panel), in the other the eccentricity is descreasing (right panel). With these plots we want to highlight how different the evolution of the satellite orbital parameters can be when the disc properties change. Furthermore, these plots show clearly some periodic oscillations that cannot be captured by the resonant torque analysis we provided so far. Such oscillations are likely related to the modes $m = 0$ of the potential (see Sec. 7.4.1), which are expected to drive the secular disc-satellite interaction. We refer to Chap. 7 for a deeper discussion about these themes and about the disc-satellite mutual evolution at long timescales.

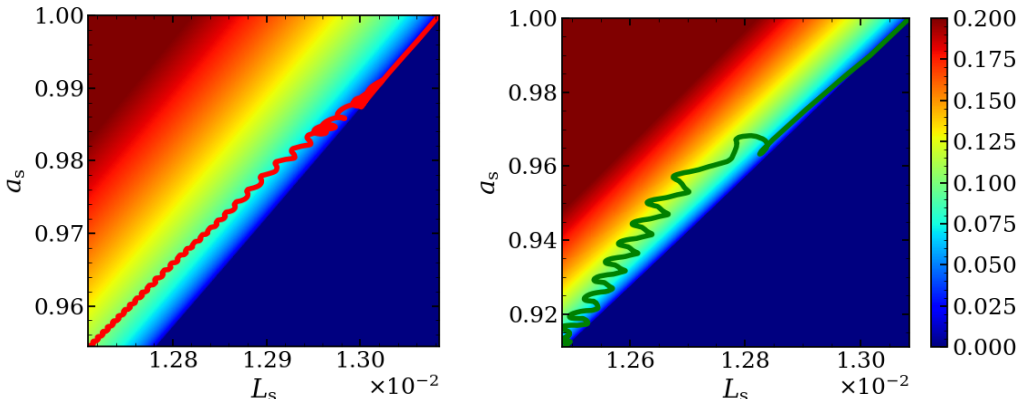


Figure 4.2: Outcome of two numerical simulations of migrating satellites interacting with their accretion discs for two different disc masses (1/5 of the satellite mass, left panel, and 3/5 of the satellite mass). The solid lines represent the path of the two satellites in the angular momentum vs. semi-major axis (L_s, a_s) plane as the simulations evolve. The colour-plot shows the corresponding satellite eccentricity. The left panel shows a satellite whose eccentricity is progressively growing as it migrates inward. The right panel shows a satellite whose eccentricity grows very fast in the initial stages and then decreases at later times. Some periodic oscillations of the eccentricity are clearly visible in both plots. Such oscillations will be discussed in more detail in Chap. 7.

4.4 Type I migration

Satellites that are not able to carve a gap in the disc structure are subject to the so called Type I migration. In this migration regime, the satellite is completely embedded in the gaseous disc. The torque exerted by the disc on the satellite is given by the sum of the torques exerted at each resonant location. It can be shown that the eccentricity damping action of Lindblad co-orbital resonces is stronger than that provided by any of the other resonances (Ward 1988; Artymowicz 1993b; Tanaka & Ward 2004). Furthermore, the eccentricity damping time is much shorter than the migration time (Artymowicz 1993b). These facts imply that it is reasonable to assume that type I migration occurs on circular orbits. Under this assumption the satellite and the disc exchange angular momentum only at circular resonant locations. The total torque exerted by the disc on the sattelite is thus given by

$$\Gamma_{\rm tot} = -\left(\sum_{m=2}^{\infty} T^{mm}_{L,\rm ILR} + \sum_{m=1}^{\infty} T^{mm}_{L,\rm OLR} + \sum_{m=1}^{\infty} T^{mm}_{\rm CR}\right). \tag{4.23}$$

The torque exerted on the satellite by OLRs and ILRs, has opposite sign. In particular: ILRs resonances exert a positive torque on the satellite increasing its angular momentum and causing an outward migration; on the other hand, OLRs exert a negative torque on the satellite causing an inward migration. In general, the contribution of the OLRs is stronger than that provided by ILRs, overall causing the satellite to migrate inward. However, as discussed in Sec. 3.7.2, the

sign of CR torque and its level of saturation play a fundamental role in determining the direction of satellite migration.

Assuming that the total contribution for the corotation torque $T^m_{\rm CR}$ is given by the horseshoe drag, a numerical estimate of Eq. (4.23) is given by Paardekooper et al. (2010)[6]

$$\Gamma^{\rm adia}_{\rm tot,2D} = \left[-2.5 - 3.4q + 0.1p + 1.1\left(\frac{3}{2} - p\right) + 7.9\frac{\xi}{\gamma}\right]\Gamma_0\gamma^{-1}, \tag{4.24}$$

where, as in Sec. 3.7.1 and 3.7.1 $\xi = 2q - (\gamma - 1)p$, p is the power of the surface density profile ($\Sigma \propto R^{-p}$), q relates with the temperature profile as $T \propto R^{-2q}$, and using an adiabatic equation of state. The variable ξ represents the power of the radial profile of the entropy changed in sign. Γ_0 (see Eq. 3.72) sets the order of magnitude of the total torque.

In recent years, predictions about the migration rate have been widely discussed in the context of planet formation. For a reasonable choice of the disc parameters, in particular for parameters of the MMSN (mininum mass solar nebula, Weidenschilling 1977,Hayashi 1981) the migration of the satellite occurs inward (Paardekooper et al. 2010).

Knowing that for a satellite on a circular orbit $L_{\rm s} = M_{\rm s}\Omega_{\rm s}a_{\rm s}^2$, and assuming $\dot{L}_{\rm s} = \Gamma_0$ we can make a qualitative estimate of the migration rate $\dot{a}_{\rm s}$ that takes the form:

$$\dot{a}_{\rm s}^{\rm typeI} = \frac{2\Gamma_0}{L_{\rm s}}a_{\rm s}. \tag{4.25}$$

Computing the typical timescale for type I migration, we get the following result

$$\tau_{\rm typeI} \equiv \frac{a_{\rm s}}{|\dot{a}_{\rm s}^{\rm typeI}|} \approx \frac{L_{\rm s}}{|\Gamma_0|} = q^{-1}\frac{M_\star}{\Sigma_{a_{\rm s}}a_{\rm s}^2}\left(\frac{H}{R}\right)^2\Omega_{\rm s}^{-1}. \tag{4.26}$$

Substituting in the previous equation the values predicted for the MMSN $\Sigma_{a_{\rm s}} = 1700 \times (a_{\rm s}/1{\rm au})^{-3/2}\ {\rm g\,cm^{-2}}$, and assuming the disc thickness $H/R = 0.05$, an earth-like mass planet (i.e. $M_{\rm s} = M_\oplus \to q \approx 3 \times 10^{-6}$) located at $a_{\rm s} = 1\,{\rm au}$, one obtains $\tau_{\rm typeI} < 1\,{\rm Myr}$, i.e. lower than the typical lifetime of accretion discs surrounding a forming star, which is expected to be $\approx 1-10\,{\rm Myr}$.

Based on this result, a large variety of studies have been conducted in order to find possible further physical mechanisms able to slow down the migration, so that planets can form and survive across the entire lifetime of the disc.

Magnetic effects (as mentioned in Sec. 3.6.1) might provide a reversal of the migration rate (Terquem 2003; Fromang et al. 2005). Migration have been observed to reverse for moderate values of eccentricity of the planet $e > H/R$ Papaloizou (2002).

Abandoning the assumption of isothermal discs, the dependence of the horseshoe drag on a non vanishing entropy or temperature gradient (see Sec. 3.6.4 and 3.7.2) provides in fact a possible mechanism to slow down, or even reverse the migration: as previously observed in the simulations by Paardekooper & Mellema (2006) when

[6]The same result has been discussed for 3D locally isothermal discs by D'Angelo & Lubow (2010), see their Eq. (9).

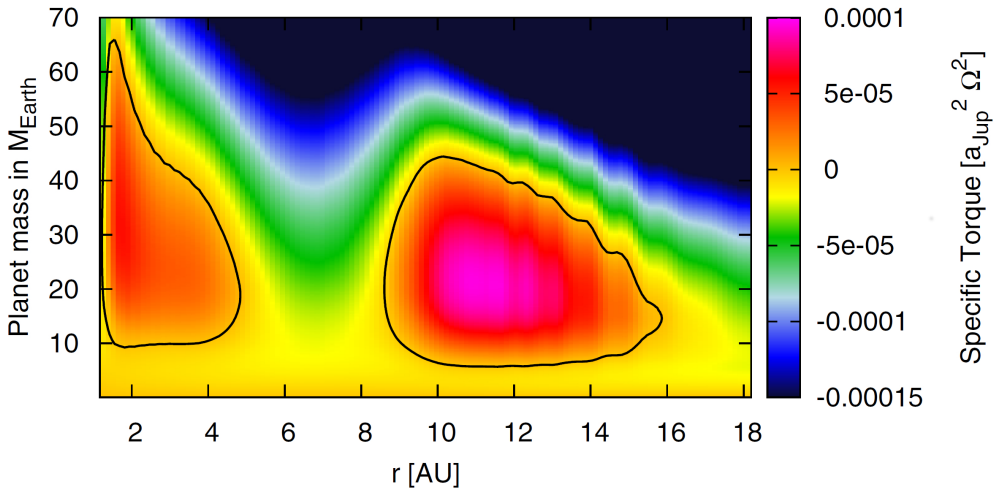


Figure 4.3: Colour plot of the torque values in the (M_s, a_s) parameter space for a disc with radiative cooling, stellar irradiation and viscous heating. Solid lines enclose the parameter regions where outward migration appears to occur. The satellite mass plays a role mainly in setting the degree of saturation of the corotation torque. Original figure Bitsch et al. (2013a).

considering non isothermal disc, (see also Baruteau & Masset 2008, Paardekooper & Papaloizou 2008, Casoli & Masset 2009, Paardekooper et al. 2011).

Bitsch et al. (2013a) performed a set of numerical simulations with a realistic modelling of the disc thermodynamics including viscous heating, radiative cooling and stellar irradiation, in order to define the regions in the (M_s, a_s) parameter space in which inward and outward migration take place. Figure 4.3 shows the torque values they find in the (M_s, a_s) parameter space.

Radiative effects or the heating due to the accretion of material on to the planet might stop the inward migration of small mass planets (Lega et al. 2014; Benítez-Llambay et al. 2015).

Finally 3D simulations performed by Fung et al. (2015) showed that the modification of the morphology of streamlines in the satellite surroundings gives rise to an additional torque that slows down the inward migration rate.

4.5 Type II migration

When the satellite satisfies the criteria presented in Sec. 4.1, it carves a stationary gap in the disc. The gas depletion in the co-orbital region makes the contribution of co-orbital resonances negligible in the computation of the total torque. This fact, in addition to the reduction of the number of Lindblad resonances involved in the disc-satellite interaction, results in a slowing down of the satellite migration rate.

In order to qualitatively understand this migration regime, we can assume that the satellite is locked inside the gap and that the torque it exerts on the disc

bridges the two sides of the disc. For this reason, the satellite location constitutes a sort of boundary condition for the equations ruling the disc evolution. Starting from Eq. (2.20), in order to conserve the angular momentum, the satellite has to satisfy the following equation (Ivanov et al. 1999, Eq. 42):

$$\frac{1}{2}M_{\rm s}\Omega_{\rm s}a_{\rm s}\dot{a}_{\rm s} + 2\pi a_{\rm s}^2\left[\frac{3}{2}\nu\Sigma_0\Omega_s + \Sigma_0\Omega_{\rm s}a_{\rm s}\dot{a}_{\rm s}\right] = 0, \tag{4.27}$$

where Σ_0 is the unperturbed surface density in the gap; all the quantities showing an "s" subscript are computed at the satellite location. The assumption for the gas drift velocity $v_R \approx \dot{a}_{\rm s}$ has also been used, implying that in this model the satellite constitutes an insurmountable barrier for the gas. Equation (4.27) represents in fact the balance between the angular momentum loss rate of the satellite and the angular momentum the satellite need to constantly inject into the disc in order to maintain the cavity depleted.

Rearranging equation (4.27) we obtain the following migration rate:

$$\dot{a}_{\rm s}^{\rm typeII} = \frac{4\pi a_{\rm s}^2\Sigma}{M_{\rm s}+4\pi a_{\rm s}^2\Sigma_0}\frac{3}{2}\frac{\nu}{a_{\rm s}^2}a_{\rm s} = \frac{M_{\rm d}}{M_{\rm s}+M_{\rm d}}\frac{a_{\rm s}}{t_\nu(a_{\rm s})}. \tag{4.28}$$

where $M_{\rm d} = 4\pi\Sigma_0 a_{\rm s}^2$ qualitatively quantifies the mass of the inner disc, and $t_\nu(a_{\rm s})$ is the viscous timescale (Eq. 2.28)

$$t_{\rm typeII} = \frac{a_{\rm s}}{|\dot{a}_{\rm s}|} = \frac{M_{\rm s}+4\pi a_{\rm s}^2\Sigma_0}{4\pi a_{\rm s}^2\Sigma_0}\frac{2}{3}\frac{a_{\rm s}^2}{\nu} = \frac{M_{\rm s}+M_{\rm d}}{M_{\rm d}}t_\nu(a_{\rm s}), \tag{4.29}$$

From Eq.s (4.28) and (4.29) it becomes clear that in this migration regime two main cases can be identified. First, the satellite mass satisfies $M_{\rm s} \lesssim M_{\rm d}$: the inertia of the satellite is smaller or at most comparable to that of the disc. The satellite result thus to be locked into the gap it has carved and the migration rate $\dot{a}_{\rm s,typeII}$ occurs at the nominal radial gas drift velocity of the gas due to the disc viscous evolution (Lin & Papaloizou 1986):

$$\dot{a}_{\rm s}^{\rm typeII} = -\frac{a_{\rm s}}{t_\nu(a_{\rm s})}, \tag{4.30}$$

so that the typical timescale type II migration for $M_{\rm s} \ll M_{\rm d}$ is:

$$t_{\rm typeII} = \frac{a_{\rm s}}{|\dot{a}_{\rm s}^{\rm typeII}|} = t_\nu(a_{\rm s}). \tag{4.31}$$

Armitage (2010) proposes a simple analogy in order to understand this case of type II migration: the satellite is surrounded by "brick walls" at the inner and outer edges of the cavity; at the inner edge, the tidal torques remove the exact amount of angular momentum needed to prevent the gas to cross the wall due to viscous diffusion, the same mechanism is reproduced at the outer edge where instead the angular momentum is removed from the satellite and transfered to the gas.

Second, the satellite mass satisfies $M_{\rm s} \gg M_{\rm d}$: in this case the inertia of the satellite is much larger than that of the disc and the migration rate is much

slower; this implies that the timescale for this case of type II migration becomes (Ivanov et al. 1999; Armitage 2007):

$$\tau_{\rm typeII} = \frac{M_{\rm s}}{M_{\rm d}}\tau_{\nu}. \tag{4.32}$$

Some recent numerical simulations have questioned the validity of the timescales in Eq. (4.29) and (4.31). Edgar (2007), Duffell et al. (2014) and Dürmann & Kley (2015) have shown that for small satellites ($M_{\rm s} < M_{\rm d}$) carving gaps might undergo migration on timescales shorter than the viscous time. This result is likely due to the fact that some material crosses the gap in 2D and 3D simulations, in contrast with the assumption under which the analytical model was developed (i.e. that the inward velocity of the gas v_R is equal to the velocity of the satellite $\dot{a}_{\rm s} = v_R$). Anyway, in the massive satellite regime ($M_{\rm s} > M_{\rm d}$) 2D simulations in Ragusa et al. (2018) have shown good agreement with the theoretical model.

Recently Kanagawa et al. (2018) showed that the migration rate in type II regime is inversely proportional to the minimum density value in the gap $\Sigma_{\rm min}$ as soon as the transition from type I migration to type II occurs. In this paper the authors found

$$\dot{a}_{\rm s}^{\rm typeII} = -150\frac{\Gamma_0}{\Omega_{\rm s} a_{\rm s} M_{\rm s}}\frac{1}{K} \tag{4.33}$$

where K have been previously defined in Eq. 4.11 and can be used to quantify the gap depth. This expression appears to be consistent with the migration timescales obtained by Duffell et al. (2014) and Dürmann & Kley (2015) for small satellites, but also with Eq. (4.29) for more massive satellites.

Finally, it is worth to mention that, performing a large number of numerical simulations, Kanagawa et al. (2018) were able to infer a single general expression of the migration timescale that can be used both for type I and type II migrating satellites as an only function of K (see their Eq. 29). We report in Fig. 4.4 the summarizing plot from Kanagawa et al. (2018), where the migration timescales from numerical simulations are compared with their theoretical predictions.

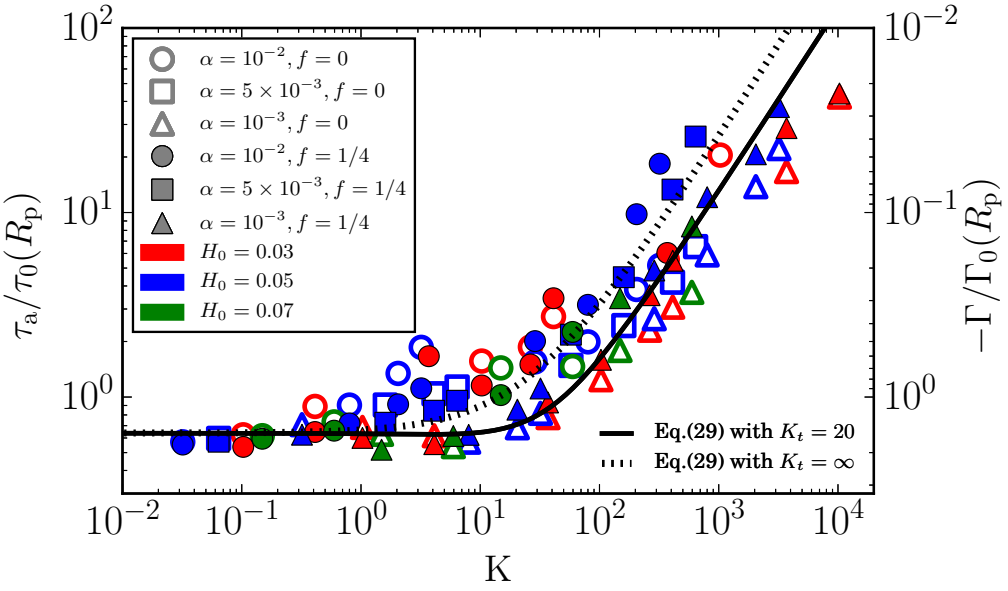


Figure 4.4: Plot of the general expression valid for both type I and type II migration timescales by Kanagawa et al. (2018) as a function of K (see Eq. 4.11); Γ_0 is given in Eq. (3.72) and $\tau_0 = (M_s \Omega_s a_s)/(2\Gamma_0)$. Solid and dashed curve represent two possible ways to prescribe the corotation torque cut-off due to the gap depletion. The points in the plot represent the outcome of their simulations; different colours and shapes represent different parameters in the simulations, in particular: α is the Shakura & Sunyaev (1973) viscous parameter, f is the flaring index of the disc ($H(R)/R = H_0(R/R_0)^f$) and H_0 is the disc aspect ratio.

Part II

Disc-satellite interaction in black hole binary systems

Chapter

5

Black hole binaries

"...C'è chi si mette degli occhiali da sole
Per avere più carisma e sintomatico mistero..."

Franco Battiato, Bandiera Bianca

In this Chapter we review the formation mechanisms of both stellar and supermassive black hole binaries (Sec. 5.1 and 5.2).

We introduce some other relevant concepts in the context of black hole binary mergers. In particular, in Sec. 5.3 we discuss how the emission of gravitational waves modifies the semi-major axis and eccentricity of the binary, while in Sec. 5.4 we present the concept of "kick velocity" of the merger remnant.

In Sec. 5.5, we will also emphasise the relevance of the accretion rate on the binary for determining the electromagnetic and gravitational wave emission properties, and how the tidal torque the binary exerts on the disc can reduce the accretion rate in these systems.

5.1 Stellar mass black hole binaries

As discussed in Sec. 1.3.2, stellar mass black holes form as endpoints of the life of massive stars. In this context, stellar mass black hole binaries have been proposed to form following two main scenarios. On the one hand, they might have formed as massive binary stars (Postnov & Yungelson 2014), becoming a black hole binary at the end of their lives; on the other hand, they formed by gravitational capture in dense stellar clusters (Benacquista & Downing 2013).

In the first case, often referred to as the "field formation" scenario, main sequence isolated binary stars are the starting point. One of the two stars becomes a black hole following a supernova event or, if the star is sufficiently massive, direct

collapse into a black hole. Then, as a consequence of Roche lobe overflow of the second star, a phase of common envelope takes place that shrinks the separation of the binary. This stage is critical in the framework of the merger of stellar mass black hole binaries: the emission of gravitational waves is not sufficient to provide a merger timescale shorter than the Hubble time if the black hole binary is not close enough. Finally, the second object becomes a black hole and the emission of gravitational waves reduces the separation until merger occurs (Belczynski et al. 2016). It must be mentioned that in this context the natal kick of the black holes, i.e. the initial velocity of the black hole following an asymmetric supernova event, is extremely relevant to determine the survival of the binary.

In the second case, often referred to as the "dynamical formation" scenario, the black holes originate from a dense stellar cluster. Most massive objects sink toward the center of the cluster due to dynamical friction with the other stars. In this dense environment, the black holes that are present in the field undergo a large set of possible three body interactions leading to the formation of binary black holes[1] (Rodriguez et al. 2016).

However, it must be said that in dense environments black holes might undergo second generation mergers with the remnants of first generation ones (Gerosa & Berti 2017), further complicating the formation scenario of stellar black hole binaries.

Binary black holes are currently receiving great attention from the scientific community due to the detection of gravitational waves signals from compact objects mergers (Abbott et al. 2016a,b,d, 2017a,d,b,c) after advanced LIGO and Virgo became operative. Besides the large number of tests for general relativity these events provide, they can also be exploited as an observational tool to probe the origin and evolution of these systems (Gerosa 2018; Belczynski et al. 2017), since at the moment of the merger their orbital properties carry the memory of their past evolution. Currently, out of a total of 7 gravitational wave merger events, one was a NS-NS merger (GW170817, Abbott et al. 2017c), one LVT151012 had 13% probability to be noise (Abbott et al. 2016a), 5 were confirmed stellar black hole binary mergers.

5.1.1 Determining the formation mechanism from gravitational wave detections

Constraints about the formation mechanism of the binary can be inferred from the gravitational wave detections. The mass distribution of merging black holes can provide some insights about the evolutionary path of the black holes, nevertheless a large number of detections is required in order to put some constraints using mass measurements only.

Three of the events involved black holes with masses relatively large with respect to those observed in X-ray binaries. See Fig. 5.1 for a graphic comparison between the masses of the gravitational wave events and the those of X-ray binaries. The large masses of these merging black holes ($M_\bullet \gtrsim 25\mathrm{M}_\odot$) suggest that their progenitors were low metallicity stars, thus characterized by weak winds (Heger

[1] See the video `https://www.youtube.com/watch?v=chD3_G9sXy0` by Aaron Geller for an example of the dynamics of a three body encounter in a dense stellar cluster.

et al. 2003; Kowalska-Leszczynska et al. 2015; Giacobbo et al. 2018). Such progenitors could be well represented by those formed in low mass galaxies (Lamberts et al. 2018) undergoing a prompt merger, or by population III stars with strongly delayed mergers (Kinugawa et al. 2014; Abbott et al. 2016e).

On the other hand, as discussed in Sec. 1.3.2, observing black hole progenitors with masses $52\,\mathrm{M}_\odot \lesssim M_\bullet \lesssim 133\,\mathrm{M}_\odot$ would suggest that the origin of the black hole binary must be dynamical, involving necessarily second generation black holes, due to the pair-instability in supernova events that completely evaporate the progenitor without leaving any remnant. However, a statistical based analysis on detection rates performed by Fishbach & Holz (2017) claimed that the lack of detections of stellar black holes in the pair-instability mass regime among the events we already have implies that such black holes likely do not exist, i.e. confirming the gap in black hole mass distribution.

Relative spin orientation is expected to give a cleaner indication of the formation scenario (Farr et al. 2017; Gerosa 2018; Belczynski et al. 2017). One of the spin-related quantities to which the gravitational waveforms are most sensitive is the effective spin[2], which reads

$$\chi_{\mathrm{eff}} = \frac{c}{G(M_1+M_2)}\left(\frac{\boldsymbol{S}_1}{M_1}+\frac{\boldsymbol{S}_2}{M_2}\right)\cdot\frac{\boldsymbol{L}}{|L|} \tag{5.1}$$

where $\boldsymbol{S}_{1,2}$ are the spins of the individual black holes (Eq. 1.26) and $\boldsymbol{L}/|L|$ is the unit vector pointing in the direction of the binary orbital angular momentum. This scalar quantity is a mass-weighted average of the aligned component of the individual spins.

The effective spin is a constant of motion in post-Newtonian dynamics and can be inferred from the gravitational wave pattern. This quantity provides a direct measure of the alignment of the spins of the black holes before the merger with the orbital angular momentum. However, the information provided by χ_{eff} about the total angular momentum budget of the binary is limited: the inclination of the spins and their intrinsic magnitude are degenerate. Measures of the precession rate of the orbital angular momentum vector are needed in order to break the degeneracy and to better constrain the individual quantities $\boldsymbol{L}$, $\boldsymbol{S}_1$, $\boldsymbol{S}_2$; nevertheless, in order to measure the precession, long waveforms are needed, since the precession rate is much longer than the orbital period. Furthermore, in order to resolve the signal modulation produced by the precession, a precise measure of the wave amplitude is required as well, but it is not easy to achieve given the very low signal-to-noise ratio of the detections.

The gravitational wave events detected so far have been associated to low values of effective spin. The statistical distribution of the values of χ_{eff} can help in identifying the formation mechanism of these objects. However, for statistical reasons, a relatively large number of events is required in order to discriminate between the field or the dynamical scenario.

On the one hand, in the field formation scenario, the progenitors of the binary black holes have spent all their life together in a binary system. This implies that

[2]The main effect of χ_{eff} on the waveform can be understood by noting that the spin-orbit alignment affect the radius of the ISCO (see Sec. 1.3.2), i.e. the frequency of the wave when the merger takes place.

the relative orientation of the spins with respect to the orbital plane of the binary is affected by mechanisms able to align (tidal interaction) or misalign (natal kicks, O'Shaughnessy et al. 2017 and Wysocki et al. 2018) the spins, that leave their peculiar footprint in the spin orientation. The effective spin distribution expected for a sample of black holes formed through the field scenario is peaked at high values of $\chi_{\rm eff}$, assuming spin magnitudes as high as those estimated in several X-ray binaries (Fragos & McClintock 2015). However, the magnitude of the spin when a black hole forms is thought to depend on the mass of its progenitor. In particular, Belczynski et al. (2017) showed that low native spins are perfectly consistent with high-mass progenitors. As a consequence, within the field scenario, the effective spin distribution is expected to be peaked at low values if the binary components have massive star progenitors.

On the other hand, binary black holes that form through the dynamical scenario are expected to show random isotropic spin orientation. Therefore, the associated $\chi_{\rm eff}$ value distribution, in a sample of gravitational wave events of binaries formed through the dynamical scenario, is expected to be uniform. This implies that a significant fraction of the gravitational wave events detected from such a sample will be characterized by low effective spin values, analogously to what expected for the field formation scenario of black hole binaries with massive progenitors.

In conclusion, in light of these considerations, the measurement of small values of $\chi_{\rm eff}$, as those of all the gravitational wave events detected so far, does not allow us to discriminate between the two formation scenarios, since the effective spin distributions in both cases are characterized by a large fraction of events with low effective spin. A larger number of detections is required in order to better constrain the origin of the binary progenitor.

However, we note that Gerosa & Berti (2017) safely excluded (evidence of 2σ) the hypothesis that the gravitational wave events detected so far involved binaries composed by two second generation black holes. Indeed, at the merger, the residual angular momentum of the binary is converted into spin of the remnant black hole. It can be shown that the remnant spin is almost completely independent from the initial spin of the merging black holes; in particular, after the merger, the remnant black hole always has an intrinsic spin $\chi_{\rm s} \approx 0.7$, which is not consistent with any of the $\chi_{\rm eff}$ measured in the observed events.

5.2 Supermassive black hole binaries

The ΛCDM cosmological model predicts the ubiquitous formation of supermassive black hole binary systems. According to this model, galaxies in the near universe were assembled via the hierarchical merger of smaller structures at high redshift (see Fig. 5.2for two examples of merging galaxies). Since we know that most galaxies host in their central regions a supermassive black hole (Kormendy & Ho 2013), it is absolutely natural to assume that after a galaxy merger two supermassive black holes coexist in the same galaxy.

Few observation of dual AGNs with separations of the order of $1-10\,$kpc (Comerford et al. 2015, and references therein) appears to support this hypothesis. At shorter separations, apart from the detection of two dual AGN sources with a

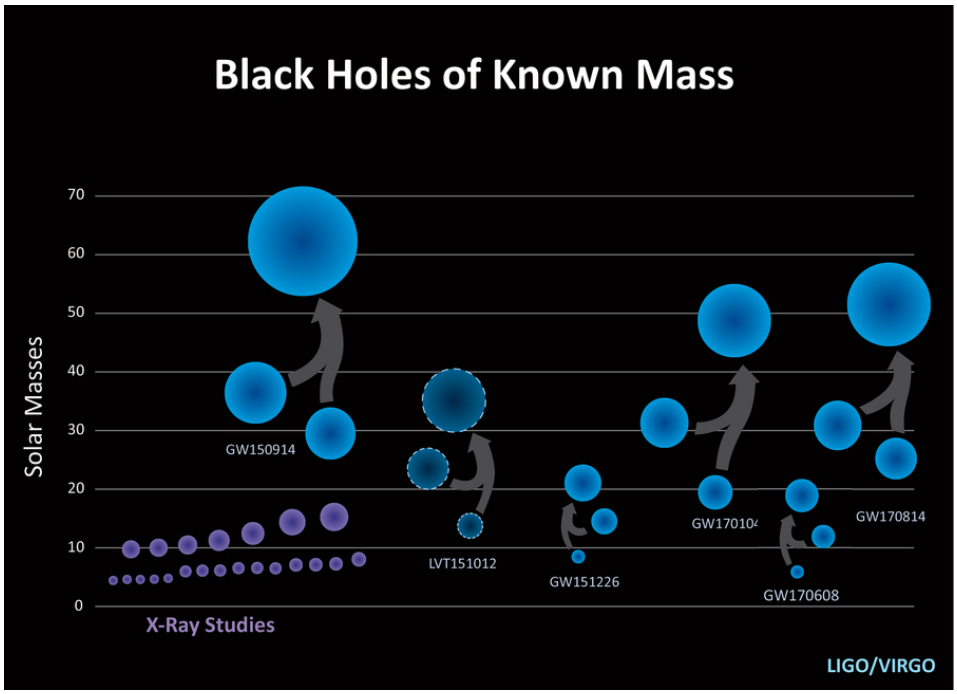


Figure 5.1: Masses of the progenitors and remnants of the first 5.87 LIGO/Virgo black hole merger detections (large blue circles, the dashed one had 13% probability of being consistent with noise fluctuation), compared to the masses of the observed X-ray binaries (violet circles). Edit of the original images by LSC/LIGO/Caltech/Sonoma State (Aurore Simonnet).

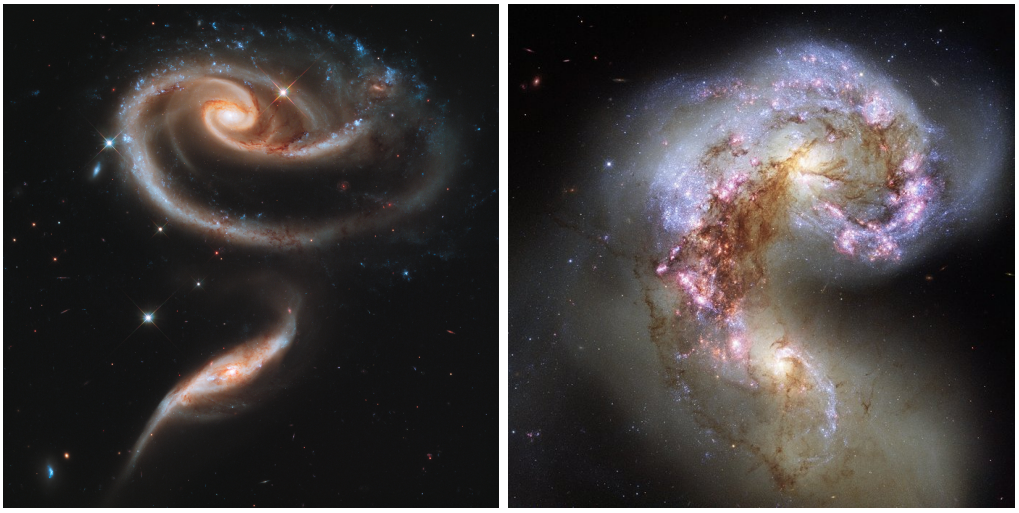


Figure 5.2: Left panel: Arp 273, an example of two interacting galaxies (UGC1810 above and UGC1813 below) located at 350 million ly in the Andromeda constellation, showing prominent spiral features, typically observed in tidally interacting objects. Right panel: Antennae galaxies, two merging galaxies located at 65 million ly in the Corvus constellation. Images credit: NASA/ESA Hubble Space Telescope.

projected separation of 7.3 pc (B2 0402+379, Rodriguez et al. 2006) using the Very Large Baseline Array (VLBA), and one with a projected separation of 0.35 pc (NGC 7674, Kharb et al. 2017) using the Very Long Baseline Interferometry (VLBI), no other direct observations of black hole pairs at separations below 1 kpc have been provided so far; in this respect, D'Orazio & Loeb (2018b) recently claimed that using the Very Large Baseline Interferometry (VLBI) it should be possible to resolve a number of periodic sources in the close universe ($z \lesssim 0.5$).

A few possible candidates of sub-parsec black hole binaries have been inferred with indirect observations. Such candidates were obtained using two main techniques: firstly, using measurements of offset broad emission-lines, that are expected to be emitted from the gas orbiting the individual black holes, carrying information about the orbital motion of the binary (Bogdanović 2015; D'Orazio et al. 2015; Wang et al. 2017); secondly, using photometric measurements of quasi-periodic sources: the expectations of periodic gas accretion in these system (see the next chapter) implies that periodic electromagnetic emission in AGNs represents the possible signature of black hole binaries (e.g. the quasar OJ287 show a qausi-periodic light curve with a periodicity of 12 years, which appears to be perfectly consistent with a model SMBHB+accretion disc Valtonen et al. 2008).

Regarding this second technique, a number of possible candidates of unresolved black hole binaries with sub-pc separations have been suggested in recent years thanks to two large surveys that searched for stationary optical transients in the sky for the past nine years. Firstly, the Catalina Real-time Transient Survey (CRTS, Drake et al. 2009): its data were used to identify a total of 111 candidates of supermassive black hole binaries (Graham et al. 2015) in $\sim 250 \times 10^3$ quasars covered by the survey; secondly, the Palomar Transient Factory (PTF, Rau et al. 2009): it identified 33 candidates supermassive black hole binaries in 35.5×10^3 observed quasar Charisi et al. (2016).

Recently, Krause et al. (2018) suggested that the presence of a supermassive binary black hole might cause the precession of radio jets with a characteristic frequency (driven by the orbital motion and geodetic precession of the binary), from which the orbital period of the binary can be inferred. In particular, using the VLA radio maps of the object Cygnus A, they found evidence of the precession of the radio-jet with a timescale of $0.5-3$ Myr, which appears to be consistent with the presence of a sub-parsec supermassive black hole binary with an orbital period of 18 yr.

As a general result, the paucity of binary systems at sub-pc separations suggests a rapid orbital decay of the binary, until the supermassive black holes coalesce on a reasonably short time-scale (Begelman et al. 1980). This explanation has been recently strengthened by the constraints placed on the GWB (gravitational wave background) through the PTA (Pulsar Timing Array) technique (Shannon et al. 2015; Lentati et al. 2015; Arzoumanian et al. 2016), suggesting that the binary shrinking could be even faster than so far predicted. Furthermore, it should be kept in mind that the GWB (Gravitational Wave Background) implied by the CRTS and PTF is in tension with the upper limits constrained by the PTA (Pulsar Timing Array) data. In particular, the hypothesis that the candidate from these surveys are false positives is preferred over supermassive black hole binary

hypothesis with an evidence $\gtrsim 2.3\sigma$ (Sesana et al. 2018). Recently Kelley et al. (2018) provided new estimates for the detection rates in the CRTS and in the forthcoming Large Synoptic Survey Telescope (LSST) of variable AGN sources. The new estimates predict that only ~ 5 supermassive black hole binaries can be detected in the CRTS as variable sources (consistently with the limits on the GWB put by the PTA data). In the LSST the number of massive black hole binaries detectable as variable sources is expected to grow to ~ 100 sources, thanks to the increase in the sensitivity and coverage of the survey.

5.2.1 Binary hardening in supermassive black hole binaries

A number of physical processes cause black hole pairs (the black holes are not gravitationally bound yet) to sink toward the centre of the newly formed galaxy (see Dotti et al. 2012; Colpi & Sesana 2017; Celoria et al. 2018 for comprehensive reviews). In particular, four main mechanisms have been suggested to drive the inward motion of the black holes (Begelman et al. 1980): the dynamical friction (active at separations a spanning $10\,\mathrm{pc} \lesssim a \lesssim 10\,\mathrm{kpc}$), the interaction with the gas ($1\,\mathrm{pc} \lesssim a \lesssim 100\,\mathrm{pc}$), the interaction with individual stars ($1\,\mathrm{pc} \lesssim a \lesssim 10\,\mathrm{pc}$), the gravitational waves emission ($a \lesssim 10^{-2}\,\mathrm{pc}$).

The interaction of the black holes with stars in the newly formed galactic core provides the dynamical friction (Chandrasekhar 1943) on each black hole necessary to drive the inspiral motion on a time scale given by (Yu 2002):

$$t_{\mathrm{df},i} \sim \frac{4 \cdot 10^6}{\log N_\mathrm{c}} \left(\frac{\sigma_\mathrm{c}}{200\,\mathrm{km\,s^{-1}}}\right) \left(\frac{r_\mathrm{c}}{100\,\mathrm{pc}}\right)^2 \left(\frac{10^8 M_\odot}{M_{\bullet,i}}\right) \mathrm{yr}, \tag{5.2}$$

where σ_c is the stars dispersion velocity, r_c is the radius of the core, N_c the number of stars hosted in the core and $M_{\bullet,i}$ is the mass of the i-th black hole. With the increase of the black holes velocities, the dynamical friction becomes progressively ineffective in driving the inspiral motion; the two black holes become gravitationally bound in a "soft" binary configuration, i.e. the kinetic energy of the binary becomes comparable to that of the surrounding stars; this occurs at separations a_soft of the order of

$$a_\mathrm{soft} \approx \frac{G(M_{\bullet,1} + M_{\bullet,2})}{\sigma_\mathrm{c}^2} \sim 11.2 \left(\frac{M_{\bullet,1} + M_{\bullet,2}}{10^8\,M_\odot}\right) \left(\frac{200\,\mathrm{km\,s^{-1}}}{\sigma_\mathrm{c}}\right)^2 \mathrm{pc}; \tag{5.3}$$

at this separation, the interaction with individual stars as three-body encounters becomes the dominant mechanism providing the inward motion of the black holes.

When the binary finally becomes "hard" ($a_\mathrm{hard} \approx a_\mathrm{soft}/4$), the stars responsible for the loss of energy and angular momentum are very close to the binary orbit, and are part of the set of stars called "loss-cone".

At the same time, for separations $a \lesssim 100\,\mathrm{pc}$, if the parent galaxies were originally gas rich, the interaction of the two black holes with the surrounding gaseous environment[3] concurs to reduce their separation and the eccentricity of the black holes by gas friction (Escala et al. 2005; Dotti et al. 2006; Mayer et al. 2007; Dotti

[3] A gaseous disc forms in the centre of the new galaxy as a consequence of angular momentum conservation (Barnes 2002), if the parent galaxies are gas rich (wet merger).

et al. 2007, 2009), even though the role of gas friction during this phase has been recently put into question (Pfister et al. 2017). This mechanism is conceptually similar to type I migration (see Sec. 4.4) since the black holes are completely embedded in the gaseous disc but no resonances can be found in the disc (at these distances the binary is not bound yet). Gas friction also provides the circularization of the orbits of the two black holes. When the binary becomes hard, the gap opening criterion is always satisfied[4] and the binary is expected to further evolve via type II migration (Haiman et al. 2009b).

Finally, at separations $a \lesssim 10^{-2} - 10^{-3}\,\mathrm{pc}$, the emission of gravitational waves becomes the dominant migration mechanism (see next section). The emission of gravitational waves further shrinks the separation of the binary and drives it towards the coalescence. For gaseous rich galaxies, the separation at which this occurs is generally referred to as "decoupling radius": the disc keeps evolving on the viscous timescale, while the evolution of the binary separation evolves at a faster rate $\dot{a}_{\mathrm{gw}}$ (see next section).

5.2.2 The final parsec problem

As soon as the binary has ejected all the stars involved in the interaction ("loss-cone depletion", which typically occurs at separations $\approx 1\,\mathrm{pc}$ for $M_\bullet \approx 10^8\,M_\odot$) the binary stalls (Quinlan 1996). Furthermore, the gaseous driven type II migration is expected to sensibly slow down for sufficiently high secondary-to-disc mass ratios (see Eq. 4.29). If this was the case, most galaxies in the Universe should host in their central regions a stalled supermassive black hole binary whose merging timescale is larger than the Hubble time. This has constituted over the past years the so called "final parsec problem".

It is reasonable to believe that in a relatively short timescale some new mechanism sets in, allowing the further shrinkage until the binary reaches the final gravitational wave driven migration toward the coalescence.

One possible solution that has been suggested is that the refill of the loss-cone occurs in much shorter timescales in triaxial cores (Merritt & Poon 2004; Vasiliev et al. 2015; Khan et al. 2016) than in spherical ones (Milosavljević & Merritt 2001, 2003; Makino & Funato 2004), implying that in some galaxies the interaction with stars might actually be effective in bringing the binary also at sub-parsec separations. Massive perturbers such as giant molecular clouds or clusters are also expected to accelerate the replenishment of the loss-cone (Perets & Alexander 2008), preventing the stall of the binary.

Another possibility that has been suggested is that a new galaxy merger while the two black holes are stalled at separations $\approx 1\,\mathrm{pc}$ might lead to the formation of black hole triplets (Bonetti et al. 2018a,b; Ryu et al. 2018) allowing the binary to further reduce its separation.

The interaction of the binary with the environment gas has been also proposed to play a role here (see Mayer 2013 for a review). Mechanisms such as those predicted for planetary migration in young solar systems (Armitage & Natarajan 2002; MacFadyen & Milosavljević 2008; Cuadra et al. 2009; Roedig et al. 2012;

[4] Apart from extreme mass ratio inspirals (EMRI).

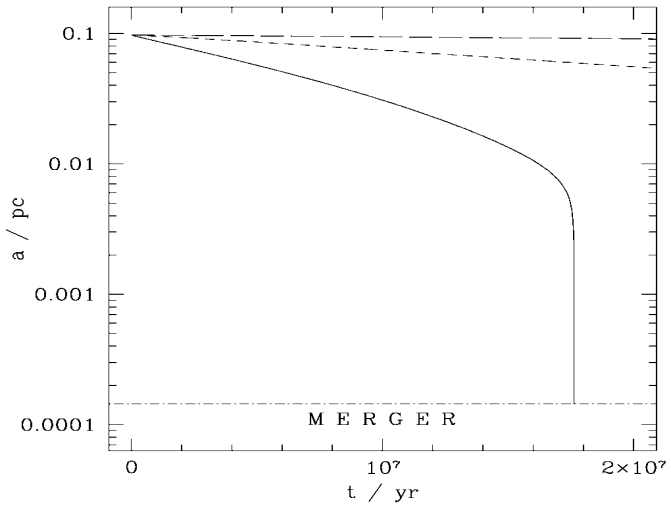


Figure 5.3: On the y-axis the binary distance in pc, on the x-axis time in years. The plot shows the inspiral motion for a $q = 0.02$ super-massive black hole binary embedded in a gaseous disc, with primary mass $M_1 = 10^8\,\mathrm{M}_\odot$. Long dashed, short dashed, thick lines are curves respectively for the accretion rates $\{10^{-2}; 10^{-1}; 1\}\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$. It can be clearly seen that, after a gaseous-torques driven migration, for $a < 0.01\,\mathrm{pc}$, the gravitational waves emission becomes the dominant mechanism, shortening drastically the migration time-scale. Original figure from Armitage & Natarajan (2002).

Tang et al. 2017a, see also Sec. 4.5 of this manuscript), "faster" type II migration (Duffell et al. 2014), retrograde discs with respect to the binary orbital motion (Nixon et al. 2011) or the interaction with gaseous clumps (Goicovic et al. 2016) have been shown to be successful in driving the binary to sub-parsec separations.

However, in this context, the possible fragmentation of the disc, that occurs at large radii where the disc is gravitationally unstable, might slow down or even stop the migration much before the gravitational wave emission phase (Lodato et al. 2009). AGN feedback may also remove the gas from the binary surrounding stalling the binary evolution (del Valle & Volonteri 2018). We also report that Muñoz et al. (2018) recently claimed that circum-individual discs are able to produce a positive torque on the binary producing an outward migration, thus preventing the merger of the two black holes. This result is in contrast with what found by Tang et al. (2017a). The treatment of the accretion prescription onto the sinks appears to be a key ingredient to properly describe the evolution of the system; however, the origin of the discrepancy between the two papers deserves further investigations.

5.3 Gravitational wave driven inspiral

Black hole binaries are powerful sources of gravitational waves. Since gravitational waves carry energy and momentum along their path, enforcing the conservation of the total energy and momentum of the system during their emission causes the binary to lose some orbital energy and orbital angular momentum. This results in an effective negative torque that shrinks the binary up to the coalescence at the

following rate $\dot{a}_{\rm gw}$ and $\dot{e}_{\rm gw}$ (Peters 1964):

$$\dot{a}_{\rm gw} = -\frac{64}{5}\frac{G^3 M_{\bullet,\rm bin}^3}{c^5 a^4}\frac{q}{(1+q)^2}a\left(1+\frac{73}{24}e^2+\frac{37}{96}e^4\right)(1-e^2)^{-7/2}, \tag{5.4}$$

$$\dot{e}_{\rm gw} = -\frac{304}{15}\frac{G^3 M_{\bullet,\rm bin}^3}{c^5 a^4}\frac{q}{(1+q)^2}e\left(1+\frac{121}{304}e^2\right)(1-e^2)^{-5/2}. \tag{5.5}$$

where q is the binary mass ratio, $M_{\bullet,\rm bin}$ is the total mass of the binary, c is the speed of light, e is the binary eccentricity and G is the gravitational constant. It should be noted that during the gravitational wave inspiral, the orbit of the binary tends to circularise

$$\frac{de}{da} \approx \frac{19}{12}\frac{e}{a} \rightarrow e \sim a^{19/12}. \tag{5.6}$$

For mergers taking place in environments rich of gas, the binary undergoes type II migration over a timescale that is typically longer than the viscous time t_ν (see Eq. 4.29). As a consequence, the material in the circumbinary disc accumulates beyond the orbit of the secondary black hole and spreads inward at the same rate of the binary. When the gravitational wave emission starts providing a migration rate of the binary that is faster than the viscous spreading of the disc, the evolution of the binary decouples from that of the disc: the disc is left "behind", while the binary approaches the coalescence. The binary separation at which this occurs is referred to as "decoupling radius" $a_{\rm dec}$.

By equating the viscous migration rate $v_R^{\rm visc} = -a/t_\nu$ with $\dot{a}_{\rm gw}$ in Eq. (5.4) we get the following estimate[5] of $a_{\rm dec}$ (Armitage & Natarajan 2002)

$$a_{\rm dec} \approx 2300\, R_{\rm g}\left(\frac{H/R}{10^{-3}}\right)^{-4/5}\left(\frac{\alpha}{0.1}\right)^{-2/5}\frac{q^{2/5}}{(1+q)^{4/5}}\frac{(1-e^2)^{7/2}}{1+\frac{73}{24}e^2+\frac{37}{96}e^4}, \tag{5.7}$$

where we recall that $R_{\rm g} = GM_{\bullet,\rm bin}/c^2$ is the gravitational radius (Eq. 1.27).

The decoupling radius sets the reference separation for the beginning of the gravitational wave inspiral. Figure 5.3 shows an example of the rapid time evolution of the semi-major axis of the binary after the binary crosses the decoupling radius.

5.4 Kick velocity

Depending on the properties of the black hole binary, the emission of gravitational waves might have a complex multipole structure that result in an anisotropic momentum flux. Such a flux of momentum imprints a recoil velocity to the merger remnant, also known as "kick velocity". A semi-analytical expression for the recoil velocity $v_{\rm kick}$, based on fully non-linear numerical relativity simulations and post Newtonian analysis of the problem (Fitchett 1983; Kidder 1995), was provided by Campanelli et al. (2007) (see also Lousto & Zlochower 2008 and Rezzolla 2009),

[5] See also Milosavljević & Phinney (2005) and Gold et al. (2014)

which reads

$$\boldsymbol{v}_{\rm kick}(q, \boldsymbol{S}_1, \boldsymbol{S}_2) = v_m\, \hat{\boldsymbol{e}}_1 + v_\perp(\cos\xi \hat{\boldsymbol{e}}_1 + \sin\xi \hat{\boldsymbol{e}}_2) + v_\parallel \frac{\boldsymbol{L}}{|\boldsymbol{L}|} \tag{5.8}$$

$$v_{\rm m} = A\frac{q^2(1-q)}{(1+q)^5}\left(1 + B\,\frac{q}{(1+q)^2}\right), \tag{5.9}$$

$$v_\perp = H\frac{q^2}{(1+q)^5}\left(a^{\rm s}_{2,\parallel} - q a^{\rm s}_{1,\parallel}\right), \tag{5.10}$$

$$v_\parallel = K\cos(\Theta - \Theta_0)\frac{q^2}{(1+q)^5}\left(a^{\rm s}_{2,\perp} - q a^{\rm s}_{1,\perp}\right), \tag{5.11}$$

where $A = 1.2 \times 10^4\,{\rm Km\,s^{-1}}$, $B = -0.93$, $H = 6.9 \times 10^3\,{\rm Km\,s^{-1}}$, $K = 6 \times 10^4\,{\rm Km\,s^{-1}}$, are numerical parameters; $\hat{\boldsymbol{e}}_1$ and $\hat{\boldsymbol{e}}_1$ are orthogonal vectors lying in the orbital plane of the binary; $\boldsymbol{S}_{1,2}$ are the spin vectors of the two black holes; $a^{\rm s}_{1,2}$ are the dimensionless spin parameters (Eq. 1.26) of the primary and secondary black holes; the subscripts $\parallel$ and $\perp$ refer to the spin projection on the orbital angular momentum vector $\boldsymbol{L}$.

The velocity v_m is the component of the recoil velocity that arises when the two black holes have different masses, $v_\perp$ accounts for the recoil velocity due to asymmetries in the spins of the black holes, $v_\parallel$ is the component of $\boldsymbol{v}_{\rm kick}$ parallel to the vector of the binary orbital angular momentum.

The dimensionless parameter ξ is the angle between the direction of the contribution to $\boldsymbol{v}_{\rm kick}$ given by $v_{\rm m}$ and the one of $v_\perp$, it depends on the configuration of the system. Lousto & Zlochower (2008) report that $\xi \approx 145°$ for a large set of quasi-circular configurations. The quantity $\Theta - \Theta_0$ is the angle between the in-plane component of the spins $S_2^\perp + S_1^\perp$ and the infall direction at merger (the direction along which the black holes approach each other).

Eq. (5.8–5.11) tell us that the recoil velocity depends on the degree of asymmetry between masses and spins of the two black holes. Configurations with non spinning black holes ($\chi^{\rm s,1} = \chi^{\rm s,2} = 0$) can lead to maximum recoil velocity $|\boldsymbol{v}_{\rm kick}| \approx 175\,{\rm Km\,s^{-1}}$ for mass ratios $q \approx 0.36$. However, for spinning black holes the recoil velocities might become as big as $|\boldsymbol{v}_{\rm kick}| \approx 4000\,{\rm Km\,s^{-1}}$ in the so called "super-kick" configuration[6], where the black holes are maximally spinning and have their spin anti-aligned, lying in the orbital plane ($\chi^\perp_{\rm s,1} = -\chi^\perp_{\rm s,2} = \pm 1$).

High recoil velocities after the merger might cause the newly formed black hole to be ejected from the sites where the merger occurred. Typically this happens for kick velocities $|\boldsymbol{v}_{\rm kick}| \approx 50\,{\rm Km\,s^{-1}}$ for mergers from stellar black holes progenitors inhabiting star clusters and $|\boldsymbol{v}_{\rm kick}| \approx 800\,{\rm Km\,s^{-1}}$ for supermassive black holes progenitors inhabiting galactic nuclei (Komossa 2012; Colpi & Sesana 2017). If the recoil velocity is not sufficient to escape the gravitational potential of the parent stellar cluster or galaxy, the remnant is expected to slowly fall back after it has dissipated the kinetic energy imparted during the merger. In the case supermassive black hole binaries, the long timescales required for the remnant to

[6]Note that Lousto & Zlochower (2011) found even faster kick velocities $|\boldsymbol{v}_{\rm kick}| \approx 5000\,{\rm Km\,s^{-1}}$, in the so called "hangup" configuration with spin partially misaligned with respect to the orbital plane.

return back to its original location will likely produce supermassive black holes appearing displaced with respect to the centre of the host galaxy.

5.5 The relevance of the accretion rate for black hole binaries

The peculiar dynamics of the gas around the two black holes is of fundamental importance in the observational characterization of these sources.

Predictions regarding their luminosity and in general their electromagnetic outputs need necessarily to rely on the accretion rate of gas on to the two black holes (Perna et al. 2016; de Mink & King 2017; Farris et al. 2015a; Tang et al. 2018). Furthermore, the relativistic motion of the material in the system shapes the peculiar Doppler-shifted nature of line emission of their electromagnetic signals (D'Orazio et al. 2015; Tang et al. 2018).

Finally, the accretion rate influences the black hole spin alignment process (Bogdanović et al. 2007; Miller & Krolik 2013; Lodato & Gerosa 2013; Gerosa et al. 2015). As a consequence, it determines the gravitational wave frequency pattern and the recoil velocity of the black hole remnant after the coalescence, with important consequences for the detection of offset AGN emission produced by kicked supermassive black holes (Blecha et al. 2016) or, more generically, the peculiar delayed electromagnetic emission after the merger (Rossi et al. 2010).

An accurate determination of the accretion rate on to black hole binaries is thus required in order to understand their dynamical evolution and define the signatures of their presence in galactic nuclei.

5.5.1 EM emission from stellar black hole binaries

The only evidence that we have of the existence of stellar black hole binaries are the few gravitational wave events detected to date. Gamma ray bursts are expected to be associated with the merger of binaries of compact objects. However, a large fraction of the scientific community agrees that, in order to have an electromagnetic counterpart to gravitational wave events, at least one of the elements of the binary must be a neutron star. This requirement is dictated by the need of filling the environment with some material responsible for the electromagnetic emission (Giacomazzo et al. 2013). This prediction was confirmed by the recent spectacular detection of a gamma ray burst (Fermi-GBM, Goldstein et al. 2017; INTEGRAL, Savchenko et al. 2017b; AGILE, (Verrecchia et al. 2017b)) associated with the detection of the gravitational wave event GW170817, interpreted as a merger of two neutron stars (Abbott et al. 2017c).

To date, only two gravitational wave detections from black hole binary mergers have been reported to show an electromagnetic counterpart (GW150914 Abbott et al. 2016b by Fermi GBM Connaughton et al. 2016, and GW170104 Abbott et al. 2017a by AGILE Verrecchia et al. 2017b, see below in this section for further details). Suggesting the idea that stellar black hole binary mergers might occur in gas poor environments and, for this reason, not to show an electromagnetic signature.

However, this issue is far from being assessed. Neither clear evidence of the presence of gas in the environment nor a reliable model of the mechanism producing a prompt signal during mergers of stellar black hole binaries are available yet.

On the one hand, the lack of electromagnetic counterparts from black hole mergers might imply that stellar black hole mergers are typically "dry", and that no material is present around the binary at the time of the merger. On the other, all the gravitational wave events so far detected were extra-galactic, so that the emission might be too faint to be detected by the current observational facilities.

Anisotropic electromagnetic emission have also been suggested as a possible explanation of the lack of electromagnetic counterparts. A statistical analysis based on the properties of the black hole binary population from the isolated binary formation channel might suggest that the lack of electromagnetic counterparts of the events so far detected is consistent with a beamed emission within an angle $\lesssim 50^\circ$ (Perna et al. 2018), using as a reference electromagnetic luminosity the one associated with GW150914 detected by the Fermi GBM (Connaughton et al. 2016).

However, we report that the claimed[7] detection by the Fermi Gamma Ray Monitor (GBM) of a short gamma ray signal (Connaughton et al. 2016) associated with the gravitational wave event GW150914 (Abbott et al. 2016b), and by AGILE (Verrecchia et al. 2017a) for GW170104, has triggered a lot of effort in the scientific community to speculate about possible scenarios producing prompt electromagnetic emission also in black hole binary mergers.

In this context, besides the need to explain the origin of the gas surrounding the binary, which is not obvious, a mechanism promptly activating the emission is required. Indeed, the accretion of material in a "stationary" regime is not sufficient for the detection of electromagnetic radiation.

The main argument regarding the presence of environmental gas in these systems comes from their progenitors. Assuming the field scenario as responsible for the formation of the black hole binary (the dynamical one does not allow the survival of a circumbinary disc, de Mink & King 2017), the stellar progenitors of the black holes are supposed to shed in the surroundings a fraction of their mass with many possible mechanisms, among them: mass loss through the outer Lagrange points, stellar winds, common envelope ejection, Roche lobe overflow (a circumbinary disc forms, de Mink & King 2017 and references therein), supernova fallback (a circum-secondary/primary disc forms around the individual black holes Perna et al. 2016; Murase et al. 2016; Kimura et al. 2017; Perna et al. 2018 and references therein).

In the case of a circumbinary disc, the binary is initially starved (as we will

[7] After a re-analysis of the data collected by Fermi-GBM telescope with an alternative technique, Greiner et al. (2016) concluded that the transient, associated with an electromagnetic counterpart of GW150914, was consistent with a background fluctuation. However, Connaughton et al. (2018), reported some inconsistencies in the analysis performed by Greiner et al. (2016). They re-applied the analysis used by Greiner et al. (2016) and found that the false alarm hypothesis associated with the transient can be excluded within $\sim 3\sigma$, in fact confirming the reliability of their previous result. Analogously, the detection of an electromagnetic signal possibly associated with the event GW170104 by AGILE (Verrecchia et al. 2017a), has been put in question due to the non-detection by Fermi GBM and INTEGRAL Savchenko et al. (2017a).

discuss in the next chapter, see also the recent paper by Martin et al. 2018) and the accretion rate on to the black holes is very low. As the merger takes place, the prompt mass loss of the binary due to the emission of gravitational waves causes the material to move on eccentric orbits around the remnant (Lippai et al. 2008): since the disc is cold, the reduction of the central mass alters the centrifugal balance causing the gas orbits to become eccentric; the intersection of elliptical orbits produces hypersonic shocks that rapidly heat up the disc, being responsible for the prompt emission (Rosotti et al. 2012; de Mink & King 2017). Furthermore, the recoiling black hole remnant is expected to strip a portion of the circumbinary disc. The motion of the gas around the remnant is turbulent and chaotic, providing typical delayed electromagnetic emission (Rossi et al. 2010).

A second possible scenario consists of the formation of a circum-primary/secondary disc following a supernova fallback. The material falls back on one of the two black holes forming a cold (Menou et al. 2001) quiescent disc. As the black holes approach the merger, tidal effects start heating up the disc, triggering the prompt emission. Just prior[8] to the merger, the remnant of the accretion disc is squeezed between the two black holes producing a rapid increase in the accretion rate and a bright peak in the electromagnetic luminosity of the system (Chang et al. 2010; Tazzari & Lodato 2015; Cerioli et al. 2016; Perna et al. 2016; Murase et al. 2016). We report that this effect has been put into question (Baruteau et al. 2012; Fontecilla et al. 2017).

For completeness we mention that, right after the detection of the gamma ray burst (Connaughton et al. 2016) associated with the gravitational wave event GW150914 (Abbott et al. 2016b), a possible scenario where a binary black hole forms directly in the centre of a massive star was suggested (Loeb 2016; D'Orazio & Loeb 2018a). In this scenario, as a consequence of core fragmentation, the intra-star formation of the binary promote a high accretion rate and electromagnetic luminosity. However, this scenario has received strong criticism by Woosley (2016) since it is based on the not documented assumption that the progenitor star has a rapidly rotating He core.

5.5.2 EM emission from supermassive black hole binaries

As discussed in Sec. 5.2, no direct evidence of the existence of supermassive black hole binaries have been provided so far, even though some indirect hints and cosmological considerations point in that direction. The currently operating gravitational wave detectors are not sensitive to the typical frequencies of gravitational wave inspirals of supermassive black holes: they are much lower than those of stellar black holes and thus require much larger interferometers to be detected, such as the Laser Interferometer Space Antenna (LISA, planned launch date 2034). However, in contrast with stellar black hole binaries, supermassive black holes are known to be the engines that make active galactic nuclei (AGN) very bright sources of electromagnetic radiation, so that their electromagnetic signature might be detected well before the merger.

[8] Conversely, Kimura et al. (2017) claims instead that the circum-individual disc becomes bright much before the merger.

The characterization of the electromagnetic signature of supermassive black hole binaries is thus of fundamental importance in order to identify possible candidates among AGN sources.

The electromagnetic emission in supermassive black hole binaries is produced by the same mechanisms taking place in AGN accretion discs, with the main difference that the perturbation of the axial symmetry of the system can introduce peculiar features in the light-curve and in the SED of the source.

The presence of a binary object interacting with a gaseous disc is expected to modulate the accretion rate, introducing a periodic variability of the luminosity with a frequency comparable to that of the binary $\omega \approx \Omega_{\rm bin}$, where $\Omega_{\rm bin}$ was defined in Eq. (1.16) (D'Orazio et al. 2013; Farris et al. 2014; Ragusa et al. 2016; Miranda et al. 2017). Furthermore, as we will discuss in the next chapter, the formation of circum-individual discs can also act as a buffer for the accretion, smoothing out the variability and introducing an additional modulation of the emission with a lower frequency $\omega \approx \Omega_{\rm bin}/7$ (Farris et al. 2014; Ragusa et al. 2016).

The relativistic motion of the black holes is expected to provide a characteristic, frequency dependent, Doppler-shift of the emission coming from the circum-secondary disc if the source is sufficiently inclined with respect to the line of sight (D'Orazio et al. 2015; Tang et al. 2018). The presence of dense circum-primary and circum-secondary discs has also been found to provide a characteristic increase in the luminosity in the high energy end of the SED (Farris et al. 2015a).

As discussed in Sec. 5.3, during the gravitational wave inspiral phase, the binary decouples from its gaseous circum-binary disc. Numerical simulations (Farris et al. 2015b; Tang et al. 2018) have shown that the electromagnetic emission from the system remains bright also after the decoupling, as the binary approaches the merger: the typical light-curves expected are characterized by a small reduction of the accretion rate (following the decoupling); then, a normal accretion rate is recovered right after the merger, as soon as the disc viscously spreads and reaches the black hole remnant.

Finally, large recoil velocities both parallel and perpendicular to the disc plane might result in the excitation of strong shock fronts producing some post-merger luminosity (Lippai et al. 2008; Haiman et al. 2009a; Rossi et al. 2010).

As we will discuss in the next chapter, a reduction of the accretion rate due to the dam effect provided by the tidal torque exerted by the binary might change our perspective of detecting signals from these systems. Indeed if only a small amount of material is allowed to leak through the disc cavity, the AGN activity around the individual black holes is not fueled, resulting in a reduction of the luminosity.

5.5.3 Setting the direction of the spins

As mentioned in Sec. 5.4, the spins of the black holes at merger are expected to play a fundamental role in determining whether large kick velocities are imparted to the remnant. Furthermore, the frequency pattern of gravitational waves is affected by the spin dynamics, so that indirect information about the processes in action can be obtained by the analysis of the wave-forms.

The interaction of the black holes with some surrounding gas on their way to the merger might significantly affect the spin orientation. Indeed, if the gaseous disc forming around individual black holes is not aligned with their equatorial plane, it is expected to change their spin orientation. In general, if this mechanism is effective on a relatively short timescale, the spins are expected to align with the orbital angular momentum vector of the binary (i.e. with its orbital plane) before the merger takes place. The physical mechanism responsible for the spin alignment is the (Bardeen & Petterson 1975) effect, which acts over a timescale (Gerosa et al. 2015)

$$t_{\mathrm{al},i} \simeq 3.4 \frac{M_{\bullet,i}}{\dot{M}_i} \alpha \left(\frac{\chi_{\mathrm{s,i}}}{\alpha_2} \frac{H}{R} \right)^{2/3}, \tag{5.12}$$

where $M_{\bullet,i}$ is the mass of i-th black hole, $\dot{M}_i$ is the accretion rate on to it, α is the Shakura & Sunyaev (1973) viscous parameter, $\chi_{\mathrm{s,i}}$ is the dimensionless spin (Eq. 1.26) and α_2 is the so called "vertical viscosity", that is related to the α-prescription by:

$$\alpha_2 = \frac{1}{2\alpha} \frac{4(1+7\alpha^2)}{4+\alpha^2}. \tag{5.13}$$

It becomes clear from Eq. (5.12) that small accretion rates on the black holes lead to long timescales for spin alignment and thus larger recoil velocities of the remnant. As a consequence, small accretion rates (or strong asymmetry between the accretion rates on individual black holes) increase the likelihood of ejection of the remnant from the galaxy or star cluster. Furthermore, if the two black holes accrete material at different rates (differential accretion Farris et al. 2014; Young et al. 2015; Young & Clarke 2015), one of the spins does align with the angular momentum vector of the binary while the other does not.

Much debate is still ongoing regarding the alignment timescale of the black hole spins. In particular, two main lines of thought can be identified: the timescale for spin alignment is short, implying mergers with aligned spins and small recoil velocities of the remnant in gas rich galaxy mergers (Bogdanović et al. 2007; Dotti et al. 2010; Miller & Krolik 2013); in contrast, other authors claim that the timescale for spin alignment is long and depends on the accretion rate on to individual black holes, implying mergers with misaligned spins and high recoil velocities of the remnant (Lodato & Gerosa 2013; Gerosa et al. 2015).

Chapter 6

On the suppression of the accretion in black hole binary systems

Based on the paper by ***Enrico Ragusa****, Giuseppe Lodato & Daniel J. Price (2016) "Suppression of the accretion rate in thin discs around binary black holes", Monthly Notices of the Royal Astronomical Society, 460, 1243*

"...And if the dam breaks open many years too soon
And if there is no room upon the hill
And if your head explodes with dark forbodings too
I'll see you on the dark side of the moon..."

Pink Floyd, Brain Damage

The issue of the mass flow within the cavity formed by a binary in its surrounding disc has been mostly studied in the protostellar case (e.g. Artymowicz & Lubow 1994, 1996) where the disc aspect ratio H/R is relatively large. In recent years, the interest has moved to the supermassive black hole binary case. However, mostly for numerical reasons, simulations of these systems have used $H/R \sim 0.1$, appropriate for protostellar binaries, but two orders of magnitude larger than the values expected for a disc surrounding a supermassive black hole binary. These studies all conclude that the binary potential does not prevent the gas from flowing within the cavity so that the accretion rate on to the binary, $\dot{M}_{\rm bin}$, is comparable to the equivalent rate for a single object, $\dot{M}_0$ (Roedig et al. 2012; Shi et al. 2012; D'Orazio et al. 2013; Farris et al. 2014; Shi & Krolik 2015). Here we investigate how this conclusion is modified when one adopts more realistic values for the disc temperature and aspect ratio.

This chapter is organized as follows: in Section 6.1 we summarize the known results for the accretion dynamics of gas orbiting a binary system. Section 6.2

describes the numerical method used in our simulations and the initial conditions. In Section 6.3 we first reproduce the results of earlier works using a large H/R and then show how the results change when reducing H/R. We discuss our results and draw conclusions in Section 6.4.

6.1 Accretion dynamics in binary systems

Adapting classical accretion disc theory, developed for a single central massive object, to binary systems is far from trivial. The tidal torques produced by the disc-satellite interaction (Lin & Papaloizou 1979; Goldreich & Tremaine 1980), strongly perturb the disc structure (Artymowicz & Lubow 1994) repelling the gas from the corotation region and clearing an annular gap across the orbit of the secondary object.

For sufficiently high mass ratios, the gap becomes so wide that the binary resides in a depleted cavity surrounded by a circumbinary disc. The effects of these strong tidal forces on the accretion are uncertain. Do they act as a dam or is the gas able to leak into the cavity from the circumbinary disc?

An analytical treatment of the gas dynamics in binary systems has been attempted (Pringle 1991; Liu & Shapiro 2010), predicting a suppression of the accretion rate for high values of the binary mass ratio $q = M_2/M_1 > 0.01$, where M_1 and M_2 are the primary and the secondary mass, respectively. However these models assume, being 1D, axial symmetry, which is not the case for a rotating binary potential.

Starting from the mid 1990s, the problem of accretion in binary systems has been treated numerically for both binary-star/planet-star systems (Artymowicz & Lubow 1996; Bate & Bonnell 1997; Günther & Kley 2002; Ochi et al. 2005; Hanawa et al. 2010; de Val-Borro et al. 2011; Dunhill et al. 2015) and supermassive black hole binaries (Hayasaki et al. 2007; MacFadyen & Milosavljević 2008; Cuadra et al. 2009; Roedig et al. 2012; Shi et al. 2012; D'Orazio et al. 2013; Farris et al. 2014; Shi & Krolik 2015).

These simulations showed that for aspect ratios $H/R \sim 0.1$ one or two streams of material — the number depending on the mass ratio q — flow through the edge of the cavity. These streams connect the binary with the edge of the cavity; their formation is periodic with characteristic frequencies, $\omega_{\rm stream}/\Omega_{\rm bin} \simeq 1 - 2$, where $\Omega_{\rm bin}$ is the binary frequency.

Various authors (MacFadyen & Milosavljević 2008; D'Orazio et al. 2013; Farris et al. 2014) noted the development of an eccentric cavity associated with an overdense lump of material at the edge of the cavity (see also Chapter 9). This lump orbits around the binary on an elliptic Keplerian orbit. When it reaches the pericentre of its orbit it causes a boost in the accretion, therefore adding also a lower characteristic frequency $\omega_{\rm lump}/\Omega_{\rm bin} = 2/9$. Recently Farris et al. (2014) showed that the accumulation of gas around each object in circum-individual "mini-discs" acts as a buffer for accretion, smoothing the faster periodicity and increasing the power of the $\omega_{\rm lump}/\Omega_{\rm bin} = 2/9$ one, with important consequences for observations.

The accretion rate on to the binary $\dot{M}_{\rm bin}$ has been found to be comparable to that predicted by classic disc accretion theory $\dot{M}_0$ in absence of the binary

companion (Shi et al. 2012; Roedig et al. 2012; D'Orazio et al. 2013; Farris et al. 2014; Shi & Krolik 2015), giving a normalized accretion rate $1 < \dot{M}_{\rm bin}/\dot{M}_0 \lesssim 1.6$, implying, de facto, that no significant suppression of the accretion occurs due to the presence of the binary. However, as we discuss, it is not surprising that lower values of $\dot{M}_{\rm bin}/\dot{M}_0$ are associated with lower viscous torques (MacFadyen & Milosavljević 2008; D'Orazio et al. 2013), or in general with less effective angular momentum transfer mechanisms in the disc.

Shi et al. (2012) and Shi & Krolik (2015) investigated this issue, trying to constrain the uncertainty on the viscous α-parameter (that in the literature typically spans the values $0.01 \leq \alpha \leq 0.1$) by performing MHD simulations, in order to provide a self-consistent physical mechanism for the angular momentum transport through the disc. They found that the angular momentum transport operated by the turbulent motion of the gas fed by MRI, is consistent with a corresponding viscous $\alpha = 0.1$ for an equal mass, circular binary system and a globally isothermal disc with $c_{\rm s} = 0.1\Omega_{\rm bin}a$, where a is the binary separation.

Regarding the differential accretion rate, i.e. how much material is accreted by each object, recent results claim that, for a fixed mass ratio q ranging between $0.1 \leq q < 1$, the ratio $\dot{M}_2/(\dot{M}_1 + \dot{M}_2)$ gets higher as the disc gets colder (Young & Clarke 2015; Young et al. 2015), where $\dot{M}_1$ and $\dot{M}_2$ are the accretion rate on the primary and secondary object, respectively. Indeed, accretion occurs mostly on the secondary object since it is closer to the cavity wall than the primary, but for high gas temperatures pressure allows the material to cross the L1 Lagrange point, thus accreting on the primary and reducing differential accretion[1]. For $q < 0.1$ the accretion returns progressively to occur mostly on the primary object (Farris et al. 2014). Anyway, the amount of material being accreted by the secondary is expected to alter the normal replenishment of the inner regions of the disc causing a reduction of the density inside the planet orbit with respect to the unperturbed state (Lubow & D'Angelo 2006).

We consider in this chapter coplanar prograde discs. Misaligned or even retrograde discs (Nixon et al. 2013; Roedig & Sesana 2014; Lubow et al. 2015; Dunhill et al. 2014; Aly et al. 2015; Nixon & Lubow 2015; Goicovic et al. 2016) have also been considered, demonstrating the importance of disc tearing and retrograde accretion in the evolution of such systems.

6.1.1 The problem of the accretion in black hole binary systems

Analytical models of binary systems with discs assume that the presence of a binary companion gives rise to tidal torques acting on the disc. Using the *impulse approximation* (Lin & Papaloizou 1979), developed under the assumption $q \ll 1$, the tidal torque density exerted by a satellite on the circumbinary disc may be approximated by (see also Eq. 3.9)

$$\frac{\partial T_{\rm tid}}{\partial R} = 2\pi f q^2 \Omega^2 a^3 \Sigma \left(\frac{a}{\Delta}\right)^4, \tag{6.1}$$

[1]This result resolved a long standing issue regarding the discrepancies in the literature between some works (Bate & Bonnell 1997; Ochi et al. 2005; Hanawa et al. 2010), attributing the qualitative and quantitative differences in the results to the different gas temperatures used in the simulations.

where Ω is the binary orbital frequency, a is the binary separation, R is the cylindrical radius, Σ the surface density, f is a dimensionless normalization factor and $\Delta = \max[R - a, H, R_{\rm Hill}]$, where H is the disc height and $R_{\rm H}$ is the Hill's radius (Syer & Clarke 1995).

The viscous torque density, responsible for disc accretion, is given by (see r.h.s. of Eq. 2.20)

$$\frac{\partial T_{\rm vis}}{\partial R} = \frac{\partial}{\partial R}\left(2\pi\nu R^3 \Sigma \frac{\partial \Omega}{\partial R}\right), \tag{6.2}$$

where ν is the viscous shear parameter. The direction of $T_{\rm tid}$ and $T_{\rm vis}$ is opposite, with the balance between the two responsible for the opening of a gap or even a cavity in the disc.

Even though equations (6.1) and (6.2) were developed for $q \ll 1$, this approximation can provide insights on scaling laws also for higher mass ratios ($q \sim 1$). Assuming an α-prescription by Shakura & Sunyaev (1973) for the viscous shear parameter $\nu = \alpha c_{\rm s} H$, where α is a dimensionless scale parameter, $c_{\rm s}$ is the sound speed and H the disc vertical displacement, we notice that both these torque terms scale with the disc aspect-ratio H/R: in particular, integrating equations (6.1) and (6.2) over the disc, one obtains $T_{\rm tid} \propto (H/R)^{-3}$ and $T_{\rm vis} \propto (H/R)^2$.

Although 1D models predict that no material can cross the gap/cavity edge (Liu & Shapiro 2010), 2D and 3D numerical simulations showed that material is able to stream inside the cavity with accretion rates comparable or even higher to those predicted for single objects even in presence of equal mass binaries, that should provide the most intense tidal torques.

However, even though it is widely believed that the accretion rate in binary systems is not affected much by the presence of tidal torques, a suppression similar to that predicted by 1D models may occur when the viscous torque weakens for example when the disc becomes thinner. The disc internal angular momentum transport mechanisms are more effective for hot/thick discs than for thin/cold ones, both in simple α-models (Shakura & Sunyaev 1973) and in physically based mechanisms, such as MRI and gravitational instability.

It is important to mention that the formation of streams of material has been credited to pressure effects (Lubow & Artymowicz 1997), since pressure can be thought as an alteration to the effective gravitational potential, allowing the mass to overcome the tidal barrier as a consequence of the conservation of the Bernoulli constant along the streamlines. This pressure effect has been recently confirmed by D'Orazio et al. (2016), even though it becomes relevant only for $H/R \gtrsim 0.1$ in equal mass ratio binary systems.

6.2 Numerical Simulations

We performed a set of 3D SPH (Smoothed Particle Hydrodynamics) simulations using PHANTOM (Lodato & Price 2010; Price & Federrath 2010; Price 2012; Price et al. 2018a), varying the disc thickness. We simulated both the binary case and the single central object case, in order to provide a consistent reference for the accretion rate.

6.2.1 Initial conditions

Our initial conditions consist of two binary sink particles (able to accrete gas particles), and a finite circumbinary disc of $N_{\rm part} = 2 \times 10^6$ gas particles in most cases. Besides the fluid dynamical forces produced by viscosity and pressure, for which we refer to Lodato & Price (2010), the gas particles feel the acceleration produced by the sinks (e.g. Nixon et al. 2013). Each sink exerts on the i-th particle the acceleration $\boldsymbol{f}_{i,\rm pot}$

$$\boldsymbol{f}_{i,\rm pot} = \frac{GM_n(\boldsymbol{r}_n - \boldsymbol{r}_i)}{|\boldsymbol{r}_n - \boldsymbol{r}_i|^3}, \tag{6.3}$$

where $n = 1, 2$ indicates quantities related to the primary or the secondary object respectively, G is the universal constant of gravitation while $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are the positions of the two sink particles of mass M_1 and M_2 respectively. No smoothing to the potential has been applied as particles are considered accreted when the condition $|r_{1,2} - r_a| < r_{\rm sink}$ is satisfied, where the parameter $r_{\rm sink}$ is the sink radius, and their kinetic energy is not sufficient to escape the potential well (Bate et al. 1995). Note that, contrary to previous investigations, we do not prescribe the sink particles on fixed orbits: their motion is determined by the gravitational potential that one exerts on the other and by the back-reaction they receive from the interaction with the gas particles. We neglect the gas-gas gravitational interaction, i.e. no disc self-gravity.

We choose an equal mass ($q = 1$), circular (eccentricity $e = 0$) binary system in order to simplify the comparison with the literature. We use code units such that the binary orbital frequency $\Omega_{\rm bin} = 1$ (see Appendix D.4). The masses of the black holes in code units are thus $M_1 = M_2 = 0.5$, the binary separation is $a = 1$ and we set the initial velocities of the sinks to obtain circular Keplerian orbits, we use $r_{\rm sink} = 0.05$.

The gas disc is set up by placing particles between an inner radius $R_{\rm in} = 2.6a$ and an outer radius $R_{\rm out} = 5a$, where a is the binary separation, in order to obtain an initial surface density distribution of the type $\Sigma = \Sigma_0 R^{-p}$, where $p = 2$ and Σ_0 is chosen in order to have a disc mass $M_{\rm disc} = 0.005$, using an initial Monte Carlo particle placement. The initial vertical position of each particle is chosen from a gaussian distribution with standard deviation $H = c_{\rm s}/\Omega$, where $c_{\rm s}$ is the sound speed (see the next Section) and Ω the Keplerian frequency for a central mass $M_{\rm tot}$. The velocity of each particle is Keplerian corrected to account for pressure, given by

$$v_i^2 = \frac{GM_{\rm tot}}{R_i} - c_{{\rm s},i}^2 \left(p + \frac{3}{2} + \ell \right), \tag{6.4}$$

where R_i is the distance of the i-th particle from the centre of mass of the system, $c_{{\rm s},i}$ is the sound speed for the i-th particle and $\ell = 1/2$ is the power law index for the sound speed.

6.2.2 Equation of state and temperature profile

We to prescribe a locally isothermal equation of state, in order to keep the disc temperature constant in time through the entire length of the simulation

$$P = \frac{k_B T}{\mu m_p}\rho = c_{\rm s}^2 \rho, \tag{6.5}$$

where $c_{\rm s}^2$ is the squared sound speed of the gas. The temperature of the gas is then prescribed through the gas sound speed, for which we use the formulation (Farris et al. 2014)

$$c_{\rm s} = \frac{H}{R}\left(\frac{GM_1}{R_1} + \frac{GM_2}{R_2}\right)^{\ell}, \tag{6.6}$$

that implies a constant H/R throughout the disc. This prescription has the nice property of becoming a radial power law around each sink, since equation (6.6) reduces to

$$c_{\rm s} = \begin{cases} \dfrac{H}{R} v_{1,\rm K}, & \text{for } R_1 \ll R_2, \\ \dfrac{H}{R} v_{2,\rm K}, & \text{for } R_2 \ll R_1, \\ \dfrac{H}{R} v_{\rm K}, & \text{for } R_1 \sim R_2 \gg a, \end{cases} \tag{6.7}$$

where a is binary separation while $v_{1,\rm K}$, $v_{2,\rm K}$, $v_{\rm K}$ are the Keplerian velocity around the primary, the secondary and the binary objects respectively.

To compare different disc temperatures we perform simulations using seven different values of $H/R = \{0.13; 0.12; 0.1; 0.08; 0.06; 0.04; 0.02\}$. As mentioned above, the case $H/R = 0.1$ is the most used in recent literature (MacFadyen & Milosavljević 2008; D'Orazio et al. 2013; Farris et al. 2014; Shi & Krolik 2015) and will be useful for a comparison with previous results.

6.2.3 Viscosity

SPH employs an artificial viscosity term, in order to resolve shocks. This term acts as a source of viscous diffusion and therefore can be used to model the angular momentum transport in the disc (Lodato & Price 2010). However, some preliminary tests using artificial viscosity to model the disc viscosity showed an unwanted increase in the accretion rate on to the sinks. This occurs because the artificial shear viscosity $\nu_{\rm AV}$ is dependent on the density ρ of the fluid such that $\nu_{\rm AV} \propto h/H \propto \rho^{-1/3}$, where h is the SPH smoothing length. This implies that if the disc is characterized by strong density gradients, the viscous effects are subject to strong changes throughout the disc, affecting the reliability of the accretion rate. We therefore decide to introduce just the bare minimum amount of artificial viscosity, using the Morris & Monaghan (1997) switch with $\alpha_{\rm AV,min} = 0.1$, $\alpha_{\rm AV,max} = 0.5$ and $\beta_{\rm AV} = 2$ everywhere to prevent particle interpenetration.

To compute the "physical" viscosity we use the implementation of Navier-Stokes viscosity similar to that given by Flebbe et al. (1994), as described in Lodato & Price (2010) (see their section 3.2.4). We set the bulk viscosity $\zeta = 0$, while

ν is computed using an α-prescription that reduces to the standard Shakura & Sunyaev (1973) prescription around each sink:

$$\nu = \alpha_{\rm SS} c_{\rm s} \frac{H}{R} f, \tag{6.8}$$

$$f = \min\left(R_1, R_2\right), \tag{6.9}$$

where $c_{\rm s}$ is given by equation (6.6), R_1 and R_2 are the distances from M_1 and M_2 respectively; we set $\alpha_{\rm SS} = 0.1$. Our choice of $\alpha_{\rm SS}$ is equal to Farris et al. (2014) and is consistent with the equivalent value extrapolated from MHD simulations by Shi et al. (2012); Shi & Krolik (2015), although this result needs to be validated for thinner discs. Note that the main angular momentum transport process at $\approx 0.1\,$pc separation in an AGN disc might be associated with gravitational instabilities (Goodman 2003; Lodato 2012). For the thin discs in AGN, self-gravitating angular momentum transport is local and is expected to provide equivalent $\alpha \approx 0.1 - 0.3$, consistent with our choice (Cuadra et al. 2009).

Parameters employed for all the simulations in this study are listed in Table 6.1.

6.2.4 Resolution

Besides the simulations discussed in the next Sections (those denoted by S in Table 6.1), we performed a set of simulations (denoted by Reso in Table 6.1) varying $N_{\rm part}$ as a convergence test. The results for $N_{\rm part} = 1.0 \times 10^6$ converge to the higher resolution simulations. The scale height is well resolved throughout the disc for every H/R (see Table 6.1). However, the cavity region is obviously poorly resolved due to its low density. While in most cases the disc thickness is resolved even in the cavity, for $H/R = \{0.04; 0.02\}$ h/H can be much above unity ($h/H \sim \{6; 8\}$, respectively). This results in an increase of the artificial viscosity inside the cavity for $H/R = \{0.04; 0.02\}$.

We estimate this excess in the cavity region by computing the value of $\alpha_{\rm SS,AV}$ corresponding to the artificial viscosity using (Lodato & Price 2010)

$$\alpha_{\rm SS,AV} = \frac{1}{10} \alpha_{\rm AV,max} \frac{h}{H}, \tag{6.10}$$

through which we are able compare the magnitude of the artificial viscosity with respect to the "physical" one. The values of $\alpha_{\rm phys,AV}$ in the cavity region for the cases $H/R = \{0.04; 0.02\}$ are $\alpha_{\rm SS,AV} \sim 0.3$ and $\alpha_{\rm SS,AV} \sim 0.4$, respectively, to be compared to the value of $\alpha_{\rm SS} = 0.1$ that we prescribed for the physical viscosity. Thus, for the two thinnest cases the evolution of the gas within the cavity is dominated by numerical effects.

6.3 Results

The finite size of our discs causes spreading to larger radii as time passes, causing the accretion rate to vary. The evolution time-scale for accretion discs is given by

	H/R	$N_{\rm part}$	α	ν	$t_\nu/t_{\rm bin}$	$\langle h/H \rangle$	SO
S1	0.13	2.0×10^6	0.10	1.7×10^{-3}	701	0.09	yes
S2	0.12	2.0×10^6	0.10	1.4×10^{-3}	823	0.09	yes
S3	0.10	2.0×10^6	0.10	1.0×10^{-3}	1185	0.11	yes
S4	0.08	2.0×10^6	0.157	1.0×10^{-3}	1185	0.13	no
S5	0.08	2.0×10^6	0.10	6.4×10^{-4}	1851	0.13	yes
S6	0.06	2.0×10^6	0.177	6.4×10^{-4}	1851	0.15	no
S7	0.06	2.0×10^6	0.10	3.6×10^{-4}	3291	0.15	yes
S8	0.04	2.0×10^6	0.10	1.6×10^{-4}	7406	0.20	yes
S9	0.02	2.0×10^6	0.10	4.0×10^{-5}	29625	0.32	yes
Reso1	0.10	1.5×10^6	0.10	1.0×10^{-3}	1185	0.12	no
Reso2	0.10	1.0×10^6	0.10	1.0×10^{-3}	1185	0.14	no
Reso3	0.10	5.0×10^5	0.10	1.0×10^{-3}	1185	0.17	no
Reso4	0.08	1.0×10^6	0.10	6.4×10^{-4}	1851	0.16	no
Reso5	0.06	1.0×10^6	0.10	3.6×10^{-4}	3291	0.19	no
Reso6	0.04	1.0×10^6	0.10	1.6×10^{-4}	7406	0.25	no
Reso7	0.02	1.0×10^6	0.10	4.0×10^{-5}	29625	0.40	no
Reso8	0.02	5.0×10^5	0.10	4.0×10^{-5}	29625	0.51	no

Table 6.1: Summary of the simulations. For each simulation we give the aspect ratio H/R, the number of particles $N_{\rm part}$, the α viscous parameter, the shear parameter ν, the viscous time $t_\nu/t_{\rm bin}$ (equation 6.12) expressed in binary orbital periods units and the initial smoothing-length normalized to the disc thickness (disc average). The column SO is to indicate if the corresponding single object simulation has been performed. Notice that the value reported for ν is computed at $R_1 = R_2 = a$. For completeness, the parameters in each simulations are: inner disc radius $R_{\rm in} = 2.6$, outer disc radius $R_{\rm out} = 5$, total binary mass $M_{\rm tot} = 1$, disc mass $M_{\rm disc} = 0.005$, density power law index $p = 2$, sound speed power law index $\ell = 0.5$, binary mass ratio $q = 1$, and binary orbital eccentricity $e = 0$.

the viscous time t_ν

$$t_\nu = \frac{2}{3}\frac{R_{\rm out}^2}{\nu(R_{\rm out})} = \frac{5^{3/2}}{3\pi\alpha}\left(\frac{H}{R}\right)^{-2} t_{\rm bin} \tag{6.11}$$

$$\approx 11.85 \cdot \left(\frac{H}{R}\right)^{-2} t_{\rm bin}, \tag{6.12}$$

where $t_{\rm bin}$ is the binary orbital period

$$t_{\rm bin} = \frac{2\pi}{\Omega_{\rm bin}} = 2\pi\left[\frac{G(M_1+M_2)}{a^3}\right]^{-1/2}. \tag{6.13}$$

Any comparison between different regimes of disc thickness needs therefore to be done at the same t/t_ν. The time for the comparison has to be chosen to be long enough to allow the achievement of quasi-stationarity, but much smaller than t_ν in order to prevent excessive relaxation of the initial conditions. Indeed, as will be seen below, circumbinary disc properties and structure evolve differently as the time passes and a comparison at later times would not be reliable anymore.

We compare our simulations at times which are sufficiently long to overcome the initial transient and computationally tractable for the cases we simulated. However, where possible we let the discs evolve up $t \sim 0.2\,t_\nu$.

Due to its long viscous time, the $H/R = 0.02$ calculation reached just $t = 0.07\,t_\nu \sim 2000\,t_{\rm bin}$; however we believe that the results for this case provide a valid estimate for the accretion rate and for this reason should not be discarded.

6.3.1 Single object simulations

In addition to the simulations of binary systems, we have also performed seven reference simulations around a single object (marked by SO = yes in Table 6.1), one for each value of H/R used in the binary simulations.

We use the reference simulations to evaluate the accretion on the single central object $\dot{M}_0$ and thus renormalize the results obtained for the binary case. The initial setup of the disc of each reference run is the same as that described for the binary case in Section 6.2.1. The only differences are the presence of a Keplerian potential produced by a mass $M = M_1 + M_2$ in the centre of mass of the system and a sink radius $R_{\rm sink,SO} = 1$. The number of particles used was $N = 2 \times 10^6$.

In the upper panel of Figure 6.1 the results for the accretion rate on to the single central object $\dot{M}_0$ (in code units) as a function of t/t_ν are plotted: the evolution of the time-varying accretion rate occurs on a viscous time t_ν. In the lower panel of Figure 6.1, the accretion rate is renormalized to $\dot{M}_{\rm scale}$ given by

$$\dot{M}_{\rm scale} = 3\pi\nu(R_{\rm out})\Sigma(R_{\rm out}) \propto (H/R)^2, \tag{6.14}$$

which gives an analytical estimate of the order of magnitude of the accretion rate. The accretion rate is obtained computing how many particles are accreted at $R_{\rm sink,SO}$.

The accretion rates plotted in Figure 6.1 are characterized by an initial transient at the time $t \approx 0.025\,t_\nu$ for each H/R. After that, the disc reaches quasi-stationarity and the accretion rate lowers as the time passes due to the spread of the disc toward larger radii.

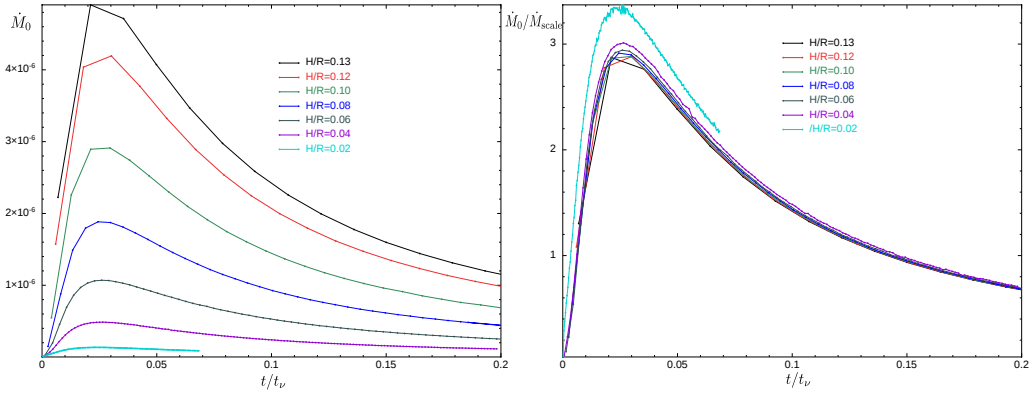


Figure 6.1: Left panel: absolute accretion rate $\dot{M}_0$ around a single object in code units, averaged over $10\,t_{\rm dyn}$, at the sink radius $R = R_{\rm sink} = 1$, as a function of t/t_ν. The different colours represents the seven cases of $H/R = \{0.13; 0.12; 0.10; 0.08; 0.06; 0.04; 0.02\}$ (black, red, green, blue, grey, purple, cyan lines, respectively) we studied. Right panel: $\dot{M}_0/\dot{M}_{\rm scale}$, where $\dot{M}_{\rm scale}$ normalization constant given by equation (6.14), again as a function of t/t_ν. It should be noticed that the absolute value of the accretion rate decreases, for lowering thicknesses, as $(H/R)^2$. In particular, from the top panel it can be easily noticed a difference of a factor ~ 25, going from $H/R = 0.1 \to 0.02$.

The lower panel of Figure 6.1 shows that the accretion rates from our simulations scale as predicted by theory. The only exception is our thinnest case with $H/R = 0.02$ that overestimates the accretion rate with respect to $\dot{M}_{\rm scale}$. Due to the higher value of $\langle h/H \rangle$ throughout the disc, the artificial viscosity gives a non-negligible contribution with respect to the physical one resulting in a slight overestimate of the accretion rate in this case. This effect is less visible but already present in the other regimes. It should be noticed that the lowest curve is the thickest case $H/R = 0.13$ (that shows the lowest $\langle h/H \rangle$), while $\dot{M}_0/\dot{M}_{\rm scale}$ progressively grows as $\langle h/H \rangle$ increases for thinner regimes.

The lower panel in Figure 6.1 confirms that the outcome of the simulations for $\dot{M}_0$ are consistent with the analytical predictions both for the scaling, i.e. $\dot{M}_0 \propto (H/R)^2$, and for the magnitude

$$\dot{M}_{0,\rm theor} \approx 3\pi\nu\Sigma. \tag{6.15}$$

6.3.2 Binary simulations

Figure 6.2 shows column density in our binary systems after $t \sim 0.2t_\nu$ for $H/R = \{0.12; 0.1; 0.08; 0.06; 0.04\}$ and at time $t \sim 0.07t_\nu$ for the case $H/R = 0.02$. We observe the formation of circumprimary and circumsecondary discs around each sink for the cases $H/R = \{0.12; 0.1; 0.08; 0.06\}$ but not for the cases $H/R = \{0.04; 0.02\}$. This is probably a numerical effect related to the previously discussed increase of the ratio h/H inside the cavity, which results in an unwanted excess of the artificial viscosity in the two thinnest cases.

The size of the cavity is larger for low H/R. This is expected. The criteria for gap opening state that, for lower H/R, the gap can be opened by progressively

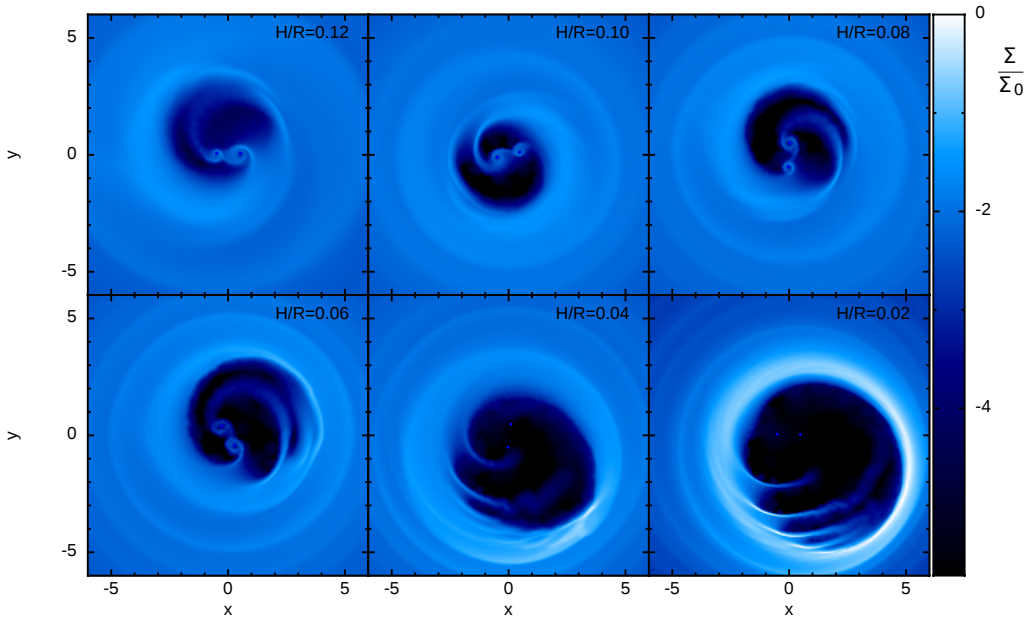


Figure 6.2: Colour plots of the column density normalized to Σ_0 of our simulations (logarithmic scale). From left on the top row $H/R = 0.12; 0.1; 0.08$, bottom row $H/R = 0.06; 0.04; 0.02$. Snapshots were taken at times $t \sim 0.2t_\nu$ in any case, except $H/R = 0.02$ which was taken at time $t \sim 0.07t_\nu$.

less massive binary companions (Lin & Papaloizou 1993; Crida et al. 2006). This occurs because the viscous torques, that are responsible for driving the gas inward, scale as $T_{\rm visc} \propto \nu \propto (H/R)^2$, while the tidal torques, responsible for the "dam" effect at the edge of the cavity, scale as $T_{\rm tid} \propto (H/R)^{-3}$. The radius at which the two contrasting torque contributions are equal approximately gives the truncation radius of the disc (for a more accurate discussion see Artymowicz & Lubow 1994).

Figure 6.2 also shows that each simulation has developed an eccentric cavity, characterized by an overdense lump of material at its edge. This effect has been widely observed in the literature (Papaloizou et al. 2001; D'Angelo et al. 2006; Kley & Dirksen 2006; MacFadyen & Milosavljević 2008; D'Orazio et al. 2013; Farris et al. 2014; Dunhill et al. 2015), interpreted as being due to the unstable growth of spontaneous deviations of the gas from circular motion, starting from super-hump theory (Lubow 1991a,b). This interpretation relates to the size of the cavity and, more specifically, to whether the resonances believed to damp this instability fall in the disc region or not, as suggested by Papaloizou et al. (2001). However, we caution that even though this interpretation applies to low mass ratios ($10^{-3} \leq q \leq 3 \cdot 10^{-2}$ Papaloizou et al. 2001; D'Angelo et al. 2006; Kley & Dirksen 2006), the extension of the model to higher companion masses is not straightforward and still needs further investigation. D'Orazio et al. (2016) pointed out that a transition between circular and lopsided discs occurs for mass ratios $q > 0.04$ as a consequence of the loss of stable orbits across the corotation region in the restricted three body problem (orbits around L4 and L5 Lagrangian points) in that range of masses, and other viscous effects.

It can be also noticed that the prominence of the lump appears to be influenced

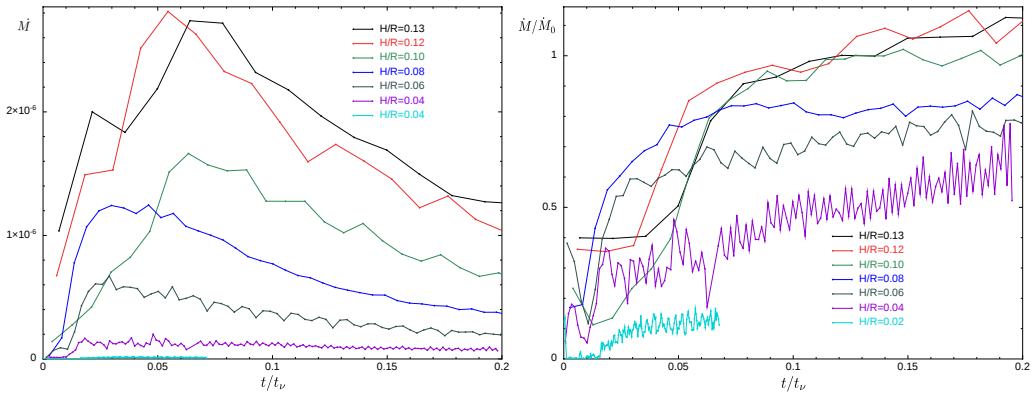


Figure 6.3: Left panel: y-axis accretion rate in code units, each point in the plot obtained averaging $\dot{M}_{\rm bin}$ over $10t_{\rm bin}$, as a function of $t/t_{\rm bin}$. Right panel: $\dot{M}_{\rm bin}/\dot{M}_0$ as a function of $t/t_{\rm bin}$. Different colours refers to different H/R regimes, in particular $H/R = \{0.13; 0.12; 0.1; 0.08; 0.06; 0.04; 0.02\}$ black, red, green, blue, grey, purple, cyan lines respectiely.

by the disc thickness, in particular the overdensity is more marked when the disc is thinner.

6.3.3 Suppression of accretion for thin discs

Figure 6.3 shows the accretion rate across the cavity edge $\dot{M}_{\rm bin}$ for our binary simulations compared to the corresponding value obtained in Section 6.3.1 for a single black hole. The left panel shows $\dot{M}_{\rm bin}$ versus t/t_ν for the various disc thicknesses. As H/R is reduced, $\dot{M}_{\rm bin}$ drops significantly. A reduction in $\dot{M}$ with thickness is expected independently of the presence of the binary (upper panel of Figure 6.1). However for our α-model, we expect that $\dot{M} \propto (H/R)^2$. The right panel of Figure 6.3 shows that the suppression of the mass accretion rate is much stronger than this. The right panel of Figure 6.3 shows the ratio $\dot{M}_{\rm bin}/\dot{M}_0$, and we recall that we have demonstrated (lower panel of Figure 6.1) that $\dot{M}_0 \propto (H/R)^2$. This represents the fraction of the unperturbed gas flow that makes it into the cavity. While for $H/R \sim 0.1$ we recover the known result that $\dot{M}_{\rm bin} \sim \dot{M}_0$, implying that the binary does not prevent matter from accreting, the situation changes drastically for lower H/R. For $H/R = 0.02$, for example, only ~ 15 per cent of the unperturbed mass makes it into the cavity. We thus conclude that

$$\dot{M}_{\rm bin} = \xi(H/R)\dot{M}_0, \tag{6.16}$$

where ξ is a function of H/R. For the thicker cases the value of $\xi(H/R)$ appears to saturate at ~ 1 for growing H/R. This is in contrast with the results obtained by Farris et al. (2014) and Shi & Krolik (2015) that found values of ξ above unity ($\xi \sim 1.6$ Farris et al. 2014, $\xi \sim 1.4$ Shi & Krolik 2015) for a unitary mass ratio binary system with $H/R = 0.1$, but in agreement with D'Orazio et al. (2013) that found $\xi = 1.015$ for the case of interest.

While for large H/R the value of $\dot{M}_{\rm bin}/\dot{M}_0$ appears to reach a well defined asymptote, for lower H/R it increases with time. This deserves a brief discussion. In our thinnest cases, the binary accretes much less than the corresponding $q = 0$

H/R	0.13	0.12	0.10	0.08	0.06	0.04	0.02
$\xi(H/R)$	0.97	0.99	0.95	0.82	0.70	0.47	0.18

Table 6.2: Values of $\xi(H/R)$ taken averaging $\xi(H/R)$ between $0.07\,t_\nu < t < 0.15\,t_\nu$ ($H/R = \{0.13; 0.12; 0.10; 0.08; 0.06; 0.04\}$). The value of ξ for the case $H/R = 0.02$ is obtained instead averaging between $0.05\,t_\nu < t < 0.07\,t_\nu$.

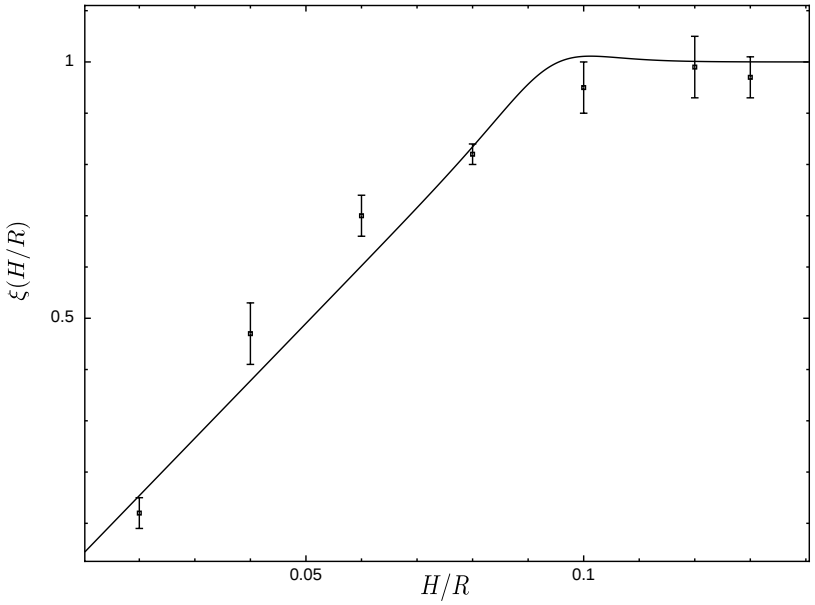


Figure 6.4: Plot of $\xi(H/R)$ as a function of H/R, with the values in Table 6.2. The error bars represent the square root of the mean quadratic error of the values assumed by $\xi(H/R)$ for times between $0.07\,t_\nu < t < 0.15\,t_\nu$ (for the case $H/R = 0.02$, $0.05\,t_\nu < t < 0.07\,t_\nu$). The black line plots an interpolating function such as equation (6.17).

discs, accumulating some material at the edge of the cavity. This translates into a slower time evolution of the accretion rate which remains almost constant after the peak (see purple and cyan lines in the left panel of Figure 6.3), while the value of $\dot{M}_0$ keeps decreasing; as a consequence, the ratio $\dot{M}_{\rm bin}/\dot{M}_0$ increases with time. For this reason the value of $\dot{M}_{\rm bin}/\dot{M}_0$ is most reliable when the discs have evolved for enough time to overcome the initial transient, but not so much that the viscous evolution of the system modifies the disc structure[2].

The left panel of Figure 6.3 highlights another important feature. The thickest cases ($H/R = \{0.13; 0.12; 0.10\}$) show a time-shift in the initial peak. This feature is likely related to our initial conditions for the velocity field of the gas, equation (6.4). In the presence of a binary potential our choice underestimates the equilibrium velocity of the gas. The thickest regimes, in which viscous forces are stronger, result then in a less steep initial transient, that reaches a lower maximum. However, this shift does not affect the evolution at later times.

Table 6.2 shows the time average of ξ for each value of the disc thickness, where the average has been taken after the initial transient but before viscous

[2]Recall that our discs do not reach a steady-state since they have a finite mass and radius.

evolution has significantly affected our results. In particular, we generally average between $0.07t_\nu$ and $0.15t_\nu$, except for $H/R = 0.02$, for which we average between $0.05 - 0.07t_\nu$.

Figure 6.4 shows the average ξ as a function of H/R, where the error bars are the square root of the mean quadratic error in the average procedure. It appears that ξ increases linearly with H/R for $H/R < 0.1$, saturating at around unity for $H/R > 0.1$. We thus write

$$\xi(H/R) \approx \begin{cases} 10 \cdot H/R, & \text{for } H/R \leq 0.1 \\ 1, & \text{for } H/R > 0.1 \end{cases} . \tag{6.17}$$

The black line in Figure 6.4 shows an interpolating function with a linear behaviour for $H/R < 0.1$ and constant $\xi = 1$ for $H/R > 0.1$.

The typical values of H/R in AGN discs are believed to span $H/R \sim 10^{-2} - 10^{-3}$ (Shakura & Sunyaev 1973; Collin-Souffrin & Dumont 1990; Natarajan & Pringle 1998; Goodman 2003; Lodato 2012; Gerosa et al. 2015) depending on the region of the disc examined and on the degree of self-gravity in the disc (Haiman et al. 2009b). If the law in equation (6.17) keeps holding also for these physical values of H/R, we expect to have $\xi \sim 0.1 - 0.01$, implying that the accretion rate is suppressed up to a factor 10^2 with respect to the normal AGN activity.

A similar effect has been recently found during the last phases of supermassive black hole binary mergers (Cerioli et al. 2016). In that case tidal torques acts as barriers for the gas during the late gravitational inspiral phase before the black hole merger: the binary companion squeezes the gas of the thin circumprimary disc toward the primary object during its orbital decay, causing an enhancement in the accretion rate and thus predicting a flare in the luminosity of the system just before the binary merger (in contrast, Baruteau et al. 2012 found that this effect does not occur for thicker discs).

6.3.4 Accretion Variability as a function of the disc temperature

In this Section we analyze the fast variability of the mass flow across the cavity edge, $\dot{M}_{\rm bin}$, and the accretion rate as computed directly from accretion on to the binary, $\dot{M}_{\rm sink}$.

Figure 6.5 shows $\dot{M}_{\rm sink}$ (red line) and $\dot{M}_{\rm bin}$ (black line). Starting from the top left panel in Figure 6.5 ($H/R = 0.1$), we notice that the values of $\dot{M}_{\rm sink}$ are much smoother than those obtained for $\dot{M}_{\rm bin}$: the formation of circumprimary and circumsecondary discs (see Fig. 6.2) acts as a buffer for the accretion mechanism, accumulating material and accreting it progressively, smoothing the variability observed instead for the mass flux at $R = a$. This buffering effect was observed by Farris et al. (2014), who studied the accretion rate in binary systems as a function of the binary mass ratio; we notice also that the qualitative behaviour of the variability of our $H/R = 0.1$ case is in very good agreement with their case $q = 1$, $H/R = 0.1$ (top panel of their Figure 10).

From the other panels in Figure 6.5 it can be noticed that the buffering effect progressively disappears as H/R is reduced: in the case $H/R = 0.04$ (bottom right panel in Figure 6.5) no circumprimary and circumsecondary discs form, and $\dot{M}_{\rm sink}$ follows the variability of $\dot{M}_{\rm bin}$.

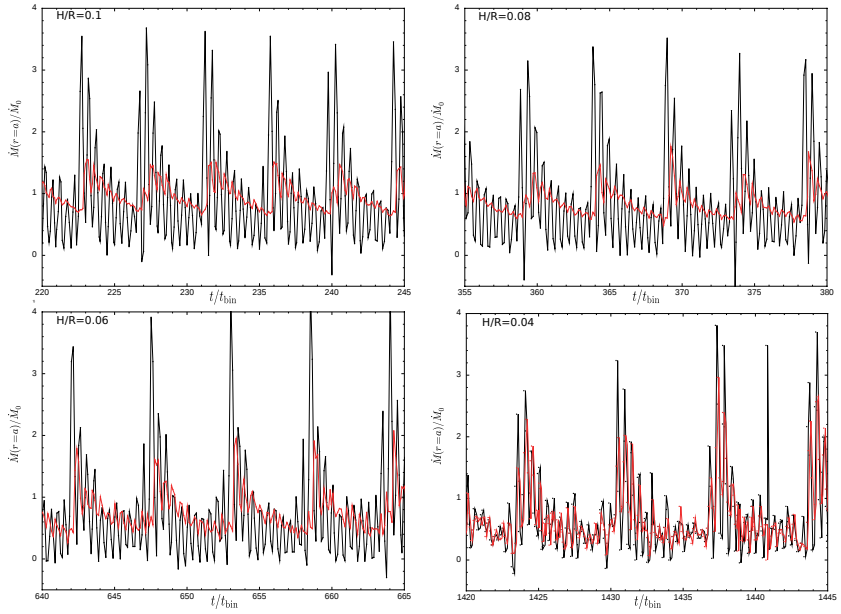


Figure 6.5: From top-left to bottom right: $H/R = \{0.1; 0.08; 0.06; 0.04\}$ respectively. Black line: Time variability of the mass flux across the radius $R = a$, $\dot{M}_{\rm bin}/M_0$ (normalized to the corresponding averaged $\dot{M}_0$), and red line: accretion rate on to the binary $\dot{M}_{\rm sink}/\dot{M}_0$. The panels show the various regimes at $t \sim 0.2t_\nu$.

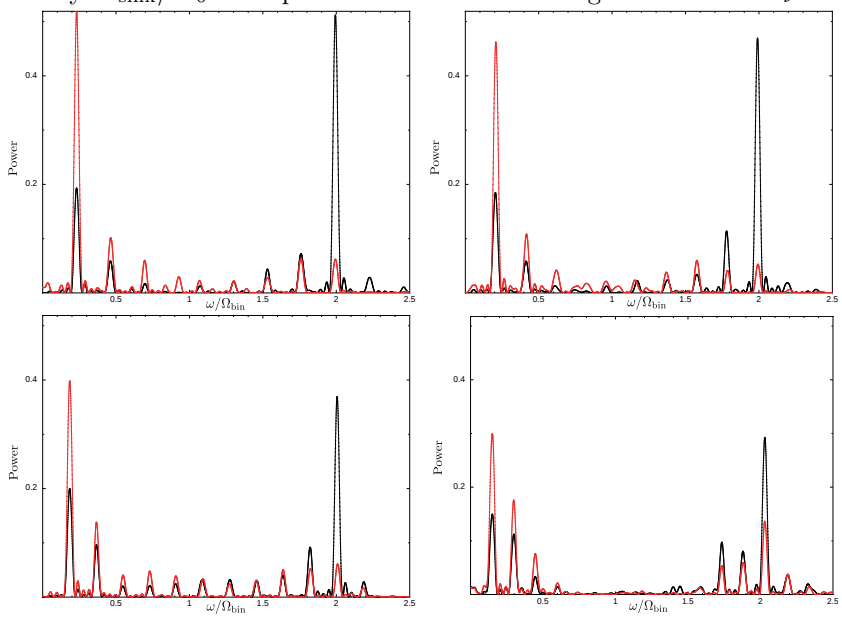


Figure 6.6: From top-left to bottom right: periodograms of the accretion rates for the cases $H/R = \{0.1; 0.08; 0.06; 0.04\}$, respectively, relative to the time intervals plotted in fig. 6.5. As in Figure 6.5, black line refer to $\dot{M}_{\rm bin}/\dot{M}_0$; red line instead is $\dot{M}_{\rm sink}/\dot{M}_0$. The frequencies reported are expressed in $\omega/\Omega_{\rm bin}$ with $\omega = 2\pi t^{-1}$ and $\Omega_{\rm bin} = 2\pi t_{\rm bin}^{-1}$.

Figure 6.6 shows the frequency analysis of the accretion rates for each H/R plotted in Figure 6.5. Red and black lines show, as previously, $\dot{M}_{\rm sink}$ and $\dot{M}_{\rm bin}$, respectively. These periodograms were obtained using Lomb-Scargle analysis on the time interval considered in Figure 6.5. Periodograms of $\dot{M}_{\rm sink}$ and $\dot{M}_{\rm bin}$ show that, in thick discs, even though the mass flow $\dot{M}_{\rm bin}$ is dominated by the frequency $2\Omega_{\rm bin}$, the accretion on to the black holes $\dot{M}_{\rm sink}$ occurs with lower periodicity, because of the accumulation of the gas into the discs.

Lowering H/R, the absence of the circumprimary and circumsecondary discs causes the power of $\dot{M}_{\rm sink}$ and $\dot{M}_{\rm bin}$ to progressively equalize at each frequency, since all the mass that enters in the cavity is accreted by the black holes faster than they are fed from the edge of the cavity. The fundamental frequency of the accretion rate returns to be $\omega = 2\Omega_{\rm bin}$ in thin discs.

It should be finally noticed that the lowest frequency of the periodogram diminishes for low H/R: this frequency represents the orbital frequency of the density lump one can observe in Figure 6.2 at the edge of the cavity; the passage of this lump at the pericentre of its orbit generates a boost in the accretion with the same periodicity of the edge of the cavity $\omega_{\rm lump} \sim (GM_{\rm tot}/R^3_{\rm lump})^{-1/2}$ where $R_{\rm lump}$ is the semi-major axis of the cavity, that results in a peak in the periodogram. Since the cavity becomes larger for low H/R, we observe a shift of this peak to lower frequencies for decreasing H/R.

Caution is required in interpreting these results. The variability of the $H/R = 0.04$ case is likely affected by the low resolution in the cavity, and the lack of circumprimary and circumsecondary discs may be attributed to the enhanced artificial viscosity in this region. However, some physical considerations also apply. Due to the low accretion rate from the edge of the cavity, circumprimary and circumsecondary discs are expected to become sparser and characterized by lower accretion rates when reducing the disc thickness, implying circumprimary and circumsecondary discs to be fainter than predicted using thicker discs (Farris et al. 2015a) and thus lower luminosities at short wavelengths.

6.3.5 The dependence of ξ on viscosity

Both viscosity and disc temperature are functions of the disc thickness. Performing simulations exclusively for different H/R does not let us disentangle whether the observed suppression of the accretion rate is due to viscous or pressure effects. For this reason we performed two additional simulations, S4 and S6 (in Table 6.1), with $H/R = \{0.08; 0.06\}$ and values of α set in order to obtain the same value of ν achieved in S3 and S5 with $\alpha = 0.1$ and $H/R = \{0.1; 0.08\}$, respectively.

The accretion rates of S4 and S6 are consistent with those obtained for S3 and S5, respectively, showing that the accretion rate does not change for the same ν despite a change in the value of H/R. Since α is independent of H/R, this suggests that the reduction of the accretion rate for low disc aspect-ratios is due to the reduction of the effective disc viscosity rather than to a variation of the gas pressure. This is consistent with the prediction of D'Orazio et al. (2016) regarding the modifications induced by pressure to the effective gravitational potential: pressure effects allow the gas to overcome gravitationally prohibited regions only for $H/R \gtrsim 0.1$.

However, the effects of the gas pressure are important for the gas inside the cavity, e.g. for the differential accretion rate (Young & Clarke 2015; Young et al. 2015).

6.4 Summary and conclusions

We performed a set of SPH simulations with a circumbinary disc around an equal mass circular binary, varying the disc aspect ratio H/R, in order to investigate the dependence of the accretion rate on the disc thickness.

We compared our simulations to a set of reference simulations with a single central object. We computed the accretion rate at the edge of the cavity $\dot{M}_{\rm bin}$ in the binary case and compared it to the reference one, $\dot{M}_0$, from the single central object simulations. Our results for our $H/R = 0.1$ case agree with recent literature, in particular they are consistent with D'Orazio et al. (2013), who obtained $\dot{M}_{\rm bin}/\dot{M}_0 \sim 1$; we also verified the results of Farris et al. (2014) regarding accretion variability and periodicity inside the cavity (although their value $\dot{M}_{\rm bin}/\dot{M}_0$ is slightly larger than ours).

While for $H/R \gtrsim 0.1$ the accretion rate in a circular equal mass binary system is $\dot{M}_{\rm bin}/\dot{M}_0 \sim 1$, for $H/R < 0.1$ a linear reduction $\dot{M}_{\rm bin}/\dot{M}_0 \sim 10H/R$ was observed. These results are summarized in Table 6.2 and Figure 6.4, reporting the values of $\xi(H/R) = \dot{M}_{\rm bin}/\dot{M}_0$.

These results have consequences for both the detection and evolution of supermassive black hole binary systems. If the relationship for the reduction in accretion rate as a function of disc thickness in equation (6.17) holds also for the thin discs expected to surround supermassive black holes ($H/R = 10^{-2} - 10^{-3}$), the accretion rate in supermassive black hole binaries $\dot{M}_{\rm bin}$ would be reduced of up to a factor 10^2 with respect to the equivalent rate on a single object. The low accretion rate implies that these systems are much fainter than normal AGN.

While our simulations do not resolve the individual circumprimary and circumsecondary discs for $H/R = \{0.02; 0.04\}$, we do expect that a reduction in $\dot{M}_{\rm bin}$ implies the formation of lower mass, lower density and lower luminosity minidiscs. This is expected to affect the spectral energy distribution from this kind of systems, since the hottest gas regions may be fainter than so far predicted; in particular reducing the short-wavelength contribution to the continuum spectrum.

Our findings also have important consequences for black hole spin alignment during the merger of a binary black hole system. Firstly, the model of Gerosa et al. (2015) implies that as the accretion rate into the cavity is suppressed, the alignment process of each black hole with its disc is slowed down accordingly. Secondly, with low accretion rates on to the binary the equalization timescale (that is, the timescale required for the two black holes to equalize their mass) is consequently increased, thus justifying a posteriori the assumption of $q = const$, made by Gerosa et al. (2015). We thus predict that the efficiency of spin alignment during merger is significantly reduced, which will affect the wave form of gravitational waves emitted during the last phases of the merging process and cause a high recoil velocity of the black hole formed after coalescence.

Low accretion rates on to the black hole binary appear to be important to

guarantee the secondary black hole reaches the binary separation at which gravitational waves emissions become dominant for the migration toward the the binary merger (Young & Clarke 2015). High accretion rates with a finite mass supply might cause the mass reservoir in the circumbinary disc to be exhausted before the binary reaches the gravitational wave inspiral phase, preventing further angular momentum extraction from the binary and thus implying the stall of the migration process. In this context the pileup of material at the cavity edge due to the suppression of accretion provides a stronger binary-disc coupling and, as a consequence, a more effective delivery of angular momentum to the gaseous disc (Rafikov 2013), shortening the migration time.

We emphasize that our results are essentially scale-free and can thus be extended to all black hole binary systems, providing predictions of accretion-dynamics in any mass regime. Equal mass, circular binaries with stellar mass black holes (such as those simulated here) are of particular interest due to the recent discovery of the gravitational wave source GW150914 (Abbott et al. 2016b). Some suggestion, although uncertain, for the occurrence of an electromagnetic counterpart to the gravitational wave emission in this particular source has also been made (e.g., Connaughton et al. 2016). Such electromagnetic counterparts are most naturally determined by gas accretion prior to, during and after the merger. Our results emphasize the important role that the disc thickness has in determining the tidal torques ability to act as a dam for the gas, which on the one hand might prevent gas flow from the circumbinary disc on to the binary, as discussed here, while on the other hand it might prevent gas flow from the individual disc to the circumbinary environment, as the gas is squeezed during the gravitational wave driven inspiral (Cerioli et al. 2016).

We finally note that in our simulations we do not inject material steadily at large radii as instead D'Orazio et al. (2013) and Farris et al. (2014) did. A constant inflow of material from large radii at a rate faster than the binary accretion rate would cause the progressive accumulation of material at the edge of the cavity. As a consequence, the viscous torque is expected to grow, possibly causing the material to overcome more easily the tidal "dam" effect provided by the binary.

Possible improvements to this work consist primarily of extending the parameter space under investigation, in particular considering the case of binary with eccentric orbits and exploring non-unitary mass ratios. Different mass ratios would provide new data for the differential accretion rates of the individual black hole, on which the model for spin alignment of Gerosa et al. (2015) is based.

Secondly, a broader range of H/R values should be explored. However, while exploring the $H/R > 0.13$ regimes would be straightforward, studying aspect ratios $H/R < 0.02$ would be more challenging since both spatial and accretion resolution were not well achieved already in our $H/R = 0.02$ case. A possible solution could be to restrict the area of the study to the region of the cavity providing a constant mass flux from outer radii as done for differential accretion rate studies Bate & Bonnell (1997); Ochi et al. (2005); Hanawa et al. (2010); Young et al. (2015). Finally, while here we discuss only the case of discs that are aligned with the orbital plane of the binary, it would also be interesting to study the problem of misaligned circumbinary discs (Lubow et al. 2015).

Part III

Disc-satellite interaction in young stellar objects

Chapter 7

Secular eccentricity evolution during disc-planet interaction

*Based on the paper by **Enrico Ragusa**, Giovanni Rosotti, Jean Teyssandier, Richard Booth, Cathie J. Clarke, & Giuseppe Lodato (2018), "Eccentricity evolution during planet-disc interaction", Monthly Notices of the Royal Astronomical Society, 474, 4460*

Ἄνδρα μοι ἔννεπε, Μοῦσα, πολύτροπον, ὅς μάλα πολλὰ
πλάγχθη, ἐπεὶ Τροίης ἱερὸν πτολίεθρον ἔπερσε...

Homer, Odissey Proemium

7.1 Evolution of planetary and disc eccentricity

The discovery of a large number of extrasolar planets has shown that the average orbital eccentricity of planets in the Galaxy is higher than that observed in our solar system (Butler et al. 2006). Two possible scenarios have been proposed during the past three decades in order to understand the origin of the orbital eccentricities observed in exoplanets. The first involves the interaction with other massive bodies in the system after the disc dispersal, in fact in a gas poor environment: for example the action of the Kozai-Lidov mechanism in the presence of massive planetary companions or a binary star companion (Naoz 2016 and references therein) or planet-planet scattering (Rasio & Ford 1996; Papaloizou & Terquem 2001; Ford & Rasio 2008; Jurić & Tremaine 2008; Mustill et al. 2017). The second involves the interaction of the planet at resonant locations with the protoplanetary disc in which it has formed (see Kley & Nelson 2012 for a review).

In this second scenario, Lindblad resonances pump the planet eccentricity, while eccentric co-rotation resonances and co-orbital Lindblad ones damp it (Goldreich & Tremaine 1980; Goldreich & Sari 2003). For a planet embedded in the disc the effectiveness of co-orbital Lindblad resonances in damping the eccentricity exceeds the pumping action of Lindblad resonances, implying that the disc-planet interaction tends to circularize the planet orbits (Cresswell et al. 2007; Bitsch & Kley 2010).

Nevertheless, gas depletion in the corotation region produced by a sufficiently massive planet (typically $M_{\rm p} > {\rm M_J}$) might lead to the growth of the eccentricity. Indeed, if the planet carve a gap which is deep enough to ensure that the disc torque on the planet is dominated by the contribution of the outer Lindblad resonance 1:3 the planet eccentricity is expected to grow Artymowicz et al. (1991).

Planets with masses as small as $M_{\rm p} \gtrsim 1\,{\rm M_J}$ have been found to be able to produce a saturation of co-orbital and corotation torque allowing the growth of the planet eccentricity provided the initial eccentricity $e_{p,0} > 0.01$ (D'Angelo et al. 2006; Duffell & Chiang 2015). These findings are in line with the theoretical production of Ogilvie & Lubow (2003) and Goldreich & Sari (2003). However, numerical simulations with massive planets ($M_{\rm p} \gtrsim 5 - 10\,{\rm M_J}$) have been observed to develop large eccentricity values also with initially circular orbits (Papaloizou et al. 2001; Dunhill et al. 2013).

The excitation mechanisms apparently stop when the planet eccentricity reaches values comparable to the disc aspect-ratio ($e \sim H/R$) for two main reasons (Duffell & Chiang 2015): first, because the epicyclic motion becomes increasingly supersonic for growing eccentricity, implying a weakening of Lindblad resonances responsible for the eccentricity pumping (Papaloizou & Larwood 2000); second, because if the eccentricity is sufficiently high, the planet hits the cavity walls rapidly damping the eccentricity. The mass of the disc and the density profile have also been shown to play a role in determining whether the planet eccentricity will grow. In particular, Dunhill et al. (2013) found that, for sufficiently massive companions ($M_{\rm p} = 25\,M_{\rm J}$), the eccentricity grows when the mass ratio between the companion and the disc is above a certain threshold and the density profile is such that the 1:3 Lindblad resonance dominates the overall torque exerted by the disc on the companion.

The general conclusion of most of these works is that the disc-planet interaction is not able to provide planet eccentricity growth above the value $e_{\rm p} \gtrsim 0.15$ (D'Angelo et al. 2006; Müller & Kley 2013; Duffell & Chiang 2015; Thun et al. 2017). However, Papaloizou et al. (2001) found for masses $M_{\rm p} \gtrsim 20 M_{\rm J}$ that the companion eccentricity might reach values of up to $e_{\rm p} \approx 0.25$.

The exchange of angular momentum between the disc and the planet causes a growth also in the disc eccentricity (Goldreich & Tremaine 1981), even when the planet has a circular orbit (Papaloizou et al. 2001;Kley & Dirksen 2006; Teyssandier & Ogilvie 2016; Teyssandier & Ogilvie 2017). The disc reacts to the presence of a planet producing an eccentricity profile decreasing with radius. We report that the growth of the disc eccentricity has been proposed as a possible explanation of non-axisymmetric features (Ataiee et al. 2013; Ragusa et al. 2017) observed in a large number of transition discs (Casassus 2016, for a review)

as an alternative scenario to the widely invoked vortex hypothesis (Regály et al. 2012; Ataiee et al. 2013; Lyra & Lin 2013). For completeness, numerical simulations in the context of binary black hole mergers have also revealed the formation of eccentric cavities with higher secondary-to-primary mass ratios (Armitage & Natarajan 2005; Shi et al. 2012; D'Orazio et al. 2013; Farris et al. 2014; D'Orazio et al. 2016; Ragusa et al. 2016), with important consequences for the modulation of the accretion rate.

Teyssandier & Ogilvie (2016) studied the normal modes solutions to the analytical equations ruling the eccentricity evolution in discs, thus making predictions about the disc eccentricity radial profiles in the presence of a planet.

It is important to notice that, mostly on account of the high computational cost of these simulations, in the aforementioned works the evolution of the eccentricity has never been explored beyond $t \gtrsim 2 \times 10^4$ planet orbits (Thun et al. 2017).

Motivated by the recent observation of CI Tau by Johns-Krull et al. (2016), Rosotti et al. (2017) performed long timescales calculations ($\sim 10^5$ orbits) in order to study the role of disc-planet interaction in exciting hot-Jupiters' eccentricity. Their simulations showed very prominent secular oscillations of the eccentricity with periodicities $\gtrsim 10^4$ orbits, superimposed on a roughly linear growth starting at a time of $\sim 4 \times 10^4$ orbits (doubling the eccentricity from ~ 0.04 over $\sim 10^5$ orbits) after an apparent stalling of the eccentricity evolution. Similar oscillations have also been observed in other works (Duffell & Chiang 2015; Müller & Kley 2013; Bitsch et al. 2013b; Dunhill et al. 2013; Miranda et al. 2017; Thun et al. 2017).

Even though the results in Rosotti et al. (2017) were not able to prove that disc-planet interaction might provide an effective mechanism to excite the eccentricities observed in hot-Jupiters, they showed clearly that the fate of the planetary eccentricity at late times cannot be determined a priori without performing simulations that cover significant fraction of the entire life time of the system.

In this chapter we present two long term numerical simulations of the disc-planet evolution for two different disc masses. The lower mass simulation is the same presented in Rosotti et al. (2017) but integrated three times longer ($\sim 3 \times 10^5$ orbits); the other has a disc mass that is a factor of three higher and is integrated for a similarly long time. We show that the initial behaviour of the planet eccentricity can be completely reversed at late times. Then we will give a physical interpretation of the peculiar evolution using a simplified toy model.

This chapter is structured as follows. In section 7.2 we present the numerical setup we used for our simulations. In section 7.3 we present the results of our simulations. In section 7.4 we discuss the results and introduce a simplified toy model to describe the evolution of the eccentricity. Section 7.5 provides an interpretation of the results in terms of the simplified toy model. In section 7.6 we draw our conclusions.

7.2 Numerical simulations

We perform two long timescale ($t \sim 3 \times 10^5$ orbits) 2D hydrodynamical simulations of a planet embedded in a gaseous disc orbiting a central star using FARGO3D

(Benítez-Llambay & Masset 2016) for two different disc masses. The simulations were run on GPUs (Nvidia Tesla K20), for a total wall clock time exceeding 6 months. We use open boundary conditions at the inner edge of the computational domain and closed at the outer one. We use a polar grid composed by $n_r = 430$ radial cells between $R_{\rm in} = 0.2$ and $R_{\rm out} = 15$ with logarithmic spacing, and $n_\phi = 580$ azimuthal cells. The outer radius $R_{\rm out}$ of the domain has been chosen to be sufficiently large to prevent the boundary conditions from affecting the dynamics; the propagation of eccentric perturbations does not reach radii $R \gtrsim 10$ (this can be noticed in Fig. 7.4, which will be discussed in the following sections). We use units in which $a_{\rm p,0} = 1$, where $a_{\rm p,0}$ is the initial semi-major axis of the planet, and $GM_* = 1$, where $M_\star$ is the mass of the star and G the gravitational constant. The simulations cover a time of $t = 3 \times 10^5 t_{\rm orb}$, where $t_{\rm orb} = 2\pi\Omega_{\rm p}^{-1}$ is the initial planet orbital period ($\Omega_{\rm p}$ Keplerian orbital frequency).

We vary the disc mass while keeping fixed the other simulation parameters. One simulation uses a disc-to-planet mass ratio $q = M_{\rm d}/M_{\rm p} = 0.2$ while the other $q = 0.65$, where $M_{\rm p}/M_\star = 0.013$ is the planet mass and $M_{\rm d}$ is the total disc mass. We will refer henceforth to the case $q = 0.2$ as the "light" case and the case $q = 0.65$ as the "massive" case. To satisfy these conditions, the initial surface density distribution is a radial power-law of the type $\Sigma = \Sigma_0 R^{-p}$ with $p = 0.3$ with the addition of an exponential taper at $R = 5$; $\Sigma_{0,\rm l} = 4.8 \times 10^{-5}$ for the light case and $\Sigma_{0,\rm m} = 15 \times 10^{-5}$ for the massive case.

The choice of the parameters for these simulations follows that used in Rosotti et al. (2017) and is based on the best fit model of the disc surrounding the star CI Tau, where an eccentric $13\,M_{\rm J}$ hot-Jupiter has been found (Johns-Krull et al. 2016).

We use a locally isothermal equation of state imposing a power-law radial temperature profile which provides a disc aspect-ratio of the type $h = H/R = 0.036R^\ell$ with $\ell = 0.215$. We use a Shakura & Sunyaev (1973) prescription for viscosity (ν) with $\alpha = 10^{-3}R^{-0.63}$, its radial depenence is set in order to obtain a stationary accretion profile throughout the disc ($\nu\Sigma = \text{const}$).

The planet is initially absent and its mass is progressively increased during the first 50 orbits. During this period of time the planet is kept on a circular Keplerian orbit at $a_{\rm p} = 1$, then its orbital parameters are left free to evolve under the action of the disc torque in order to allow the planet migration and eccentricity growth.

7.3 Results

In Fig. 7.1 two colour maps of the disc surface density at $t = 3 \times 10^4\, t_{\rm orb}$ are shown, both for the light and the massive disc case. The presence of an eccentric cavity and of a crescent shaped overdense feature at the apocentre of the cavity can be clearly noticed, consistent with the density perturbation expected for an eccentric disc (Teyssandier & Ogilvie 2016, see their eq. A31). We also report that this type of features, induced by the presence of a planet or stellar companion, has been previously discussed in the literature (Ataiee et al. 2013; Ragusa et al. 2017) to describe possible mechanisms producing the non-axisymmetric structures found in some transition discs in high resolution observations provided in the radio and

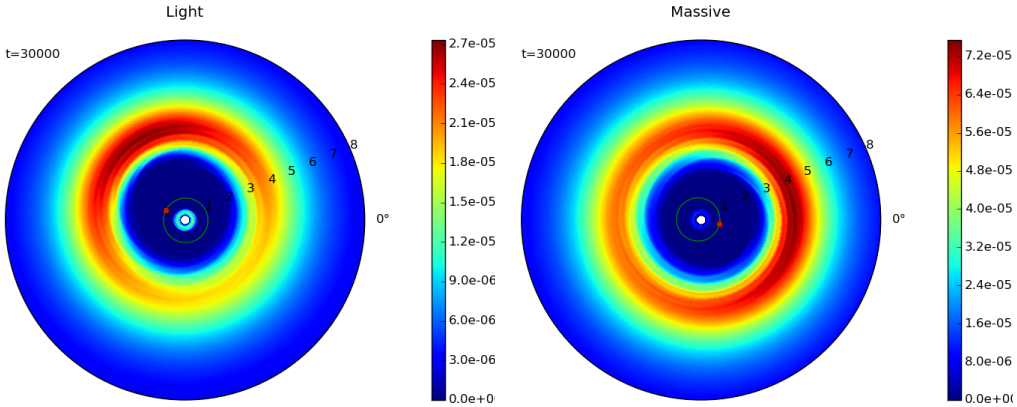


Figure 7.1: Density colour-plot for light (left panel) and massive (right panel) case at $t = 3 \times 10^4\, t_{\rm orb}$. The numbers indicate different radii. Note the formation of an eccentric cavity characterized by a horseshoe feature at its apocentre, consistent with the theoretical predictions about the density structure in eccentric discs.

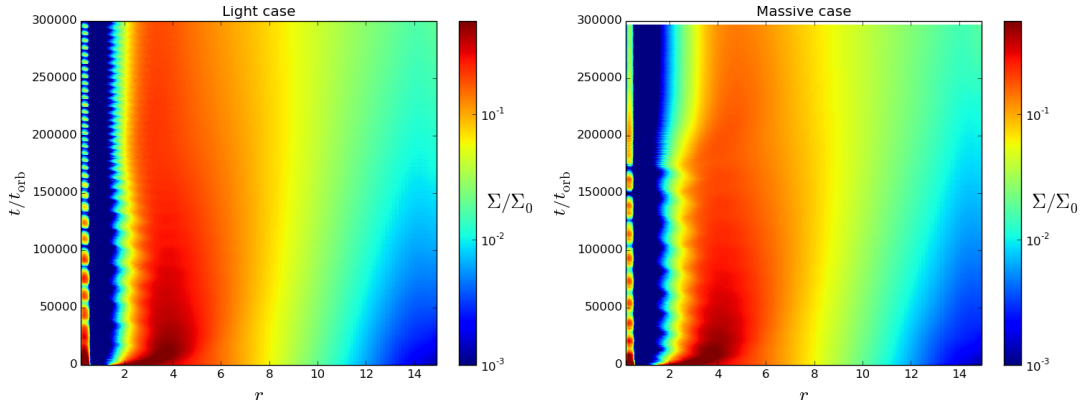


Figure 7.2: Density radial profiles Σ/Σ_0, obtained through azimuthal average, for light (left panel) and massive (right panel) case as a function of radius (x-axis) and time (y-axis), different colours represent different values of density.

NIR (Casassus 2016). In Fig. 7.2 the time evolution of the disc density radial profile is shown.

We used a Jacobi set of coordinates: thus the quantities related to the planet are computed in the reference frame of the star; while the quantities related to the disc are computed in the frame of the centre of mass (hereafter CM) of the system $M_\star + M_\mathrm{p}$. This peculiar set of coordinates is required since the disc orbits around the CM of the system. If computed in the star frame as the planet-related quantities, the disc eccentricity would be non-vanishing at large radii.

As noted by Ogilvie (2001), the eccentricity vector provides a useful tool to describe the values of eccentricity and pericentre phase of both the planet and the disc presented in this work (see also Eq. 1.10). In our 2D case this reads

$$\boldsymbol{e} = -\frac{j}{GM}\hat{\boldsymbol{u}}_z \times \boldsymbol{v} - \hat{\boldsymbol{u}}_R, \tag{7.1}$$

where j is the modulus of the z-component of the the angular momentum vector per unit mass, $\boldsymbol{v}$ is the velocity vector of the planet (or of the disc fluid element considered when computing the disc eccentricity), $\hat{\boldsymbol{u}}_z$ and $\hat{\boldsymbol{u}}_R$ are unit vectors pointing in the vertical and radial direction, respectively, and $M = M_\star + M_\mathrm{p}$. It can be shown that the modulus of $\boldsymbol{e}$ is the canonical expression for the orbital eccentricity e

$$e = \sqrt{1 - \frac{j^2}{GMa}}, \tag{7.2}$$

where a is the planet (or disc fluid element) semimajor-axis, and points in the pericentre direction.

The disc eccentricity is computed for each fluid element of the grid using Eq. 7.1. The scalar eccentricity e radial profile is then obtained through an azimuthal average of grid cells at each radius.

Besides the direct computation of the orbital eccentricity of the disc fluid elements, we can quantify the global amount of disc eccentricity using the angular momentum deficit (AMD), that is defined as follows

$$A_\mathrm{d} = \int \Sigma(R,\phi)\left[\sqrt{GMa(R,\phi)} - v_\phi(R,\phi)R\right] RdR\,d\phi, \tag{7.3}$$

where $\Sigma(R,\phi)$ is the disc surface density and $v_\phi(R,\phi)$ the azimuthal velocity and $a(R,\phi)$

$$a(R,\phi) = -\frac{1}{2}\frac{GM}{E(R,\phi)}, \tag{7.4}$$

where $E(R,\phi)$ is the mechanical energy per unit mass

$$E(R,\phi) = -\frac{GM}{R} + \frac{1}{2}v^2(R,\phi), \tag{7.5}$$

where $v(R,\phi)$ is the gas velocity map.

The AMD is the amount of angular momentum the disc is lacking in comparison with a situation where the gas orbits the CM of the system on circular orbits. The same quantity is defined for the planet as follows

$$A_\mathrm{p} = J_\mathrm{circ,p} - J_\mathrm{p}, \tag{7.6}$$

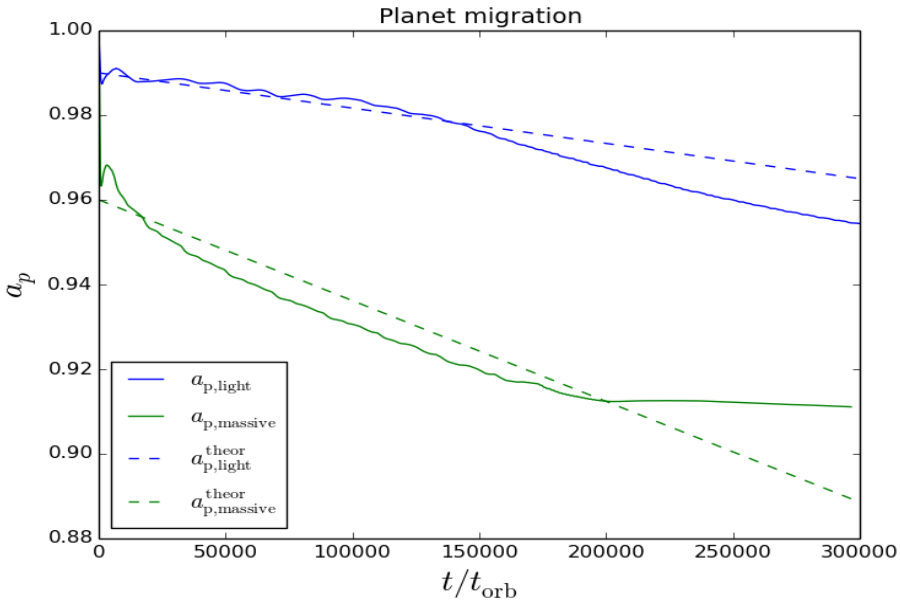


Figure 7.3: Planet migration: $a_{\rm p}$ as a function of time for light (blue curve) and massive (green curve) case. The dashed lines are two lines with slope $\dot{a}_{\rm p} = (100\,t_\nu)^{-1}$ for the light case and $\dot{a}_{\rm p} = (35\,t_\nu)^{-1}$ for the massive one, which represents the theoretical migration rate predicted by Eq. (7.9) for the two cases.

where $J_{\rm circ,p} = M_{\rm p}\sqrt{GM_\star a_{\rm p}}$ is the angular momentum that the planet would have on a circular orbit with radius $a_{\rm p}$, $J_{\rm p} = M_{\rm p} v_{\phi,p} R_{\rm p}$ is the planet angular momentum where $v_{\phi,p}$ is the planet instantaneous azimuthal velocity and $R_{\rm p}$ is its separation from the central star. It can be shown that, for small eccentricities, Eq. 7.6 can be approximated by

$$A_{\rm p} \approx \frac{1}{2} e_{\rm p}^2 J_{\rm circ,p}. \tag{7.7}$$

7.3.1 Planet migration

The disc-planet interaction also drives the migration of the planet, see Fig. 7.3. This is actually a consequence of the conservation of the total angular momentum. Indeed the total AMD $A_{\rm tot} = A_{\rm p} + A_{\rm d}$ can be written as

$$A_{\rm tot} = J_{\rm circ,p} + J_{\rm circ,d} - J_{\rm tot} \tag{7.8}$$

where $J_{circ,\rm d}$ is the angular momentum of the disc if it was circular and $J_{\rm tot}$ is the total angular momentum of the system. It follows straightforwardly that, in order to conserve the total angular momentum, any change in $A_{\rm tot}$ (which depends on the eccentricity of both the planet and the disc) in the simulations has to be accompanied by corresponding changes in $J_{\rm circ,d} + J_{\rm circ,p}$, i.e. varying the semi-major axis of the orbits both in the planet and in the gas.

We can compare the migration timescale $t_{\rm mig} = a_{\rm p}/\dot{a}_{\rm p}$ we observe in our simulation with the standard type II migration rate[1] (Syer & Clarke 1995; Ivanov et al.

[1]It should be noticed that when the planet is left free to evolve it is completely embedded in

1999, see Eq. 4.29 for its derivation)

$$t_{\mathrm{typeII}} = \frac{M_{\mathrm{d}}^{\mathrm{local}} + M_{\mathrm{p}}}{M_{\mathrm{d}}^{\mathrm{local}}} t_\nu \tag{7.9}$$

where $M_{\mathrm{d}}^{\mathrm{local}} = 4\pi\Sigma(a_{\mathrm{p}})a_{\mathrm{p}}^2$ is approximately the unperturbed amount of disc material contained inside the orbit of the planet; t_ν is the viscous timescale (Eq. 2.28)

$$t_\nu = (\alpha h^2 \Omega_{\mathrm{p}})^{-1}, \tag{7.10}$$

where $h = H/R$ is the disc aspect ratio, α is the Shakura & Sunyaev (1973) viscous parameter. Substituting the values from our simulations one gets $t_\nu \sim 1.2\times10^5 t_{\mathrm{orb}}$ for both our setups, which is perfectly consistent with the damping timescale for the semi-major axis. The timescales we obtain for type II migration from our simulations are $t_{\mathrm{typeII}}^{\mathrm{light}} \sim 100\, t_\nu$ and $t_{\mathrm{typeII}}^{\mathrm{massive}} \sim 35\, t_\nu$, plotted as dashed lines in Fig. 7.3. We find that the ratio $t_{\mathrm{typeII}}^{\mathrm{light}}/t_{\mathrm{typeII}}^{\mathrm{massive}} \sim 2.85$ is perfectly consistent with the ratio one would expect from Eq. 7.9 when comparing the migration rate of a planet embedded in two discs differing by factor 3 in the disc mass. The result is thus consistent with the classical Type II migration rate predicted by Syer & Clarke (1995) and Ivanov et al. (1999), contrary to what have been found in the studies of Dürmann & Kley (2015) and Duffell et al. (2014).

It should be noticed that the migration of the planet apparently stops at $t \sim 2\times10^5$ orbits in the massive case. As we will see, this can be reasonably attributed to the rapid broadening of the cavity that can be noticed in Fig. 7.2. A larger cavity implies a clearing of material from the region where resonances mediate energy exchange between disc and planet. These structural changes also correspond to changes in the eccentricity evolution. Interestingly, a similar behaviour has been previously observed in Papaloizou et al. (2001) in which they observed a change of migration rate (and even direction) as a consequence of the disc structure evolution. In contrast, in the light case the migration accelerates at late times. In the latter case the migration rate indeed appears to increase starting from 1.5×10^5 orbits , which (from Fig. 7.2) can be seen to coincide with the disc's inner edge moving slightly inwards.

7.3.2 Eccentricity and pericentre phase evolution

In this section we will limit our discussion to the qualitative behaviour of the eccentricity evolution in the simulations; We postpone a possible modelling and interpretation of the results to the following sections.

The colour plots in Fig. 7.4 and 7.5 show the scalar eccentricity (azimuthal average) and pericentre phase, respectively at different times (y-axis) and radii (x-axis) both for the light and massive disc case.

It is interesting to note from these plots that the disc eccentricity evolution can be considered "rigid": in Fig. 7.4, for any fixed time, an increase in the eccentricity at small radii is reflected in an increase also at larger radii. The disc

the disc, and it spends the first $\approx 10^3$ orbits undergoing type I migration, migrating at a much faster rate. This produces the impression in Fig. 7.3 that the initial $a_{\mathrm{p},0} < 1$. As soon as the cavity is cleared, it starts migrating at the slower type II rate.

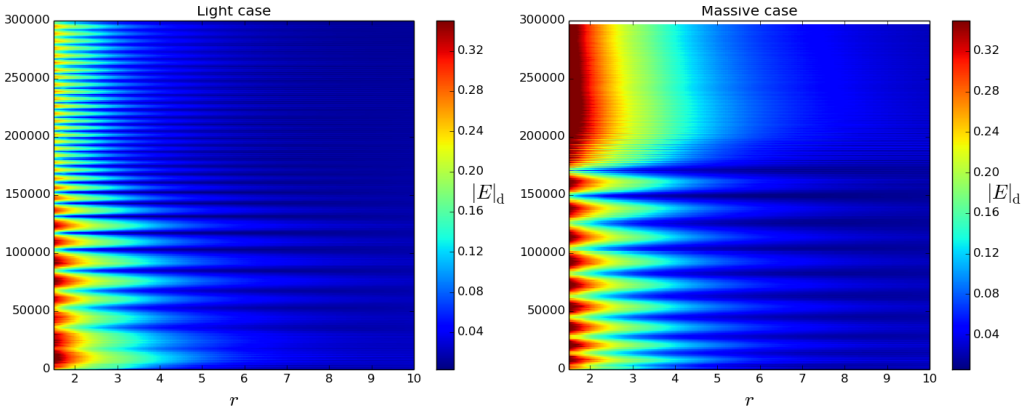


Figure 7.4: Colour plots of the scalar eccentricity (azimuthal average) as a function of time (y-axis) and radius (x-axis) for light (left panel) and massive (right panel) case.

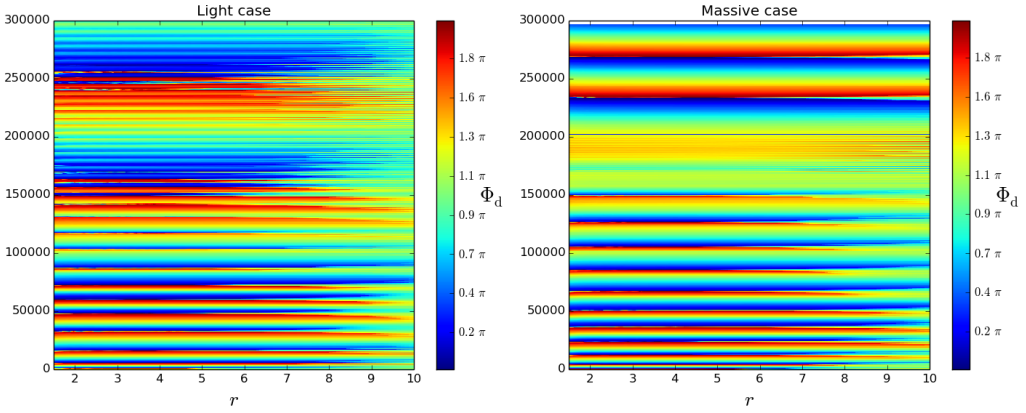


Figure 7.5: Pericentre phase colour-plot as a function of time (y-axis) and radius (x-axis) for light (left panel) and massive (right panel.) case.

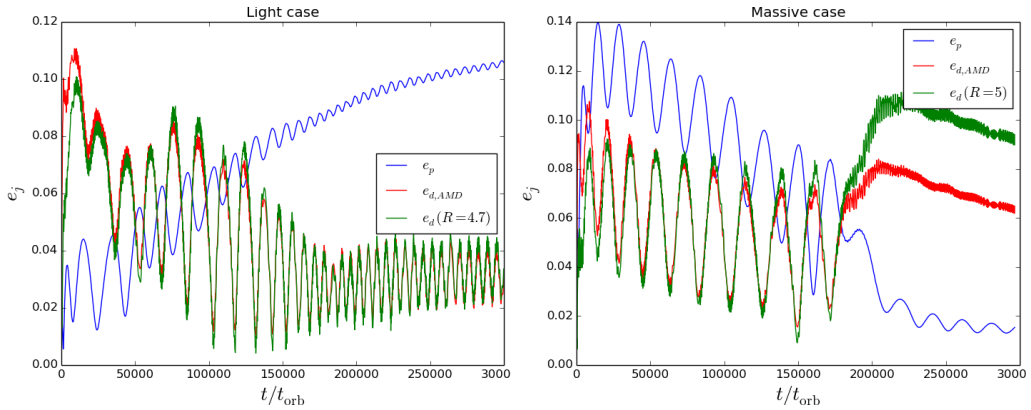


Figure 7.6: Eccentricity e as a function of time for light (left panel) and massive (right panel) case. The blue curve shows the planet eccentricity, the green curve the disc eccentricity at $R = 4.7$ in the light case and at $R = 5$ (azimuthal averages) in the massive one, while the red curve is a global measurement of the disc eccentricity starting from the AMD (see Sec. 7.3.2). The choice to use two different reference radii for the disc eccentricity is due to the slightly different size of the cavity in the two cases.

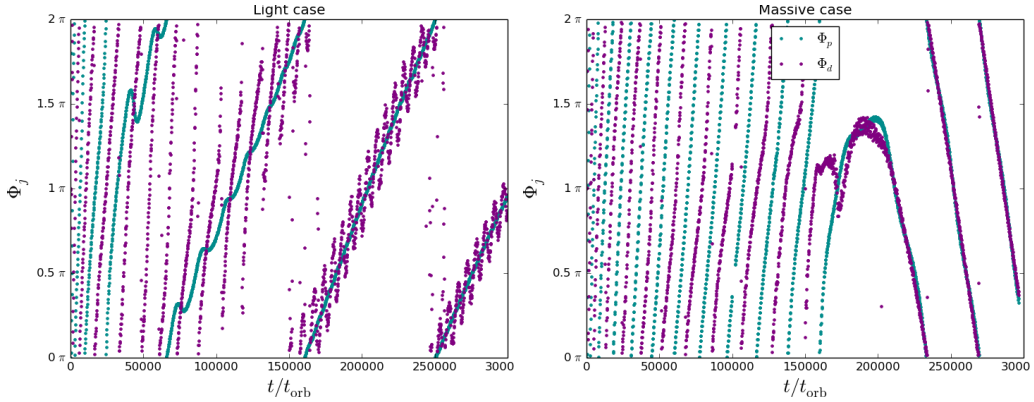


Figure 7.7: Pericentre phase as a function of time for light (left panel) and massive (right panel) case (we remind that the value of the pericentre phase is constant throughout the entire disc domain). The cyan and violet curve represent planet and disc pericentre phase. During the first $\approx 4 \times 10^4$ orbits both simulations show an anti-aligned precession ($|\Phi_{\rm p} - \Phi_{\rm d}| \approx 180°$). After $\approx 4 \times 10^4$ orbits in the light case the planet precession decouples from that of the disc, becoming much slower than the disc one. At very late times ($t \gtrsim 2 \times 10^5$ orbits) also the disc precession rate slows down, and precesses with the planet in a pericentre aligned configuration. The massive case remains in the anti-aligned configuration for much longer, even though also in this case a transition toward the slowly precessing aligned configuration takes place after $t \approx 2 \times 10^5$ orbits. In the massive case, the transition appears to be accompanied by a reversal of the precession rate, which becomes retrograde

"rigid" behaviour is even more evident looking at Fig. 7.5, at fixed time, the pericentre phase is the same at all radii throughout the entire disc[2]. Furthermore the radial profiles of the eccentricity and pericentre phase in Fig. 7.4 and 7.5 imply that the gas orbits are a set of nested, pericentre aligned eccentric orbits with an eccentricity profile decreasing with radius.

Fig. 7.6 and 7.7 show the evolution of planet and disc eccentricity and their pericentre phases, respectively. The red and green curves in Fig. 7.6 represent two different ways to estimate the disc eccentricity: the red curve ($e_{\rm d,AMD}$) is computed inverting the approximate relationship between eccentricity and AMD given in the case of the planet in Eq. (7.7) and which yields

$$e_{\rm d,AMD} = \sqrt{\frac{2A_{\rm d}}{J_{\rm d,circ}}}, \tag{7.11}$$

where $J_{\rm d,circ}$ is given by

$$J_{\rm d,circ} = \int \Sigma(R,\phi)\sqrt{GMa(R,\phi)}\, RdR\, d\phi; \tag{7.12}$$

this approach allows us to give an estimate of the disc eccentricity relying on global disc quantities, in fact treating it as if it was a second planet. We then notice that

[2]For $R > 8$ the eccentricity is almost negligible. The algorithm we used to compute the pericentre phase tends to attribute $\Phi_{\rm d} = 180°$ when $e \approx 0$ since it is not possible to attribute a pericentre in a circular orbit.

the density peak in the two simulations is located at $R \approx 4.7$ in the light case and at $R \approx 5$ in the massive one; the green curve represents the values of the eccentricity evolution at these radii. The good agreement of the two curves tells us that the global behaviour of the disc is dominated by the values the eccentricity has at these radii, the discrepancy in the massive case between the red and green curve after $\approx 1.7 \times 10^5$ orbits suggest that the reference radius for the eccentricity has migrated outward (consistently with the broadening of the cavity, Fig. 7.2).

Both simulations show that the planet and the disc exchange eccentricity through slow periodic (period $\Delta t \gtrsim 10^4$ orbits) anti-phased oscillations (a maximum in the planet curve correspond to a minimum in the disc one) superimposed on a series of roughly linearly growing and decreasing trends. It should be noted that the frequency of the oscillation is not constant throughout the entire length of the simulation, we will discuss more in detail this feature in Sec. 7.5.3.

Both the light and the massive case show a rapid exponential growth of the disc eccentricity (Fig. 7.6) during the first stages of evolution ($t \lesssim 1.5 \times 10^4$ orbits) up to values $e_\mathrm{d} \sim 0.11$, then a slower decrease at later times. Interestingly, the maximum level of disc eccentricity achieved is the same for both simulations. This might suggest that some non-linear effects prevent the disc eccentricity to grow further. The planet eccentricity in the massive case has a similar behaviour: it grows fast in the beginning, attains a value $e_\mathrm{p} = 0.14$ and starts decreasing at the same time as the disc eccentricity. The planet eccentricity in the light disc case in contrast has a completely different behaviour: its growth oscillates around $e_\mathrm{p} = 0.025$ for $t \lesssim 4 \times 10^4$ orbits, but then at later times starts growing again at constant rate; at very late times ($t \gtrsim 2 \times 10^5$ orbits) the planet eccentricity growth rate appears to slow down.

The precession of the pericentre phase (Fig. 7.7) presents some very interesting features as well. Both the massive and the light disc cases show in the initial stages ($t \lesssim 4 \times 10^4$ orbits) of the simulation an anti-phased precession of the disc-planet pericentre: the planet and disc pericentre precess at the same rate maintaining a phase difference $|\Phi_\mathrm{p} - \Phi_\mathrm{d}| \approx 180°$.

In the light case, after $\approx 4 \times 10^4$ orbits, in correspondence with the beginning of the growing trend of the planet eccentricity, the planet pericentre phase starts to precess much more slowly than the disc one. After $t \gtrsim 1.5 \times 10^5$ orbits also the disc transitions to a slower precession rate. When this condition is reached, the planet and the disc pericentre phases are aligned, precessing at the same slow rate. Some oscillations at a faster frequency in the disc pericentre phase can be noticed in this slow configuration.

In the massive case, the anti-aligned configuration (present in the light case just for $t \lesssim 4 \times 10^4$ orbits) lasts for much longer: only at $t \gtrsim 2 \times 10^5$ the transition toward the slower aligned precession rate appears to take place. However, in this case the transition appears to be accompanied by a reversal of the precession rate (which becomes retrograde), a significant slow down of the oscillation frequency of the eccentricities and also by a variation in the disc eccentricity radial profile (see the right panel of Fig. 7.4).

Another interesting quantity that is useful for the interpretation of the results presented here is the AMD (equations (7.3) and (7.6), Fig. 7.8). The evolution of

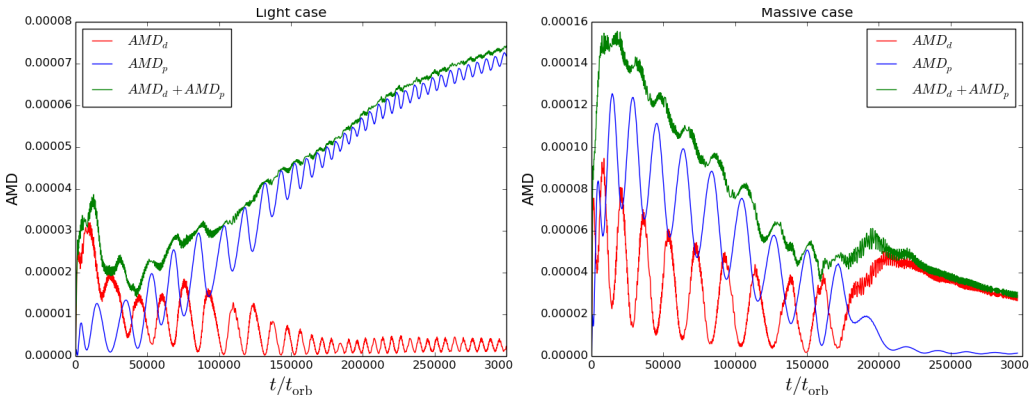


Figure 7.8: AMD $A_{\rm p}$, $A_{\rm d}$, $A_{\rm tot}$ as a function of time for light (left panel) and massive (right panel) case.

the AMD reflects the evolution of the eccentricity since $A \propto e^2$ in the limit of low eccentricity (Eq. (7.7)). The total AMD $A_{\rm tot} = A_{\rm p} + A_{\rm d}$ in the light case starts growing when the system evolves to the aligned configuration, while in the massive case it decreases up to the end of the simulation. However the most interesting feature of the oscillations observed in the eccentricity is that their amplitude is AMD conserving (see Fig. 7.8): the overall amount of AMD changes both in the planet and in the disc, but the amplitude of the oscillations is such that the $A_{\rm tot}$ is conserved during one oscillation. The implication of this is that the periodic oscillations only *exchange* angular momentum between the planet and the disc: for a given amount of angular momentum exchanged the amplitude of the fluctuations is fixed by the orbital properties of the planet and the disc.

As we will see, most of the features that we observed in these simulations (rapid initial exponential growth of eccentricity, long-term periodic oscillations, rigid precession of the pericentre phase, linear growth or decrease of the eccentricity at late times) can be interpreted in terms of either a classical linear theory of a 2-planet system (Murray & Dermott 1999), or its extension to disc-planet interactions (Teyssandier & Ogilvie 2016).

We devote section 7.4 and 7.5 to this interpretation.

7.4 Interpretation of the results

Before trying to interpret the complex evolution revealed by the hydrodynamical simulations, it is instructive to build up a qualitative picture of the evolution of dynamically coupled eccentric planet/disc systems.

The structure of an eccentric disc-planet system can in principle be described as a superposition of rigidly precessing normal modes, each of which is characterised by its disc eccentricity profile $e(R)$ (normalised to planet eccentricity), growth rate γ, precession rate ω and angular offset $\Delta\Phi$ between the line of apses of the disc and planet (Teyssandier & Ogilvie 2016).

The rate of precession ω is set by many different contributions. Among them, purely secular gravitational disc-planet interaction is expected to cause the pro-

grade precession of the pericentre phase of both planet and disc, while pressure effects in discs with standard pressure profiles ($dP/dR < 0$, where P is the pressure radial profile) are expected to cause the retrograde precession of the pericentre phase (it can be shown that the precession rate observed in Fig. 7.7 are consistent with that predicted by Teyssandier & Ogilvie 2016).

Mode growth is generically driven by resonances. With standard disc parameters, eccentric Lindblad resonances allow the disc and planet eccentricity to grow, while eccentric corotation resonances cause the eccentricity to decrease (Goldreich & Tremaine 1980; Goldreich & Sari 2003; Ogilvie & Lubow 2003). The growth or damping of the planet eccentricity thus depends on the balance between these two opposite effects. Goldreich & Tremaine (1980) showed that if the planet does not perturb significantly the surface density of the disc, the corotation torque slightly exceeds the Lindblad one, damping the planet eccentricity toward circular orbits. In contrast, if the planet carves a sufficiently deep cavity or gap around the planet (at least a factor $\sim 10^{-3}$, Duffell & Chiang 2015), and if no material replenishes the corotation region, the corotation torque saturates (Goldreich & Sari 2003; Ogilvie & Lubow 2003) so that dominance of the Lindblad resonances causes the eccentricity to grow. In addition to the effect of eccentric co-rotation resonances, the disc viscosity is also expected to circularize the gas orbits in the disc. Putting all this together implies that the disc eccentricity evolution can be expected to depend on the planet-star mass ratio, planet eccentricity, pressure and disc viscosity (Artymowicz & Lubow 1994; Crida et al. 2006).

Each of the aforementioned effects have been included in the derivation by Teyssandier & Ogilvie (2016) of the eigenmodes of an eccentric fluid disc, both with and without the inclusion of a planet. In this formalism, the real part of the eigen-values associated with each eigen-mode corresponds to the precession rate.

7.4.1 Case of no mode damping or driving

We first consider the case where mode damping and pumping can be neglected. This implies the absence of resonant interactions and viscous effects and therefore means that the gravitational influence of the planet on the disc is mediated by the secular interaction, i.e. the response of the disc to the zero frequency ($\Omega_{\rm m} = 0$) component of the Fourier decomposition of the acceleration induced by the planet.

The secular interaction can be visualised as being the response of the disc to an elliptical ring of material representing the time average of the planet's mass distribution around its orbit. If the system is in a single mode the amplitude of disc and planet eccentricity is constant in time and the entire system undergoes rigid precession at a constant rate.

If however the system exists in a superposition of modes, each with characteristic eccentricity profile and precession rate, the net eccentricity of both the planet and the disc undergoes cyclical variations, that correspond to the beats of the fundamental modes, depending on the instantaneous phase relationship of the various modes. The varying eccentricity of both planet and disc result in an exchange of angular momentum between the two components. Since for Fourier mode with frequency $\Omega_{\rm m}$ the relationship between energy exchange and angular momentum exchange is given by $\Delta E = \Omega_{\rm m} \Delta L$, it follows that the secular interaction involves

zero energy exchange between planet and disc (recalling that $\Omega_{\rm m} = 0$ for secular interaction). It is therefore convenient to consider the interaction in terms of the angular momentum deficit (AMD) defined in Eq. (7.3) and (7.6).

The differential precession of an ensemble of modes results in a variation of the AMD of disc and planet at constant energy. Total angular momentum conservation requires that the total AMD of the planet plus disc is constant.

We have seen that the eccentricity and AMD variations of the simulated disc-planet system can indeed be described in terms of such fluctuations on which slower long term trends in mode amplitude resulting from net pumping/damping are superposed.

The fact that the oscillatory behaviour is close to being sinusoidal suggests that the evolution can be understood in terms of the superposition of two dominant modes. We will find that we can gain significant qualitative insight into the behaviour of the system by considering the analogue problem of the secular interaction between two *point masses* for which (given the number of degrees of freedom in the system) there are just two modes (as known from textbook studies of celestial mechanics Murray & Dermott 1999). We however emphasise that we do not necessarily expect the mode structure to be the same in the case of the fluid disc and will indeed find that - whereas the modes in the two-planet case both undergo prograde precession - the role of pressure within the disc can induce retrograde precession in one of the modes. Nevertheless, we will find that a heuristic understanding of the nature of the two modes in the point mass case will be extremely useful in guiding our interpretation of the simulations.

7.4.2 Case of secular interaction between two point masses

In this section we describe a toy model accounting only for secular contributions to the eccentricity equations. We aim to give a simplified description of the coupled evolution of planet and disc in order to interpret some features of the planet and disc eccentricity and pericentre phase evolution discussed in Sec. 7.3.

Teyssandier & Ogilvie (2017) predict a "rigid" evolution of the eigenmodes, leading us to expect that in general the eccentricity radial profile evolves rigidly as $e_{\rm d}(t, R) = e_0(R)h(t)$, where $h(t)$ is a generic function of the time only, and that the pericentre longitude does not depend on the radius $\Phi_{\rm d}(R, t) \equiv \Phi_{\rm d}(t)$. We thus expect that a simplified description of the evolution of the system can be obtained by replacing the disc with a virtual planet, adding some terms to account for the disc eccentricity pumping and damping effects. In fact this approach consists in modelling the disc-planet interaction as a planet-planet interaction where the outer planet is a virtual mass with disc averaged orbital characteristics: semi-major axis $a_{\rm d}$, longitude of pericentre $\Phi_{\rm d}$, mass $M_{\rm d}$ and eccentricity $e_{\rm d}$. We will use $a_{\rm p}$, $\Phi_{\rm p}$, $M_{\rm p}$ and $e_{\rm p}$ to refer to the actual planet instead.

It is important to bear in mind that such a description will not be quantitatively correct for two main reasons:

a) This approach intrinsically neglects pressure (which we have seen in the previous section to have a role in determining the precession rate).

b) The approximation of a disc of nested ellipses by an equivalent point mass

particle forces us to reduce local quantities such as the density or the disc eccentricity to equivalent global quantities without a well defined prescription.

Moreover such an approach does not of course include the additional effects of viscous damping and driving of eccentricity at resonances, which would need to be added *ad hoc*.

In the following equations we will use the following notation

$$E_j = |E_j| e^{i\Phi_j},\ j = \{\mathrm{p}, \mathrm{d}\}, \tag{7.13}$$

where $|E_j| = e_j$ is the "physical" eccentricity and Φ_j is its pericentre phase, the subscripts p and d refer to the *planet* and *disc-"virtual" planet*. This formalism allows us to write one single set of equations for both eccentricity and pericentre phase.

We follow the Hamiltonian approach given by Zhang et al. (2013), in which the gravitational potential produced by the two components (in our case the real planet and the disc virtual planet) of the system is expanded up to the second order in e_p and e_d. The equations ruling the evolution of the complex eccentricities E_p and E_d have the form

$$\begin{pmatrix} \dot{E}_\mathrm{p} \\ \dot{E}_\mathrm{d} \end{pmatrix} = \mathbf{M} \cdot \begin{pmatrix} E_\mathrm{p} \\ E_\mathrm{d} \end{pmatrix}, \tag{7.14}$$

where the notation $\dot{E}_i$ indicates the time derivative, while the complex matrix $\mathbf{M}$ reads

$$\mathbf{M} = i\Omega_\mathrm{sec} \begin{pmatrix} q & -q\beta \\ -\sqrt{\alpha}\beta & \sqrt{\alpha} \end{pmatrix}, \tag{7.15}$$

where $\alpha = a_\mathrm{p}/a_\mathrm{d}$, $q = M_\mathrm{d}/M_\mathrm{p}$, $\beta = b^{(2)}_{3/2}(\alpha)/b^{(1)}_{3/2}(\alpha)$, where $b^{(n)}_{3/2}(\alpha)$ is the n-th Laplace coefficient (see also Eq. 3.40)

$$b^{(n)}_{3/2}(\alpha) = \frac{1}{\pi}\int_0^{2\pi} \frac{\cos(n\theta)}{(1 - 2\alpha\cos(\theta) + \alpha^2)^{3/2}} d\theta. \tag{7.16}$$

For $\alpha \ll 1$, $\beta \approx 5\alpha/4$ since $b^{(1)}_{3/2}(\alpha) \approx 3\alpha$ and $b^{(2)}_{3/2}(\alpha) \approx 15\alpha^2/4$ (Murray & Dermott 1999). The matrix $\mathbf{M}$ in Eq. (7.15) is purely imaginary and accounts for the secular, non-dissipative disc-planet interaction. Ω_sec is a real scaling parameter for the matrix and has the dimension of a frequency:

$$\Omega_\mathrm{sec} = \frac{1}{4}\Omega_\mathrm{p}\frac{M_\mathrm{p}}{M_\star}\alpha^2 b^{(1)}_{3/2}(\alpha). \tag{7.17}$$

The solutions to equation (7.14) is

$$\begin{pmatrix} E_\mathrm{p}(t) \\ E_\mathrm{d}(t) \end{pmatrix} = C_1 \begin{pmatrix} \eta_\mathrm{s} \\ 1 \end{pmatrix} e^{i\omega_\mathrm{s} t} + C_2 \begin{pmatrix} \eta_\mathrm{f} \\ 1 \end{pmatrix} e^{i\omega_\mathrm{f} t}, \tag{7.18}$$

where C_1 and C_2 are constants that depend on the initial conditions, $\omega_\mathrm{s,f}$ and $(\eta_\mathrm{s,f}, 1)$ are the complex eigen-values and complex eigen-vectors of $\mathbf{M}$, respectively.

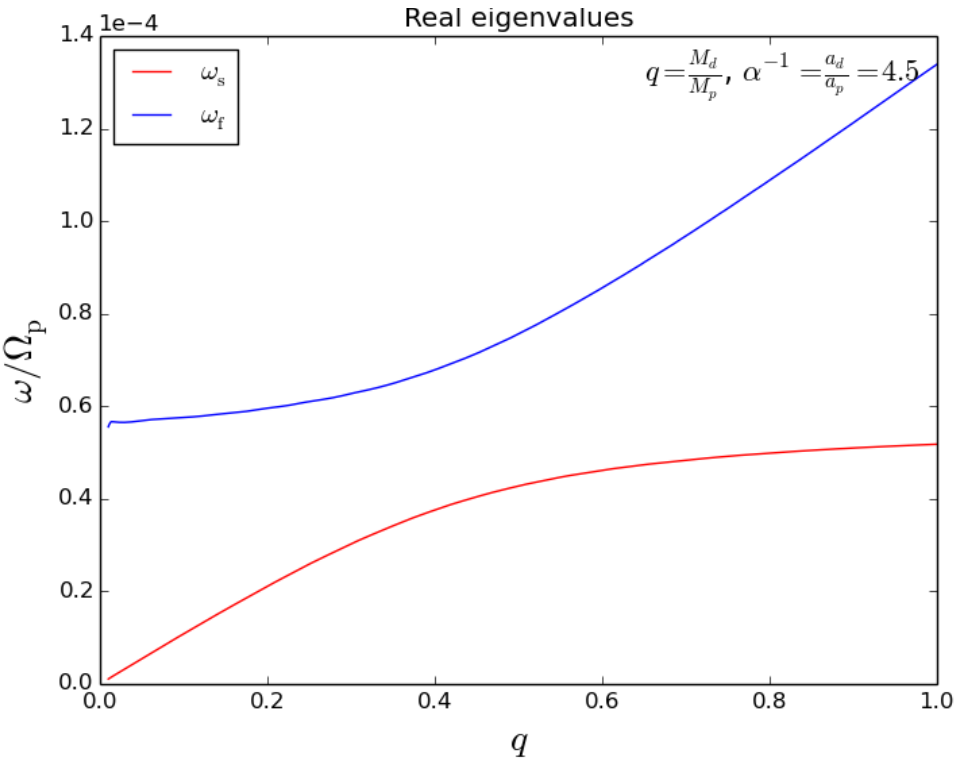


Figure 7.9: Eigenvalues as a function of q for fixed $\alpha^{-1} = 4.5$. The blue curve represents ω_f while the red one ω_s. The eigenvalues are expressed in units of the planet orbital frequency Ω_p, and represent the precession rate of the pericentre phase.

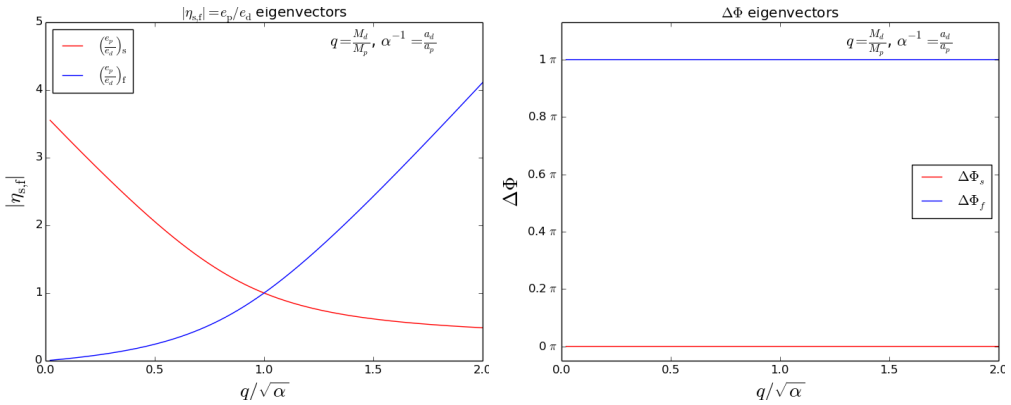


Figure 7.10: $|\eta_{s,f}| = e_p/e_d$ (left panel) and $\Delta\phi = \arg(\eta_{f,s})$ (right panel) of eigenvectors as a function of $q/\sqrt{\alpha}$. The blue curve refers to the fast mode the red curve refers to the slow one. It should be remembered that $e^{i\pi} = -1$. $\eta_{s,f}$ represent the ratio between the planet and the disc eccentricity if just one mode is present. While $\Delta\Phi$ represents the pericentre phase difference between the planet and the disc if just one mode is present.

With reference to Eq. (7.18), the eigen-values $\omega_{\rm s,f}$ of the matrix $\mathbf{M}$ in Eq. 7.15 are

$$\omega_{\rm s,f} = \frac{1}{2}\Omega_{\rm sec}\left(q+\sqrt{\alpha}\right) \mp \frac{1}{2}\Omega_{\rm sec}\sqrt{\left(\sqrt{\alpha}-q\right)^2+4q\sqrt{\alpha}\beta^2} \tag{7.19}$$

The eigen-frequency ω in units of $\Omega_{\rm p}$, plotted for varying q and fixed α in Fig. 7.9, gives us information on the precession rate of the pericentre phase. From Eq. (7.19) we can clearly see that $\omega_{\rm s} < \omega_{\rm f}$; for this reason we will refer to the s mode as the "slow mode" and to the f mode as the "fast mode". It is notable that $\Omega_{\rm sec}$ sets the timescale of the oscillations, it is independent of q and scales as α^3 for $\alpha << 1$ (equation (7.17)). The individual precession frequencies and hence the beat frequency $\Delta\omega = \omega_{\rm f} - \omega_{\rm s}$ do however depend on q; in particular, the dependence on q mostly affects the beat frequency $\Delta\omega = \omega_{\rm f} - \omega_{\rm s}$, which has a minimum when $q = \sqrt{\alpha}$ (see Fig. 7.9).

The components of the eigen-vectors of the matrix $\mathbf{M}$ instead are

$$\eta_{\rm s,f} = \frac{1}{2\sqrt{\alpha}\beta}(\sqrt{\alpha}-q) \pm \frac{1}{2\sqrt{\alpha}\beta}\sqrt{\left(\sqrt{\alpha}-q\right)^2+4q\sqrt{\alpha}\beta^2}. \tag{7.20}$$

According to Eq. (7.18), the ratio of planet to disc eccentricity when only one of the modes is present is given by $|\eta| = e_{\rm p}/e_{\rm d}$.

It should be noticed that $\eta_{\rm f} < 0$ for any parameter choice, while in contrast $\eta_s > 0$. This implies that the fast mode involves misalignment by π between the pericentre phases of the planet and virtual planet (i.e. disc) while the two orbits are aligned in the case of the slow mode. This is illustrated in the right hand panel of Fig. 7.10. It should be also noticed that for $q/\sqrt{\alpha} < 1$ one has $|\eta_{\rm s}| > 1$ and $|\eta_{\rm f}| < 1$ (left panel of Fig. 7.10), while for $q/\sqrt{\alpha} > 1$ one gets $|\eta_{\rm s}| < 1$ and $|\eta_{\rm f}| > 1$. The condition $q = \sqrt{\alpha}$ marks the condition that the two components have equal angular momentum if on a circular orbit[3].

The above inequalities imply that the component (i.e. planet or virtual planet) with the higher circular angular momentum will have the greater amplitude when the system is entirely in the slow mode ($C_2 = 0$), while the component with lower circular angular momentum will have greater amplitude when the system is entirely in the fast mode ($C_1 = 0$). In the limit that the circular angular momenta of the two components is very different (i.e.: $q \to 0$), the fast mode becomes overwhelmingly dominated by the component with the smaller angular momentum and in this limit can be envisaged as the precession of a test particle in the combined potential of the central object and the potential generated by a circular ring of material at the location of the other 'planet'. In this limit the slow mode is non precessing and has finite eccentricity contributions in both components. As q tends to $\sqrt{\alpha}$ the modes become increasingly entwined in the sense that both modes contain comparable contributions in both components. We will discuss in the following sections how the evolution of the two simulations described here can be understood in terms of the different ratios of $q/\sqrt{\alpha}$ in the two cases, and thus[3] of $J_{\rm d,circ}/J_{\rm p,circ}$.

[3]It should be noticed that the ratio $q/\sqrt{\alpha}$ is equivalent to the ratio $J_{\rm d,circ}/J_{\rm p,circ} = (M_{\rm d}\sqrt{GM_*a_{\rm d}})/(M_{\rm p}\sqrt{GM_*a_{\rm p}})$, and provides thus a measure of the relative contribution of the disc and the planet to the total amount of angular momentum of the system.

Given the large number of simulations with fixed binaries in the literature (D'Angelo et al. 2006; Müller & Kley 2013; Duffell & Chiang 2015; Miranda et al. 2017; Thun et al. 2017), we believe it is very interesting to notice that the case $q = 0$ is the reference case for those simulations of a circumbinary disc surrounding a binary system (of any mass ratio) with fixed orbital parameters. A discussion of the instructive case $q = 0$ can be found in Appendix C.1.

7.4.3 Phenomenological implementation of pumping and damping terms

The mere gravitational interaction we discussed in the previous section does not provide any mode evolution since it describes stationary modes. Nevertheless, we clearly observe instead in our simulations the growth and the decrease of the eccentricity at different stages of the simulations. This clearly implies that some modelling of this behaviour needs to be included in our simplified description.

Zhang et al. (2013) treated the damped three body problem in order to show that the damping of the eccentricity of a hot-Jupiter operated by the tidal effects of the central-star can be slowed down if a second planet orbiting on an outer orbit is present. To do so, they added some real terms in the matrix $\mathbf{M}$ in Eq. (7.14). The introduction of these terms introduces a complex component of the eigenvalues which is responsible for the exponential damping (or pumping depending on the sign) of the mode.

In our simulations, the resonances are initially very strong due to the presence of a large amount of material in the cavity region during the initial phases. For this reason the planet eccentricity in the massive case and the disc one in both cases grows very fast for $t \lesssim 10^4$ orbits following an exponential trend. However, after this initial transient, the gas depletion in the cavity region leads to the saturation of the pumping mechanism, which is also associated with the attainment of a maximum value of the disc eccentricity. At later times the pumping/damping mechanisms are such to cause a linear increase/decrease of the eccentricity with time, in contrast with the exponential trend predicted by Zhang et al. (2013) modelling to include resonant and viscous effects.

As described at the beginning of Sec. 7.4, the physical scaling of pumping and damping needs to account for viscous effects and the contribution of each single resonance. The work by Teyssandier & Ogilvie (2016) includes a detailed formulation to deal with pumping and damping effects. The viscous effects depend on the disc eccentricity radial gradient and on the viscosity prescription adopted; it provides damping of eccentricity for standard disc parameters. For the resonant interaction, each individual resonance depends differently on several factors including: the disc density profile, the planet eccentricity and pressure effects, which vary significantly throughout the simulation.

In the light of these considerations, we conclude that our ability to model the eccentricity pumping/damping in the framework of the toy model is very limited. The main difficulty resides in identifying the dependence of the pumping/damping on global planet and disc properties. We decide for this reason to include these effects in our model by prescribing a linear time evolution of the C_1 and C_2 pa-

rameters in Eq. (7.18), which becomes

$$\begin{pmatrix} E_{\rm p}(t) \\ E_{\rm d}(t) \end{pmatrix} = C_1(t) \begin{pmatrix} \eta_{\rm s} \\ 1 \end{pmatrix} e^{i\omega_{\rm s}t} + C_2(t) \begin{pmatrix} \eta_{\rm f} \\ 1 \end{pmatrix} e^{i\omega_{\rm f}t}. \tag{7.21}$$

We prescribe the time dependence of $C_1(t)$ and $C_2(t)$ to be in the form

$$C_1(t) = \max(C_1^0 + \gamma_{\rm s}t, 0.01) \tag{7.22}$$

$$C_2(t) = \max(C_2^0 + \gamma_{\rm f}t, 0.01) \tag{7.23}$$

where $\gamma_{\rm s} > 0$ and $\gamma_{\rm f} < 0$ are pumping and damping rates with the dimension of frequencies. We keep a minimum value of $C_{1,2} = 0.01$ to prevent $C_{1,2}$ from becoming negative, since in both cases oscillations are present up to the end of the simulations (indicating that both the modes maintain an amplitude $\neq 0$). The sign of $\gamma_{\rm s,f}$ is chosen on the basis of what we observe in our simulations, i.e. a transition from the fast to the slow mode.

Solutions like eq. (7.21) imply that the eccentricity $|E_{\rm p}| = e_{\rm p}$ and $|E_{\rm d}| = e_{\rm d}$ of planet and disc are

$$|E_{\rm p}| = \sqrt{C_1^2(t)\eta_{\rm s}^2 + C_2^2(t)\eta_{\rm f}^2 + 2C_1(t)C_2(t)\eta_{\rm s}\eta_{\rm f}\cos(\Delta\omega t)}, \tag{7.24}$$

$$|E_{\rm d}| = \sqrt{C_1^2(t) + C_2^2(t) + 2C_1(t)C_2(t)\cos(\Delta\omega t)}, \tag{7.25}$$

where $\Delta\omega = \omega_{\rm f} - \omega_{\rm s}$. It becomes clear that the simulataneous presence of two eigenmodes produces in the eccentricity some typical oscillations with a periodicity equal to the beat frequency of the two precession rates $\Delta\omega$.

The pericentre phase evolution of the planet $\Phi_{\rm p}(t)$ and of the disc $\Phi_{\rm d}(t)$ is given by

$$\Phi_{\rm p} = \mathrm{mod}\Bigg\{\frac{\omega_{\rm s}}{2}t + \frac{\omega_{\rm f}}{2}t + \arg\Bigg[(\eta_{\rm s}C_1(t) + \eta_{\rm f}C_2(t))\cos\left(\frac{\Delta\omega}{2}t\right) + \tag{7.26}$$
$$+ i(\eta_{\rm f}C_2(t) - \eta_{\rm s}C_1(t))\sin\left(\frac{\Delta\omega}{2}t\right)\Bigg], 2\pi\Bigg\},$$

$$\Phi_{\rm d} = \mathrm{mod}\Bigg\{\frac{\omega_{\rm s}}{2}t + \frac{\omega_{\rm f}}{2}t + \arg\Bigg[(C_1(t) + C_2(t))\cos\left(\frac{\Delta\omega}{2}t\right) + \tag{7.27}$$
$$+ i(C_2(t) - C_1(t))\sin\left(\frac{\Delta\omega}{2}t\right)\Bigg], 2\pi\Bigg\}.$$

However, more relevantly, for any complex number in the the form $E = \mathcal{A}e^{i\omega_{\rm s}t} + \mathcal{B}e^{i\omega_{\rm f}t}$, where $\mathcal{A}$ and $\mathcal{B}$ here represent the mode strength ($C_{1,2}(t)\eta_{\rm s,f}$ for the planet or $C_{1,2}$ for the disc), it can be shown that the phase $\arg(E) = \Phi$ can be approximated by

$$\Phi \approx \begin{cases} \omega_{\rm s}t + \dfrac{\mathcal{B}}{\mathcal{A}}\sin(\Delta\omega t), & \text{if } \mathcal{B} \ll \mathcal{A} \\ \omega_{\rm f}t - \dfrac{\mathcal{A}}{\mathcal{B}}\sin(\Delta\omega t), & \text{if } \mathcal{A} \ll \mathcal{B} \end{cases}, \tag{7.28}$$

implying that whether a component is predominantly precessing at the slow or fast rate (and what is the amplitude of superposed oscillations on this mean precession rate) is determined by the relative values of $\eta_s C_1(t)$ and $\eta_f C_2(t)$ for the planet solution, and by $C_1(t)$ and $C_2(t)$ for the disc one. The variation in time of these parameters implies that the system might experience a transition from the dominance of one mode to the other. This might occur at different times in the planet and in the disc depending on the absolute value of the eigen-vectors. It is important to stress that, for a given configuration, one component can be dominated by one mode while the other not. In fact, the ratio of the amplitude of the fast and the slow modes is given by

$$\mathcal{R}_p = \frac{\eta_f C_2}{\eta_s C_1} \tag{7.29}$$

and

$$\mathcal{R}_d = \frac{C_2}{C_1} \tag{7.30}$$

for the planet and the disc, respectively. Since $\eta_f \neq \eta_s$, the ratio of the two amplitudes can be different in the two components of the system. In particular, for small disc masses (small q), $\eta_f \ll \eta_s$ (see Fig. 7.10) and thus the planet can be in the slow mode while the disc resides in the fast mode.

It is important at this stage to notice that the sign of the eigen-vectors sets the disc-planet configuration of the pericentre precession: in our formulation of the toy model we used eigen-vectors of the form $(\eta_{s,f}, 1)$. For the slow mode, since $\eta_s > 0$, as we have seen in the previous section, both the disc and the planet component of the eigen-vector are positive; thus, with reference to Eq. (7.28), when both the planet and the disc satisfy $\mathcal{B} \ll \mathcal{A}$ they will precess at the slow precession rate with the disc and planet pericentres aligned. In contrast, for the fast mode, since $\eta_f < 0$, when both planet and disc complex eccentricities satisfy $\mathcal{A} \ll \mathcal{B}$ the planet and disc will precess at the fast rate with the pericentres anti-aligned. It follows that when the transition from the fast to the slow mode has been completed in both the planet and in the disc, we will observe also a transition from an anti-aligned to an aligned configuration of the disc-planet pericentre precession.

The initial values of C_1^0 and C_2^0 depend on the initial evolution of the system and, as previously said, cannot be predicted *a priori*. As mentioned in Sec. 7.1, previous works found at short timescales a dependence of the growth rate and saturation value of the planet eccentricity on three main parameters: the planet mass, the disc mass and the initial value of the planet orbital eccentricity (Papaloizou et al. 2001; D'Angelo et al. 2006; Dunhill et al. 2013; Duffell & Chiang 2015).

7.5 Interpretation in the light of the toy model

From Eq. (7.24) and (7.25) the relative amplitude of the oscillations between the planet and the disc is set by $\eta_s \eta_f = -q/\sqrt{\alpha}$. Since $\eta_s \eta_f < 0$, the oscillations in the eccentricity between the planet and the disc are anti-phased, as can be noticed in Fig. 7.6. This enables the conservation of the total AMD across the time of one oscillation as would be expected in pure planet-planet interaction in celestial

mechanics (the non-conservation of the AMD on longer timescales is due to the pumping/damping effects).

Fig. 7.7 shows clearly that at the beginning of both simulations the disc and planet pericentre phases are precessing in an anti-aligned configuration ($\Delta\Phi \approx 180°$). In the light case (left panel of Fig. 7.7), after $\approx 4 \times 10^4$ orbits the planet pericentre phase decouples from the disc one and starts precessing at a much slower rate. The same conclusion can be reached regarding the massive case, but the anti-aligned configuration lasts for much longer and is apparently broken only after $t \gtrsim 2 \times 10^5$ orbits.

In the light case this behaviour can be easily interpreted as the coexistence of the two evolving eigen-modes with positive precession rates $\omega_{\rm s,f}$ predicted in Eq. (7.21), assuming that the following relationships between C_1, C_2, $\eta_{\rm s}$, $\eta_{\rm f}$ hold:

$$C_1\eta_{\rm s} < C_2\eta_{\rm f}, \quad \text{if } t \lesssim 4 \times 10^4\, t_{\rm orb}, \tag{7.31}$$

$$C_1\eta_{\rm s} > C_2\eta_{\rm f}, \quad \text{if } t \gtrsim 4 \times 10^4\, t_{\rm orb}, \tag{7.32}$$

$$C_1 < C_2, \quad \text{if } t \lesssim 2 \times 10^5\, t_{\rm orb}, \tag{7.33}$$

$$C_1 > C_2, \quad \text{if } t \gtrsim 2 \times 10^5\, t_{\rm orb}. \tag{7.34}$$

Note that in the light case $q \ll \sqrt{\alpha}$ and thus $\eta_{\rm f} \ll \eta_{\rm s}$.

In the massive case, the same interpretation can be given but with different times delimiting the different stages in the modes evolution

$$C_1\eta_{\rm s} < C_2\eta_{\rm f}, \quad \text{if } t \lesssim 2 \times 10^5\, t_{\rm orb}, \tag{7.35}$$

$$C_1\eta_{\rm s} > C_2\eta_{\rm f}, \quad \text{if } t \gtrsim 2 \times 10^5\, t_{\rm orb}, \tag{7.36}$$

$$C_1 < C_2, \quad \text{if } t \lesssim 2 \times 10^5\, t_{\rm orb}, \tag{7.37}$$

$$C_1 > C_2, \quad \text{if } t \gtrsim 2 \times 10^5\, t_{\rm orb}. \tag{7.38}$$

Note that here $q \gtrsim \sqrt{\alpha}$ and thus $\eta_{\rm f} \gtrsim \eta_{\rm s}$.

The time at which the transition between the fast and slow modes occurs depends both on the conditions after the initial transient and on the parameters q and α involved in the model.

Although we do not know *a priori* what are the conditions in the disc at the end of the initial transient stage, we can use the picture of the mode structure outlined in Sec. 7.4.2 in order to understand how the evolution of the two simulations differs on account of different values of q (effective disc to planet mass ratio) on long timescales. The smaller the mass of the disc, the more angular momentum is concentrated in the planet, which means that the slow mode has a larger contribution from the planet eccentricity. As the fast mode is damped, therefore, the planet can be expected to make the transition to being predominantly in the slow mode at an earlier evolutionary phase in the light case than in the massive simulation. This is consistent with what we observe in our simulations: in the light case the planet makes the transition to the slow precession mode at an earlier evolutionary stage (see Fig. 7.7).

7.5.1 Evolution depending on the disc mass

We now illustrate how the difference between the two simulations can be understood purely in terms of the dependence of the eigen-vectors on q. In Fig. 7.11 and

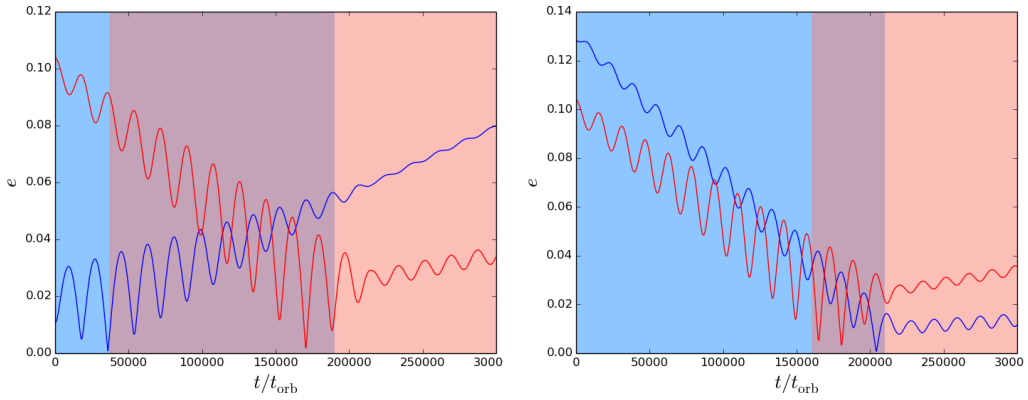


Figure 7.11: Planet (blue curve) and disc (red curve) eccentricity using Eq. (7.24,7.25). The left panel shows a reasonable choice of parameters for the light case, while the right panel a choice for the massive case. The summary of the parameters used can be found in Table 7.1. The evolution of C_2 stops when $C_2 = 0.01$ is reached, as prescribed in Eq. (7.22,7.23), in order to prevent C_2 from becoming negative. Notice that the initial conditions C_1^0 and C_2^0 and pumping/damping coefficients $\gamma_{\rm s}$ and $\gamma_{\rm s}$ are the same in the two simulations, while $\omega_{\rm s}$ and $\omega_{\rm f}$ were chosen in order to reproduce the behaviour in the simulations. The blue shaded area of the plots marks the time region where the fast f mode is dominant in both the planet and the disc, causing the pericentres to precess at the fast rate in the anti-aligned configuration; the red shaded are marks the time region where the slow s mode is dominant in both the planet and the disc, causing the pericentres to precess at the slow rate in the aligned configuration; the violet area marks the region where the slow mode is dominant in the planet but not in the disc, causing a decoupling of the precession rates. These figures are not meant to reproduce precisely the eccentricity evolution in Fig. 7.6 but to show that a change in the values of the eigenvectors produced by a different q (while keeping fixed all the other relevant parameters) can give rise to very different evolutionary path of the system.

	q	α	C_1^0	C_2^0	$\gamma_{\rm s}/\Omega_{\rm p}$	$\gamma_{\rm f}/\Omega_{\rm p}$
light case	0.2	≈ 0.25	0.004	0.1	1.60×10^{-8}	-7.15×10^{-8}
massive case	0.65	≈ 0.25	0.004	0.1	1.60×10^{-8}	-7.15×10^{-8}
	$\omega_{\rm s}/\Omega_{\rm p}$		$\omega_{\rm f}/\Omega_{\rm p}$		$\eta_{\rm s}$	$\eta_{\rm f}$
light case	7.95×10^{-6}		6.35×10^{-5}		2.3	-0.2
massive case	-1.60×10^{-5}		6.35×10^{-5}		0.4	-1.3

Table 7.1: Summary of the parameters used to produce Fig. 7.11 and 7.12.

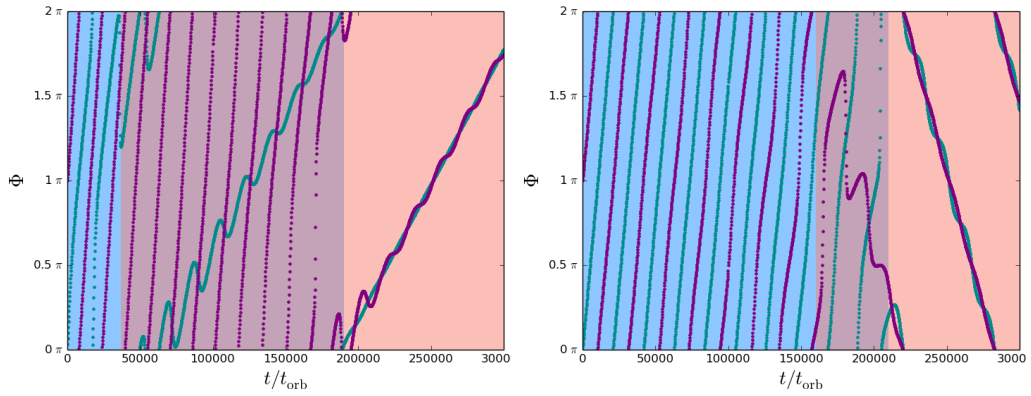


Figure 7.12: Planet (blue curve) and disc (red curve) pericentre phase using Eq. (7.26,7.27). The left panel shows a reasonable choice of parameters for the light case, while the right panel a choice for the massive case. The summary of the parameters used can be found in Table 7.1. The evolution of C_2 stops when $C_2 = 0.01$ is reached, as prescribed in Eq. (7.22,7.23), in order to prevent C_2 from becoming negative. Notice that the initial conditions C_1^0 and C_2^0 and pumping/damping coefficients $\gamma_{\rm s}$ and $\gamma_{\rm s}$ are the same in the two simulations, while $\omega_{\rm s}$ and $\omega_{\rm f}$ were chosen in order to reproduce the behaviour in the simulations. The blue shaded area of the plots marks the time region where the fast f mode is dominant in both the planet and the disc, causing the pericentres to precess at the fast rate in the anti-aligned configuration; the red shaded are marks the time region where the slow s mode is dominant in both the planet and the disc, causing the pericentres to precess at the slow rate in the aligned configuration; the violet area marks the region where the slow mode is dominant in the planet but not in the disc, causing a decoupling of the precession rates. These figures are not meant to reproduce precisely the eccentricity evolution in Fig. 7.6 but to show that a change in the values of the eigenvectors produced by a different q (while keeping fixed all the other relevant parameters) can give rise to very different evolutionary path of the system.

7.12 we plot Eq. (7.24,7.25) and Eq. (7.26,7.27) with a set of parameter choices for each simulation that qualitatively reproduce the main evolutionary features observed in our simulations (Fig. 7.6 and 7.7).

In both Fig. 7.11 and 7.12 we prescribe the same pumping/damping prescription: a linearly decreasing fast mode ($\gamma_{\rm f} = -7.15 \times 10^{-8}\,\Omega_{\rm p}$) and a growing slow mode ($\gamma_{\rm s} = 1.60 \times 10^{-8}\,\Omega_{\rm p}$). We choose $\omega_{\rm s,f}$ to be consistent with those observed in the simulations. The summary of the parameters used in Fig. 7.6 and 7.7 can be found in Table 7.1.

This choice implies $\eta_{\rm s}/\eta_{\rm f}$ changes from being > 1 to < 1 between the simulations. We notice from Fig. 7.10 that this implies that $q/\sqrt{\alpha}$ is respectively lower and greater than 1 (i.e. that the circular angular momentum is mainly in the planet in the light case and mainly in the disc in the massive case). We choose $C_1^0 \ll C_2^0$ for both our simulations (see caption to Fig. 7.11 and 7.12). Noting that the "density edge of the cavity" is located at $R \approx 4.5$ in both simulations we adopt $\alpha \approx 0.25$ in both cases. The values of q in the two simulations are 0.2 and 0.65 for the light and massive case, respectively.

Obviously our simulations are not supposed to share the initial conditions and mode pumping/damping rates, in contrast with what we prescribed. However, we remark that Fig. 7.11 and 7.12 are not meant to reproduce precisely the eccentricity evolution in Fig. 7.6. We find it more instructive to show a comparison between the two regimes while keeping fixed all the other relevant parameters in order to highlight the role of the mass ratio q in determining the evolution of the system.

We believe that these images show clearly that a change in the values of the eigenvectors produced by a different q (while keeping fixed all the other relevant parameters) can give rise to very different evolutionary path of the system, despite our inability to model properly the evolution during the initial phases of the simulation. Furthermore, we see from Fig. 7.11 and 7.12 that this simple parametrisation does an extraordinarily good job of reproducing the main features of the planet and disc eccentricity evolution on long timescales (compare with Fig. 7.6 and 7.7).

In the light case the planet makes the transition to the slow mode significantly before the disc because the simulation is in the regime $q \ll \sqrt{\alpha}$ where the two eigenvectors are very different. In the massive case the value of $q/\sqrt{\alpha}$ is closer to 1 and so the eigenvectors are more similar to each other. Consequently the disc and the planet follow more similar evolution of the eccentricity and indeed they make the transition to being predominantly in the slow mode at nearly the same time. By the end of the simulation both simulations are mainly in the slow mode and hence the ratio of eccentricities is given simply by the eigenvector of the slow mode.

The main conclusion that can be drawn from these considerations, is that, *assuming* a damping of the fast mode, the system will end up in a configuration where $e_{\rm p} > e_{\rm d}$ for all those cases in which $q/\sqrt{\alpha} = J_{\rm d}/J_{\rm p} < 1$ (light discs) and vice versa $e_{\rm p} < e_{\rm d}$ for all those cases in which $q/\sqrt{\alpha} > 1$ (massive discs). As a consequence, under this assumption one should expect that low mass discs favour the growth of the eccentricity at long timescales. We caution however

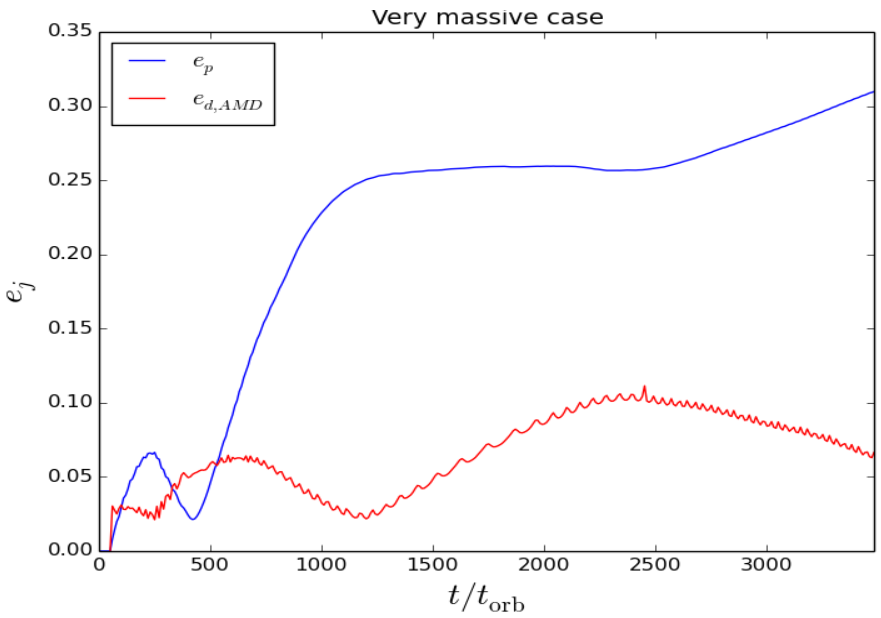


Figure 7.13: Eccentricity evolution of planet (blue curve) and disc (red curve, AMD based) during the early stages of the simulation for a $q = 2$ disc-planet system. This behaviour is consistent with what was observed in the previous work by Papaloizou et al. (2001). However our analysis suggests that this is a transient growth phase, and we expect the eccentricity to decay on a longer timescale

that the assumption about the fast mode damping is tentative since it is based only on the two simulations we performed. A larger number of simulations is required to address the reliability of this assumption and the direction that further investigations should take.

It should be noted that our simulations show that higher disc masses can pump higher levels of planetary eccentricity at short timescales (as previously pointed out by Dunhill et al. 2013). In addition, we have carried a third simulation with a much more massive disc ($q = 2$), in order to compare with the previous results obtained by Papaloizou et al. (2001), finding that such a high disc mass allows the planet to reach eccentricities as high as $e = 0.3$ during the first 3×10^3 orbits (see Fig. 7.13), consistently with what they previously found. This shows that the eccentricity of the planet acquired over hundreds of orbits does indeed increase with disc mass as one would naively expect but, if our analysis holds for such massive discs, we expect the eccentricity to decay on a longer timescale (10^5 orbits).

We do not attempt to run this simulation for as many orbits as the two cases we presented in this chapter. Due to the required computational resources, it is beyond the scope of this project to further verify whether massive discs lead to more damping of the planet on long timescales, but future work should address this statement.

7.5.2 Effects of disc viscosity and thickness

Disc viscosity (ν) and thickness (H/R) are expected to play a role in the eccentricity evolution of the disc-planet system.

The effects of viscosity might act to either increase or damp the eccentricity. On the one hand, an increase in the disc viscosity implies stronger damping effects (Teyssandier & Ogilvie 2016), providing thus an overall faster decrease of the system eccentricity. On the other hand, more viscous discs have smaller cavities, modifying the surface density at resonant locations and thus produce a stronger resonant interaction and faster evolution of the growing/decreasing trend. The pumping or damping nature of this latter effect depends on the type of resonances (Lindblad or co-rotation) that are strengthened.

Regarding the disc thickness, a higher H/R also provides a narrower gap, hence stronger resonant interaction again. Furthermore, the resonance width is broadened by pressure effects, and scales as $(H/R)^{2/3}$, so higher H/R implies that resonances operate on a broader disc region, which also increases the growth/decrease rate. Higher H/R implies a faster propagation speed of the mode, making the growth of a trapped slow mode more difficult in the inner regions (Teyssandier & Ogilvie 2016). Finally, larger H/R increases the effect of pressure, which drives retrograde precession. This opposes the gravitational secular interaction, which drives prograde precession.

Finally, large eccentricity gradients imply large fluid relative velocities, that approach the sound speed when $Rde/dR \sim H/R$ and produce the crossing of fluid trajectories when $Rde/dR \sim 1$ (Ogilvie 2001), possibly inducing shocks and large pressure gradients that might limit the further growth of the disc eccentricity. Whether this implies that the planet eccentricity scales with H/R (Duffell & Chiang 2015) still needs to be addressed with further work, although note that in our simulations we reach planet eccentricities ≈ 0.1, well in excess of the value of $H/R \approx 0.036$.

A further exploration of the parameter space is required to address the dependence of the evolution on these parameters.

7.5.3 Effects of the disc evolution

While our toy model reproduces the qualitative features we have highlighted so far, as already noticed in Sec. 7.3 a careful inspection of Fig. 7.6 reveals that the oscillation frequency of the eccentricity is not fixed in time. This effect cannot be captured within our modelling but it indicates that the viscous evolution of the disc has a role in determining the precession rate (eigen-frequencies) and modes relative strength (eigen-vectors): as the disc viscously spreads the effective q and α change, causing an evolution of the eigenfrequencies. The accretion of material at the inner edge causes a further decrease of the effective q. In addition, it is very important to remember that the eccentricity values of the planet have an effect in determining the size and the density profile of the cavity edge (Artymowicz & Lubow 1994; Thun et al. 2017) causing a variation of the effective α.

As a consequence, it should be noticed that any system naturally evolves toward a situation where $q/\sqrt{\alpha} < 1$ due to the progressive disc dispersal. This implies

the existence of a period in which the planet eccentricity grows above that of the disc. However the growth of the eccentricity occurs on a very long timescale, thus if the disc disperses too rapidly this final growth might not occur at all.

In the massive case the disc switches to the slow mode after 2×10^5 orbits. The dominance of this single mode produces an abrupt change in the disc eccentricity and density profile. In particular, in this mode the resonant region is very depleted. This not only explains the stalling of migration but also the fact that the eccentricity stops growing and is then subject just to very slow damping for the last $\sim 10^5$ orbits (likely due to viscous effects in the disc).

7.6 Conclusions

We performed two long time scale 2D hydrodynamical simulations of a planet embedded in a gaseous disc using two different disc masses (light case and massive case), in order to study the long term evolution of both the planet and disc eccentricities.

The disc-planet interaction induces an eccentricity exchange between the planet and the disc in the form of periodic oscillations of both planet and disc eccentricity superimposed on a growing or decreasing trend depending on the disc mass.

In the light disc case the planet eccentricity, after an apparent stalling of its evolution, grows linearly with time up to $e_\mathrm{p} \approx 0.12$ (reached after 3×10^5 orbits). After 2×10^5 orbits the growth appears to slow down, probably because of some saturation effects. The disc eccentricity rapidly reaches $e_\mathrm{d} \approx 0.1$ at the beginning of the simulation and then decreases linearly. At the end of the simulation the planet-to-disc eccentricity ratio is $(e_\mathrm{p}/e_\mathrm{d})_\mathrm{light} \approx 3$.

In the massive case instead the planet eccentricity grows exponentially up to $e_\mathrm{p} \approx 0.14$ during the initial phases of the simulation but then linearly decreases as a function of time. As in the low mass case, in the massive case the disc eccentricity grows rapidly during the initial phases of the simulation up to $e_\mathrm{d} \approx 0.1$ and then decreases linearly as a function of time. In this case the planet eccentricity exceeds the disc one ($e_\mathrm{p}/e_\mathrm{d} > 1$) up to $t \approx 2 \times 10^5$ orbits, when a rapid transition to $e_\mathrm{p}/e_\mathrm{d} \approx 0.3$ occurs.

In our simulations we find that the planet eccentricity can reach values $e_\mathrm{p} \gtrsim 0.1$, well in excess of the value of $H/R \approx 0.036$ suggested by Duffell & Chiang (2015) as the maximum value of the eccentricity. Furthermore, Goldreich & Sari (2003) and Duffell & Chiang (2015) have argued for the need of a non-zero initial eccentricity of the planet to ensure the saturation of the corotation torque, and subsequent growth of the eccentricity. In our simulations, both the planet and the disc are initially on circular orbits, suggesting that corotation resonances might not necessarily need to be saturated to cause eccentricity growth.

We interpret the coupled evolution of the planet and disc eccentricity in terms of a superposition of secular modes whose relative amplitudes are slowly modified by resonant pumping and viscous damping. These modes are generically a rapidly precessing mode with anti-alignment between disc and planet pericentres and a slowly precessing aligned mode. The ratio of disc to planet eccentricity in each of these modes is generically > 1 and < 1, depending on the ratio between the

planet and disc angular momenta. The dominance of the disc eccentricity in the fast mode implies that viscous damping preferentially damps the fast mode. The system thus ends up being completely in the slow mode at the end of the simulation but following very different evolutionary paths.

At the end of the low mass simulation, the planet is describable as being primarily in the slow mode which is growing very slowly whereas the disc is executing high frequency low amplitude libration about this slow precession. The high mass simulation has instead evolved to a situation where the planet is on an almost circular orbit with stalled migration and the eccentric disc undergoes retrograde precession due to pressure effects. At the end of this simulation the region encompassing the Lindblad resonances has been cleared of material so failing re-supply of this region (by accretion from the outer disc), the orbital evolution of the planet is stalled in an almost circular orbit.

We provide a simplified toy model in which we treat the disc as a second "virtual" planet undergoing the secular interaction with the real one. This model depends only on two variables $q = M_\mathrm{d}/M_\mathrm{p}$ and $\alpha = a_\mathrm{p}/a_\mathrm{d}$, predicts the presence of two eigen-modes with respective eigen-values (setting the precession rate) and eigen-vectors (setting the ratio $e_\mathrm{p}/e_\mathrm{d}$).

Under the same initial conditions and pumping/damping prescription, we are able to qualitatively reproduce with our toy model the two very different evolutions of eccentricity and pericentre phase in the light case and massive one (Fig. 7.11 and 7.12, to be compared with Fig. 7.6 and 7.7). The different behaviour of the two simulations can be understood in terms of different eigenvectors characterizing the fast and slow precession modes. Indeed in Eq. (7.20) (see also Fig. 7.10) predicts that, for the same α, low values of q produce a slow mode with $e_\mathrm{p} > e_\mathrm{d}$, high values of q predicts $e_\mathrm{p} < e_\mathrm{d}$. This is in perfect agreement with the outcome of our simulations.

Simply requiring a different mass ratio between the "virtual" planet and the real one the toy model is able to explain:

- Oscillations in the eccentricity.
- Transition from the fast to the slow mode.
- Aligned and anti-aligned configurations of the pericentre precession.
- Faster transition to the slow mode of the planet in the light case than in the massive one.
- Final values of the eccentricity when the system is fully in the slow mode.

Our model is not able to capture the intrinsic nature of the pumping and damping mechanism and thus it cannot be used as predictive tool to determine under which conditions the slow mode grows or decreases the eccentricity. However, we expect the disc thickness and viscosity to have a role in determining the intensity of the pumping and damping mechanisms as pointed out in Sec. 7.5.2. The absolute mass of the planet might be relevant as well affecting the strength of the resonant disc-planet interaction. In contrast it provides larger cavities, and some resonances might be saturated.

Nevertheless, the model predicts some useful relationships that apparently hold between the planet and disc eccentricity depending on the disc-to-planet mass ratio and disc cavity size.

If we assume that the damping of the fast mode on very long timescales is a general result, massive discs appear to disfavour values of planet eccentricity higher than those in light discs at late stages of their evolution. For relatively high $M_\mathrm{d}/M_\mathrm{p}$, and a disc density profile comparable with that produced by a $13\,M_\mathrm{J}$ planet, the system ends up in a slow mode configuration characterized by $e_\mathrm{p}/e_\mathrm{d} < 1$. In contrast, light discs are expected to produce a slow mode with $e_\mathrm{p}/e_\mathrm{d} > 1$, in fact favouring higher values of e_p. This goes in the opposite direction of what is often found in the literature, where high planet eccentricities have been observed to develop in presence of high disc masses on short timescales (Papaloizou et al. 2001): they found that a $30\,M_\mathrm{J}$ planet can reach $e_\mathrm{p} \approx 0.3$ in less than $10^3\,t_\mathrm{orb}$ with $M_\mathrm{d}/M_\mathrm{p} \approx 2$. To support the results by Papaloizou et al. (2001), we report that some preliminary simulations we performed reached eccentricity values as high as $e_\mathrm{p} \approx 0.3$ for $M_\mathrm{d}/M_\mathrm{p} \approx 2$ after $t \approx 10^3 t_\mathrm{orb}$ (see Fig. 7.13). If the damping of the fast mode at long timescales is confirmed to be a general feature, this implies in fact a reversal of the dependence on the disc mass of the eccentricity evolution on short timescales.

We caution that our last claim is tentative and supported only by two numerical simulations. A larger number of simulations exploring a wider range of disc masses is required in order to properly investigate the issue. In general, a more complete understanding of the origin of the mechanisms for pumping or damping of the eigen-modes is obviously required in order to make quantitative predictions about the eccentricity evolution at very late times and constitutes a possible follow up of this work.

It should be also considered that in principle any planetary system passes through a phase in which the ratio $M_\mathrm{d}/M_\mathrm{p} \ll 1$, due to the progressive disc dispersal and accretion of material on the planet, which increases its mass. If we for now adopt the assumption that planets attain larger eccentricities in the case of *lower* mass discs, then it raises some interesting possibilities about planet eccentricity evolution during disc dispersal. Whereas rapid dispersal (i.e. on a timescale $\ll 10^5$ orbits) would simply freeze the planet's eccentricity at its previous value, slow dispersal could instead cause the planet eccentricity to rise in the last stages. However, further investigations are required in order to understand how the secular eccentricity evolution is affected by substantial changes in the disc parameters throughout its lifetime.

Finally, we believe that this work demonstrates the importance of carrying out long timescale simulations when studying the disc-planet interaction in protoplanetary discs: both our simulations indeed undergo a complete inversion of the evolutionary trend on long timescales with respect to those shown in the initial phases. This however does not occur before 5×10^4 orbits, which is beyond the timescales explored in previous simulations.

Chapter

8

Transition Discs

"...The goal has never been to be famous,
my goal has been to be a star."

Lady Gaga, (...but it suits transition discs)

The recent advent of the Atacama Large Millimiter Array (ALMA) and the Spectro-Polarimetric High-contrast Exoplanet REsearch (SPHERE) instrument, mounted on the Very Large Telescope (VLT), has allowed the observation of a large number of sources in the nearby star forming regions (Taurus, Lupus, Ophiucus, Upper Scorpius, Orion). The extensive imaging campaigns conducted in these regions resolved with unprecedented detail a large number of protostellar discs, revealing a variety of structures such as: spirals (Muto et al. 2012; Grady et al. 2013; Garufi et al. 2013; Benisty et al. 2015; Pérez et al. 2016; Stolker et al. 2016; Benisty et al. 2017), shadows (Garufi et al. 2014; Avenhaus et al. 2014; Benisty et al. 2017; Avenhaus et al. 2017), circular gaps (ALMA Partnership et al. 2015; Hendler et al. 2018; Fedele et al. 2018; Dipierro et al. 2018a), cavities, rings and non-axisymmetric features (Dutrey et al. 2008; Brown et al. 2009; Andrews et al. 2011; Isella et al. 2013; Zhang et al. 2014; van der Marel et al. 2016b; Canovas et al. 2016; Fedele et al. 2017; Pinilla et al. 2017; van der Marel et al. 2018; Casassus et al. 2018; Long et al. 2018; Liu et al. 2018).

Protoplanetary discs presenting cavities are generally referred to as "transition discs" (also "transitional" is used). These systems are a special category of class II accretion discs (see Sec. 1.4) surrounding young pre-main sequence stars. They were originally thought to bridge class II and class III objects, where the disc is progressively dispersed as the system ages. This interpretation has been recently put into question, following the observation of these features also in very young systems.

Many of these systems present very large horseshoe structures at the edge of their cavity in the dust continuum, and spiral density features in the infrared, that have been suggested to be hints of the presence of forming planets.

In this chapter we will review both the observations and possible physical origin of the cavities and structures in these sources, emphasising the role planet disc interaction may play for their origin; in the next chapter we will better discuss the models at the basis of the formation of the horseshoe structures, proposing a new scenario that allows us to relax some assumptions currently required in the most widely invoked model.

8.1 Observations of transition discs

The existence of systems with dust cavities comes from four main obervational diagnostics, probing different disc components, that we briefly review in this section: SED in the IR, imaging in the sub-mm and mm continuum (probing mm dust grains), IR imaging (probing μm dust grains) and line emission. However, these systems are still very far from being completely understood: some systems can be identified as transition discs in all the channels, others just in some of them. We refer to the papers by Espaillat et al. (2014), Owen (2016) and Ercolano & Pascucci (2017) for detailed reviews of the observations.

8.1.1 SED surveys in the IR

These objects were identified for the first time by Strom et al. (1989); Skrutskie et al. (1990), due to their very peculiar SED, characterized by a flux reduction at near-mid-infrared wavelengths (NIR/MIR, $\lambda \approx 1$–$20\,\mu$m)[1]. The advent of the Spitzer Space Telescope and, in particular, the usage of the InfraRed Spectrograph (IRS), allowed the characterization of a large number of young star spectra. The general outcome is that a number ranging between 10–20% of the total sources presents these peculiar features in their infrared colours (Luhman et al. 2010; Ercolano et al. 2011).

The electromagnetic flux at those wavelengths is dominated by thermal emission coming from small dust grains with sizes ranging between 1–20 μm, as a consequence of their opacity peak at those wavelengths. This straightforwardly leads us to interpret these peculiar spectra as protostellar systems where some physical mechanisms have partially (pre-transition discs are expected to show an inner ring of dust close to the star), or completely, cleared the inner disc region from small dust grains, causing a change in the optical depth of the emission in those regions of the disc. The presence of the 10μm emission feature (see Fig. 8.1) indicates that inside the hole there is still some micrometric dust (the feature can come only from dust grains with size $\approx 1\,\mu$m). Fig. 8.1 shows the typical SED of transition and pre-transition discs and a sketch of the dust distribution producing them.

The SEDs of many of these objects show a strong UV excess produced by accretion shocks on the stellar surface (Espaillat et al. 2014; Owen 2016), denoting

[1] It should be noted that a sub-category, showing only MIR wavelengths depletion, is referred to as "pre-transition" (Espaillat et al. 2007).

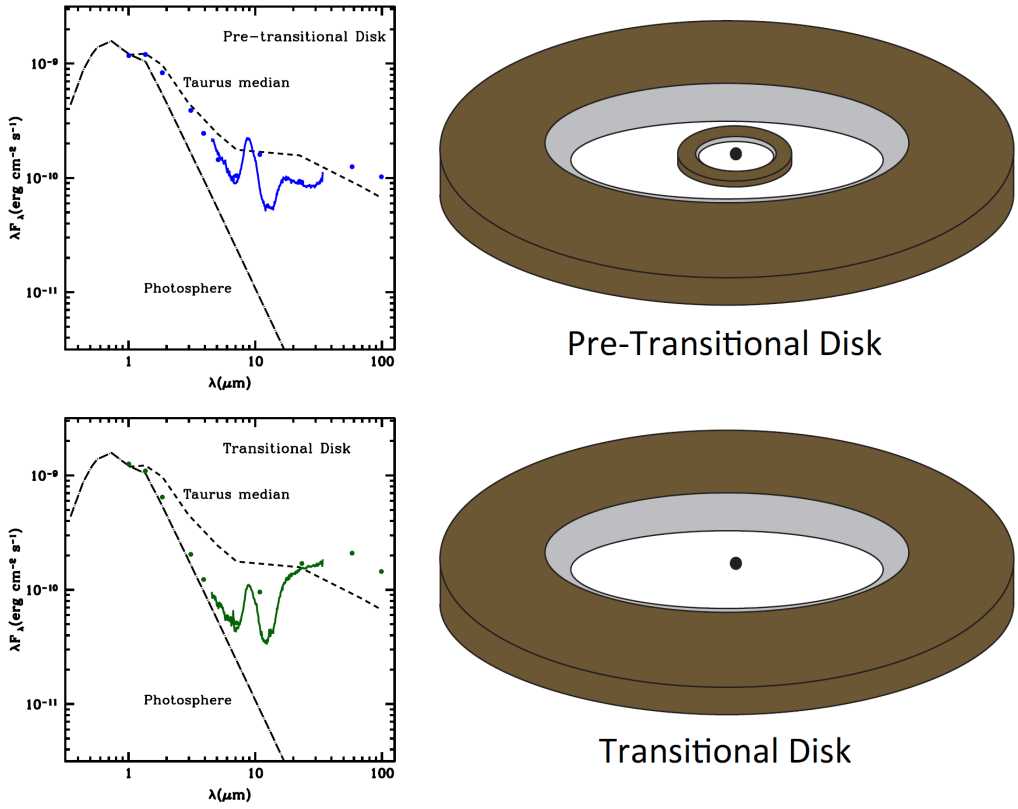


Figure 8.1: Left panels: top, SED from pre-transition disc LKCa 15 (Espaillat et al. 2007); bottom, SED from transitional disc GM Tau (Calvet et al. 2005); the green and blue lines are the data collected using the Spitzer Infrared Spectrograph (IRS), the long dashed lines represent the star photospheric emission, the short dashed lines represents the median Taurus SED (D'Alessio et al. 1999), representing the typical optically thick emission from a disc in the Taurus star forming region. We note the emission bump at $\lambda = 10\,\mu$m called the "10 μm silicate emission feature". Right panels: sketches of the dust distributions for a pre-transition (top) and transitional (bottom) disc: in brown the dust distribution; discs said to be "pre-transition" have an inner ring of optically thick material close to the star that still provides some emission at NIR wavelengths. Image adapted from Espaillat et al. (2014).

that gas is still accreting on to the central star with a relatively high rate. This is surprising: small dust grains are expected to be strongly coupled with the gas; a depletion in the small dust grains in the inner regions of the disc would thus suggest that also the gas is depleted, which is actually not the case.

8.1.2 (sub-)mm dust continuum imaging: probing millimetric dust grains

The advent of radio interferometric facilities such as the Very Large Array (VLA), the Small Millimetre Array (SMA) (Dutrey et al. 2008; Brown et al. 2009; Andrews et al. 2011; Isella et al. 2013), and, in particular, the recent advent of the Atacama Large Millimetre Array (ALMA) (Zhang et al. 2014; van der Marel et al. 2016b; Canovas et al. 2016; Fedele et al. 2017; Pinilla et al. 2017; Fuente et al. 2017; van der Marel et al. 2018), enabled the imaging of these systems at mm and sub-mm wavelengths. One of the observing channels of these facilities is sensible to the dust continuum emission from ∼mm dust grains. The images show holes as large as ≈ 100au in the (sub-)millimetric emission (see Fig. 8.2), confirming in fact the hypothesis that transition discs spectra are caused by an important reduction of dust opacity in the inner regions of the disc, but also a broad variety of non-axisymmetric features whose origin will be the subject of the next chapter (Casassus et al. 2013; Pérez et al. 2014; Casassus et al. 2015a; van der Marel et al. 2016a; van der Plas et al. 2017).

In some sources, the size of the observed cavities grows with the wavelength of the observations (Pinilla et al. 2015, 2017), moreover the azimuthal extent of non-axisymmetric structures is smaller for large dust grains (van der Marel et al. 2013; Casassus et al. 2015b). This has been suggested to be a consequence of the different aerodynamical coupling that depends on the grain sizes. As expected, larger dust grains are more easily trapped in gas pressure maxima (dust trapping, see Sec. 2.7.2).

Andrews et al. (2011) conducted a large imaging survey of bright sources in close star forming regions. They found that at least 1 out of 3 sources presents a cavity, i.e. much more than the previous result of 10–20% obtained by spectroscopic methods surveys (Luhman et al. 2010; Ercolano et al. 2011). The reliability of this result has been put into question since the sample of these surveys is biased toward highly luminous sources. However, more recent surveys of unbiased disc samples appear to confirm that not all cavities are associated with SED features (van der Marel et al. 2018), implying in fact that transition discs are a larger population than thought before, even though they appear to be more common in bright sources.

The detection of (sub-)mm dust cavities in sources that are not showing any feature in the SED can be understood by noting that the 80% of the dust emission at wavelengths $\approx 10\,\mu\text{m}$ comes from within $R \lesssim 1\,\text{au}$ (D'Alessio et al. 2006) so that SED surveys detected dust holes (transitional discs), without being able to detect large gaps (pre-transition discs). Furthermore, it is not straightforward to link (sub-)mm imaging with IR SEDs: observations at (sub-)mm wavelengths probe the optically thin thermal emission of millimetric dust grains, that only partially contribute to the opacity at μm wavelengths. Millimetric dust grains are

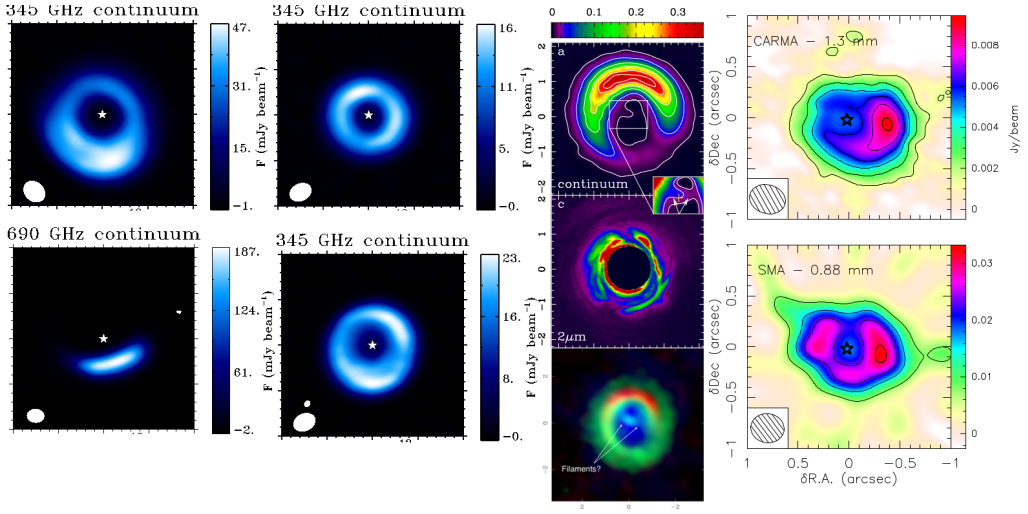


Figure 8.2: Dust continuum observations of several sources (adapted from van der Marel et al. 2016a, Casassus et al. 2013,Casassus et al. 2015a and Isella et al. 2013). Left column (blue style plots): 345 GHz dust continuum images, from top-left to bottom-right panel, HD135344B, DoAr44, IRS48 and SR21 (adapted from original image in van der Marel et al. 2016a). Middle column: 345 GHz ALMA dust continuum image of HD142527 (top panel), 2 μm GEMINI observation of scattered light from the same object (middle panel) (original images from Casassus et al. 2013), summary (bottom panel) of the 345 GHz dust continuum (in red), moment 0 map of ^{12}CO (3-2) emission (in blue), moment 0 map of HCO^+ (4-3) emission (original image from Casassus et al. 2015a). Right column: CARMA 1.3 mm (top) and SMA 0.88 mm (bottom) dust continuum observations of LkHα330 (original image from Isella et al. 2013).

also expected to have a different dynamical behaviour with respect to smaller dust grains, so the depletion of mm grains is not necessarily reflected in the μm grains.

We finally report that two broad families of transition discs can be identified based on their luminosity at mm-wavelengths (Owen & Clarke 2012): one is "mm-bright", associated with massive discs, large accretion rates on to the central star ($\dot{M}_\star \sim 10^{-8}\,M_\odot\,yr^{-1}$) and large cavities ($R_{cav} \gtrsim 20\,au$); the other "mm-faint", associated with low disc mass, low accretion rates $\dot{M}_\star \lesssim 10^{-9}\,M_\odot\,yr^{-1}$ and small cavities $R_{cav} \lesssim 20\,au$. While the second family can be easily associated with discs in the process of being dispersed, the first cannot, possibly suggesting that their origin has to be found in different physical processes.

8.1.3 IR imaging and gas line emission: probing μm dust grains and gas distribution

IR imaging became available after the operation commencement of instruments such as NACO (short for NAOS/CONICA), the High Contrast Instrument for the Subaru next generation Adaptive Optic (HiCIAO), and the Spectro-Polarimetric High-contrast Exoplanet REsearch (SPHERE) instrument mounted on the Very Large Telescope (VLT). These instruments probe the polarized scattered light from μm dust grains in the surface layers of discs at wavelengths 1–10 μm, and

obtain the images subtracting the non-polarized contribution coming from the star photosphere (Polarimetric Differential Imaging, PDI).

Some of the sources with large cavities in the (sub-)mm observations show a very different behaviour at μm wavelengths: although some sources present cavities also for smaller grains, many others, in contrast, show a much shallower depletion or even no depletion at all. Furthermore, sources with cavities in (sub-)mm images, in the IR have been often found to show a set of non-axisymmetric structures, like shadows (Garufi et al. 2014; Avenhaus et al. 2014; Benisty et al. 2017; Avenhaus et al. 2017) or spirals (Muto et al. 2012; Grady et al. 2013; Garufi et al. 2013; Benisty et al. 2015; Stolker et al. 2016; Benisty et al. 2017), see the top row in Fig. 1.3.

As previously mentioned, an indirect evidence of the presence of gas is its high rate of accretion on the central star, detected as UV excess in many of these systems.

The gas distribution can be probed at high resolution by looking at molecular gas rotational lines in the (sub-)mm such as the HCO^+ (4-3), or the CO: ^{12}CO and its isotopologues, ^{13}CO and $C^{18}O$ (Bruderer et al. 2014a; Zhang et al. 2014; Perez et al. 2015; van der Marel et al. 2016a; Canovas et al. 2016; Boehler et al. 2017; Fedele et al. 2017; Tang et al. 2017b), or, less frequently, HCO^+ and its isotopologue DCO^+, HCN and its isotopologue DCN, or N_2H^+ and N_2D^+ (Casassus et al. 2013; van der Plas et al. 2017; Salinas et al. 2017; Kastner et al. 2018). Moreover, the observation of roto-vibrational lines in the IR such as the fundamental CO transition at $\lambda = 4.6\,\mu m$ and ionized NeII $\lambda = 12.8\,\mu m$ (Salyk et al. 2007; Najita et al. 2008, 2009) act as a probe of the presence of the gas in the inner regions ($R \approx 1\,au$) of the disc. As a general result, a discrepancy between the distribution of gas and dust cavities imaged at (sub-)mm and IR wavelengths is present: in particular gas cavities have been generally found to be smaller than those observed in the dust (van der Marel et al. 2016a; Canovas et al. 2016; van der Marel et al. 2018).

Nevertheless, the reader should keep in mind that molecular emission acts as tracer for the gas, but the abundance of molecular species with respect to the hydrogen (that constitutes the most abundant species) throughout the disc might change significantly as a function of the local conditions, so that the lack of emission in some regions might have a chemical origin instead of dynamical.

8.2 Physical mechanisms producing disc cavities

Different mechanisms have been proposed to be the cause of the cavity structure observed in transition discs. Among them, the most credited ones are: dust grain growth (Dullemond & Dominik 2005; Birnstiel et al. 2012), photoevaporation (Clarke et al. 2001; Alexander et al. 2014) or dynamical clearing (implying the interaction of the disc with one or multiple planets/sub-stellar companion Lin & Papaloizou 1979; Zhu et al. 2011, see Chapter 4).

In general, a reliable mechanism must be able to explain at least three main characteristics of transition discs: strong dust depletion in the cavity region with respect to the outer disc, large cavities ($\gtrsim 20$ au), high accretion rates on to the

central star $\gtrsim 10^{-8}\,\mathrm{M_\odot\,yr^{-1}}$.

Finally, the reader should also keep in mind that although all the three mechanisms mentioned before might be responsible, or more generally to concur (Rosotti et al. 2013; Jennings et al. 2018), for the creation of the cavity, only the presence of a planet/companion can possibly explain the non-axisymmetric structures that have been widely observed accompanying transition discs. In this section we briefly review them, in order to highlight their working principle, while we defer to the next section a more thorough discussion regarding the perturbations in the dust and gas density producing the asymmetries.

8.2.1 Dust grain growth

The dust grain population is expected to evolve as the time passes depending on two competing mechanisms: coagulation and fragmentation. The relative motion provided by Brownian motion, the differential settling (i.e., different grain sizes show different rate of relaxation of the dust distribution toward the midplane) and turbulence enable the dust grains to meet each other (Dullemond & Dominik 2005; Birnstiel et al. 2012). Attractive forces, such as van der Waals or electrostatic forces (Dominik & Tielens 1997), stick dust grains together forming larger ones. Collisions with energies greater than the binding energies of the stuck grains cause them to fragment. The balance between these mechanisms sets $a \approx 1\,\mathrm{cm}$ as the maximum size the grains can grow by this mechanism[2] (Windmark et al. 2012).

Numerical simulations by Birnstiel et al. (2012) have shown that most of the mass of the grains is concentrated in larger grains. The growth of dust grains to sizes $a_\mathrm{d} \approx$ mm–cm produces a reduction of the emissivity at μm wavelengths, due to the different opacity function of large dust grains (i.e. the emission is optically thicker at larger wavelengths). This produces a dip in the SED at μm wavelengths that can explain the SED discussed in Sec. 8.1.1. However, in the same work they also note that the abundance of larger grains would still be detectable at mm wavelengths, failing in fact to explain the large cavities observed by ALMA and other facilities in the (sub-)mm.

8.2.2 Photoevaporation

Photoevaporation is a physical process where discs lose material through photoevaporative winds (Clarke et al. 2001; Alexander et al. 2014). High energy photons coming from the central star heat up the gas surface layers of the disc. Three main photon wavebands have been identified for their different heating mechanisms, far-ultraviolet (FUV, $E_\mathrm{phot} \approx 6$–$13.6\,\mathrm{eV}$), extreme-ultraviolet (EUV, $E_\mathrm{phot} \approx 13.6$–$100\,\mathrm{eV}$) or X-ray ($E_\mathrm{phot} \approx 0.1$–$10\,\mathrm{keV}$). In this process the disc surface temperature can grow significantly up to $T_\mathrm{hot} \approx 10^3$–$10^4\,\mathrm{K}$. The rise in the gas temperature makes the surface layers gravitationally unbound, so that they "evaporate" producing a photoevaporative wind and contributing to the disc dispersal at the end of the disc life.

[2]It should be noted that still an effective mechanism able to grow the dust grain size from cm to m has not been provided yet, see Sec. 1.4.2.2.

The regions farther from the star are the most susceptible to this mechanism, for this reason we can define the critical radius $R_{\rm crit}^{\rm phot}$ where the material becomes gravitationally unbound

$$R_{\rm crit}^{\rm phot} \simeq \frac{GM_\star}{c_{\rm s,hot}^2} \approx 1.8 \left(\frac{M_\star}{M_\odot}\right) \left(\frac{T_{\rm hot}}{10^4}\right)^{-1} {\rm au}, \tag{8.1}$$

where $c_{\rm s,hot}$ is the sound speed in the heated surface layers.

In this picture, numerical studies of the viscous evolution of discs accounting for photoevaporative winds have shown that the disc dispersal occurs as an inside-out process (Alexander et al. 2006; Gorti et al. 2009; Alexander et al. 2014). At early times, the disc accretion rate on to the central star is much higher than that at which the disc photoevaporates; as the disc evolves, material accretes on to the central star and spreads at larger radii. After a few Myr its mass decreases and the accretion rate becomes comparable to the mass loss due to photoevaporation. At this stage, for radii $R \gtrsim R_{\rm crit}^{\rm phot}$ the inward motion of the material is significantly slowed down, failing to supply the inner region that instead is still accreting at normal rate. This causes the inner disc to be rapidly ($\approx 10^5\,{\rm yr}$) emptied creating a disc cavity both in the gas and in the dust. During this process the dust grains rapidly drift on to the central star leaving behind a purely gaseous inner disc (that is still accreting[3]) and a dust hole. The disc is finally completely dispersed in $\approx 10^5\,{\rm yr}$.

The short time the disc takes in order to accrete the material, form a cavity and then to finally disperse the disc compared to the entire life of the accretion disc implies that the observation of systems exactly in this phase of their evolution is statistically unlikely.

Such an evolutionary scenario for discs naturally provides the formation of a cavity in the dust; furthermore, it seems to be consistent with the statistical constraints on the transition discs population. Nevertheless, photoevaporation alone can explain only relatively small cavities (up to $R_{\rm cav} \approx 20\,{\rm au}$) and low accretion rates ($\approx 10^{-9}\,M_\odot{\rm yr}^{-1}$). The observation of higher accretion rates on to the central star, the large cavities observed in these systems and the rapid disc dispersal after the carving of the cavity all appear to be in contrast with the hypothesis that photoevaporation is the main physical process producing transition discs (Owen et al. 2011), we refer to Sec. 3.1 in Owen (2016) for a more thorough discussion.

However, despite not being responsible alone for the creation of transition discs, photoevaporation still remains an important ingredient to be considered in other scenarios such as the dynamical clearing by planets that we will discuss in the next section (Rosotti et al. 2013; Jennings et al. 2018).

8.2.3 Dynamical clearing

As discussed in Chapter 4, the presence of secondary companions such as one (or multiple) planets, or a second binary star, are expected to carve a gap or a cavity in the accretion disc. This idea comes directly from the studies conducted

[3]However, the viscous accretion rate must be smaller than the photoevaporation in order to clear the cavity, i.e. the viscous accretion rate must satisfy $\dot{M} \lesssim 10^{-8}\,M_\odot{\rm yr}^{-1}$.

starting from the late '70s about the disc-satellite interaction (Lin & Papaloizou 1979; Goldreich & Tremaine 1980) and it is very well suited for explaining the dust cavities, but also spirals and non axisymmetric features detected in transition discs.

In general, what is most appealing in the dynamical clearing hypothesis is that, as discussed in section 2.7.2, the trapping of dust grains in pressure maxima provides an effective way to explain the discrepancy between the dust and gas cavity size: large dust grains concentrate in the gas pressure maximum produced outside the planet orbit by the perturbed density structure of the gas, even though the gas distribution extends closer to the planet, this process is often referred to as "dust filtration".

Furthermore, in some numerical simulations the formation of large gas streams has been observed linking the edge of the gas cavity with the central star. This provides an interesting solution that enables the feeding of the central star with a large amount of material despite the formation of a large dust and gas cavities.

We finally remark that all the mechanisms discussed in these sections do not necessarily exclude each other. The combined action of dynamical clearing with dust growth (Zhu et al. 2012b) and/or photoevaporation (Rosotti et al. 2013; Jennings et al. 2018) has been proposed to improve the clearing of the inner cavity from small dust grains, possibly solving the problem of dust filtration arising within the multiple planet system scenario and the inability of one single planet to open a cavity.

Below, we review the results regarding both (sub-)stellar and planetary companions as possible explanations of the observed cavities in transition discs.

Clearing by a sub-stellar companion

Bodies with relatively large satellite-to-star mass ratios ($q \gtrsim 0.04$ D'Orazio et al. 2016) are expected to clear cavities. Secondly, accretion streams of material linking the cavity with circum-individual discs provide an effective mechanism to feed the central star allowing us to explain the accretion related UV excess observed in these systems (Farris et al. 2014; Ragusa et al. 2016).

To further strengthen this hypothesis, (sub-)stellar companions have been detected in transition disc cavities. Namely, the system CoKu/Tau4, associated with a transition disc (Forrest et al. 2004; Nagel et al. 2010), has been found to host an almost equal mass binary with a binary separation $a \approx 8\,\mathrm{au}$ (Ireland & Kraus 2008).

HD142527 hosts an M-dwarf companion (Biller et al. 2012), orbiting the central star on an eccentric, inclined orbit (Lacour et al. 2016) with a semi-major axis $a \approx 30$–$50\,\mathrm{au}$ and a mass constrained by evolutionary models of 13–30 $\mathrm{M_J}$. In particular, this system have received a lot of attention by the scientific community for its peculiar features in all the observational bands discussed in the previous sections, including gas streams across the cavity visible in the HCO^+ line emission (Casassus et al. 2013, see the bottom panel of the central column in Fig. 8.2).

GGTau is a well known quadruple system composed by a close, almost equal mass binary (GGTau A). This system shows a clear circumbinary disc structure with an inner dust cavity in the IR (Beck et al. 2012; Yang et al. 2017).

The main criticism raised about the hypothesis concerning the clearing of the cavity (sub-)stellar companion comes from the fact that cavities in transition discs have been detected to be as large as $R_{\rm cav} \approx 100\,{\rm au}$ but the presence of such a companion have been excluded up to much smaller radii: this would in principle set a limit on the maximum semi-major axis of the companion, preventing the formation of such large cavities. However, it cannot be excluded that these massive companion might reside on inclined or/and eccentric orbits, so that their projected separation appears to be much closer to the central star than its real semi-major axis, as recently demonstrated in Price et al. (2018b) for the system HD142527.

Clearing by a planetary companion

Lower satellite-to-star mass ratios ($q \approx 10^{-3}$) have been found to open gaps rather than cavities, implying that one single planet is not expected to be able to carve a large cavity alone[4]. Scenarios where a system with multiple planets embedded in the disc carve the cavity have been explored (Dodson-Robinson & Salyk 2011; Zhu et al. 2011; Dong et al. 2015; Duffell & Dong 2015; Dong & Dawson 2016). However, one general result from these studies is that dust cavities produced by multiple systems of planets are generally shallower than those carved by single planets (Zhu et al. 2011; Duffell & Dong 2015), and therefore they are not able to reproduce the level of dust depletion in observations; moreover, the presence of multiple planets significantly reduces the accretion rate on to the central star with respect to the unperturbed state, in contrast with observations (Zhu et al. 2011). Finally, even considering the most optimistic occurrence rates of giant planets forming at separations between $3-30\,{\rm au}$ from exoplanets data, there are not enough planets with the characteristics to open such gaps (Dong & Dawson 2016).

From the observational point of view, few transition discs have been found to host forming planets strengthening the dynamical clearing scenario. In particular,

1. Two planet candidates at separations of $\approx 10\,{\rm au}$ (inside the dust cavity) and $\approx 50\,{\rm au}$ (in the outer disc) have been observed in HD 100546 Quanz et al. 2013; Currie et al. 2015.

2. Two planets at separations $\approx 14\,{\rm au}$ and $\approx 18\,{\rm au}$ have been observed in LkCa15 (Kraus & Ireland 2012; Sallum et al. 2015).

3. Besides the presence of ring-like structures and of a cavity (Osorio et al. 2014; Fedele et al. 2017) in the continuum observations of HD 169142, a number of point-like sources have been detected at IR wavelengths (Biller et al. 2014; Ligi et al. 2018), suggesting that this object hosts a multiple-planet system.

4. A planet candidate with mass $\lesssim 5\,{\rm M_J}$ orbiting at $\approx 20\,{\rm au}$ the central star within a non-completely depleted cavity has been detected in MWC 758 (Reggiani et al. 2018).

[4]We note that recently Bae et al. (2018) showed numerically that, properly modeling the dust grain growth physical processes, also systems with one planet with $\sim 1\,{\rm M_J}$ might be characterized by cavities at radio wavelengths.

5. The system PDS 70 appears to host a massive companion of $\approx 17\,M_J$ orbiting at $\approx 20\,au$ (Keppler et al. 2018; Müller et al. 2018), well within the dust cavity.

It must be reported that some candidates have been questioned due to the strong data processing needed to obtain high contrast images necessary to claim the detection of a planet (e.g., the innermost candidate in HD 100546 (Rameau et al. 2017; Mendigutía et al. 2017; Follette et al. 2017), the same applies to one of the point-like sources in HD 169142 (Ligi et al. 2018)).

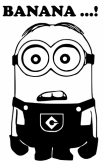
However, although the planet dynamical clearing mechanism appears to be very promising to model the formation of transition discs, some issues still remain unsolved. Dust filtration works well but has difficulties to explain the NIR deficit in moderately accreting transition discs (Zhu et al. 2012b). Indeed, while large dust grains are effectively trapped beyond the planet orbit, small dust grains are expected to cross the planet orbit following the dynamics of the gas. This replenishes the cavity with small grains, whose depletion is expected to provide a NIR deficit in the SED, being in fact not consistent with the observations. Furthermore, circumplanetary discs surrounding large planets are expected to be detectable with long integration times in ALMA band 7 or band 9 (Szulágyi et al. 2018; Zhu et al. 2018), but they still remain elusive.

Chapter

9

On the origin of horseshoes in transition discs

Based on the paper by ***Enrico Ragusa****, Giovanni Dipierro, Giuseppe Lodato, Guillaume Laibe & Daniel J. Price (2017), "On the origin of horseshoes in transitional discs", Monthly Notices of the Royal Astronomical Society, 464, 1449*



"Space bananas?!"

Minions, Despicable horseshoes

9.1 Non-axisymmetric structures in transition discs

Recent spectacular observations of dust and gas in nearby protoplanetary discs have revealed substructures in the form of spirals, gaps, cavities and ring-like features (ALMA Partnership et al. 2015; Andrews et al. 2016; see recent review by Casassus 2016). Whether, and how, such structures are created is critical to understanding the planet formation process.

One of the most spectacular first results with the Atacama Large Millimetre/Submillimetre Array (ALMA) was the observation of a non-axisymmetric 'horseshoe' in the dust continuum emission in Oph IRS 48 (van der Marel et al. 2013), with subsequent observations revealing asymmetric structures in several other transition discs (van der Marel et al. 2016a, see Fig. 8.2). Such features are

most commonly interpreted (including by van der Marel et al. 2013) as vortices, for example arising from the Rossby Wave Instability (RWI) at the edge of the gap formed by a young planet (Lovelace et al. 1999; Lyra et al. 2009; Lyra & Lin 2013; Zhu & Stone 2014; Hammer et al. 2017, 2018) or as a result of internal dynamical processes associated with the presence of a weak magnetic field (Ruge et al. 2016) and deadzones (Regály et al. 2012, 2017). RWI arises in sufficiently inviscid discs, with equivalent α parameters (Shakura & Sunyaev 1973) of the order of $\alpha \lesssim 10^{-4}$. Such a low viscosity allows the vortex to survive for thousands and up to 10^4 orbits (de Val-Borro et al. 2007; Ataiee et al. 2013; Zhu & Stone 2014; Fu et al. 2014a). Vortices can effectively trap dust particles, leading to a more azimuthally and radially concentrated dust density distribution of larger grains at the center of the vortex[1] (Barge & Sommeria 1995; Birnstiel et al. 2013; Lyra & Lin 2013; Zhu et al. 2014; Ruge et al. 2016). However, the combined effect of the dust settling and trapping inside the vortex produces an enhanced dust-to-gas mass ratio in the vortex, leading to an increase of the dust back-reaction. This produces an alteration of the coherent vorticity pattern and destroys the vortex (Johansen et al. 2004; Fu et al. 2014c).

Lopsided discs have been identified in high-resolution observations at (sub-)mm wavelengths. While in the case of IRS 48 (van der Marel et al. 2013) millimetre grains appear to be more concentrated in the horseshoe region compared to smaller sizes, in other cases (SR 21 and HD135344B; Pinilla et al. 2015), dust trapping is not observed, further challenging the vortex scenario. HD142527 shows a large horseshoe in mm continuum emission but whether or not dust trapping occurs is more controversial (Perez et al. 2015; Muto et al. 2015; Casassus et al. 2015b).

Here, we investigate an alternative explanation for the development of non-axisymmetric gas and dust structures based on studies of discs around black hole binaries (Shi et al. 2012; Farris et al. 2014; D'Orazio et al. 2016; Ragusa et al. 2016; Miranda et al. 2017). These showed that, for mass ratios $q \gtrsim 0.04$, the wide cavity around the primary object carved by the companion becomes eccentric and develops a strong overdensity at the cavity edge, orbiting at the local Keplerian frequency. This arises naturally even in relatively viscous discs in the presence of a sufficiently massive companion. In the protostellar case, Ataiee et al. (2013) showed with 2D hydrodynamic simulations that, for lower mass ratios ($\sim 10^{-3}$), the asymmetries at the cavity edge are weaker than in the vortex scenario, resulting in ring-like rather than horseshoe morphologies. It is therefore timely to explore the case with higher mass ratio to determine whether the horseshoe-like density features revealed by ALMA observations might be explained by the presence of a massive companion inside the cavity. We explore this hypothesis using global, 3D smoothed particle hydrodynamics (SPH) simulations of gas and dust evolution in a circumbinary disc, where the binary consists of a young star and either a massive planet or low mass stellar companion. We demonstrate the formation of crescent-like structures with emissivity contrast up to ~ 10, sufficient to explain many of the 'dust horseshoes' and other non-axisymmetric features observed in

[1] It can be shown that the velocity field in a vortex is sustained by a pressure configuration with a maximum when the vortex is anticyclonic and by a minimum when the vortex is cyclonic, this leads to particle trapping according to Eq. (2.58) (Youdin 2010).

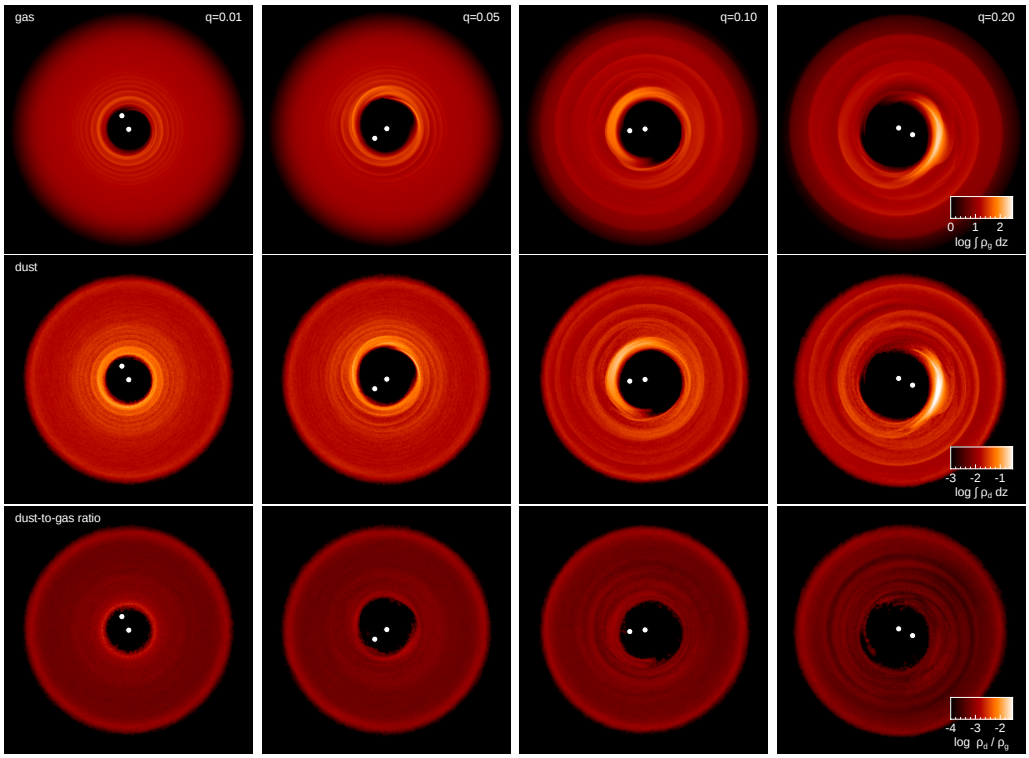


Figure 9.1: Gas (top row) and dust (middle row) surface density in units of g/cm^2 in logarithmic scale after 140 binary orbits for four different binary mass ratios; $q = \{0.01, 0.05, 0.1, 0.2\}$ (left to right, respectively). High mass ratio binaries drive the formation of a large eccentric cavity leading to non-axisymmetric overdensities in both gas and dust ($q \gtrsim 0.05$; right columns). Low binary mass ratios, by contrast, produce more axisymmetric overdensities around a smaller central cavity ($q \lesssim 0.05$; left columns). The bottom row shows the column averaged dust-to-gas ratio in logarithmic scale for the different mass ratios. Note that, for the millimetre size particles we simulate, no dust trapping occurs in the overdense region. Simulated observations of these calculations are shown in Figure 9.2.

transition discs.

The idea of transitional discs as circumbinary discs has been recently explored by Ruíz-Rodríguez et al. (2016). They found that the spectral energy distributions of $\sim 40\%$ of transitional discs in their sample can be explained as being produced by the flux emission of discs orbiting around binary systems.

9.2 Methods

We perform a set of 3D gas and dust numerical simulations of a circumbinary disc surrounding a binary object (to be interpreted both as star-star or star-planet system), using the PHANTOM SPH code (Lodato & Price 2010; Price & Federrath 2010; Price 2012). The binary is represented by two sink particles (e.g. Bate et al. 1995; Nixon et al. 2013) that exert the gravitational force on each other and on the gas particles. The sink particles are free to move under the backreaction force

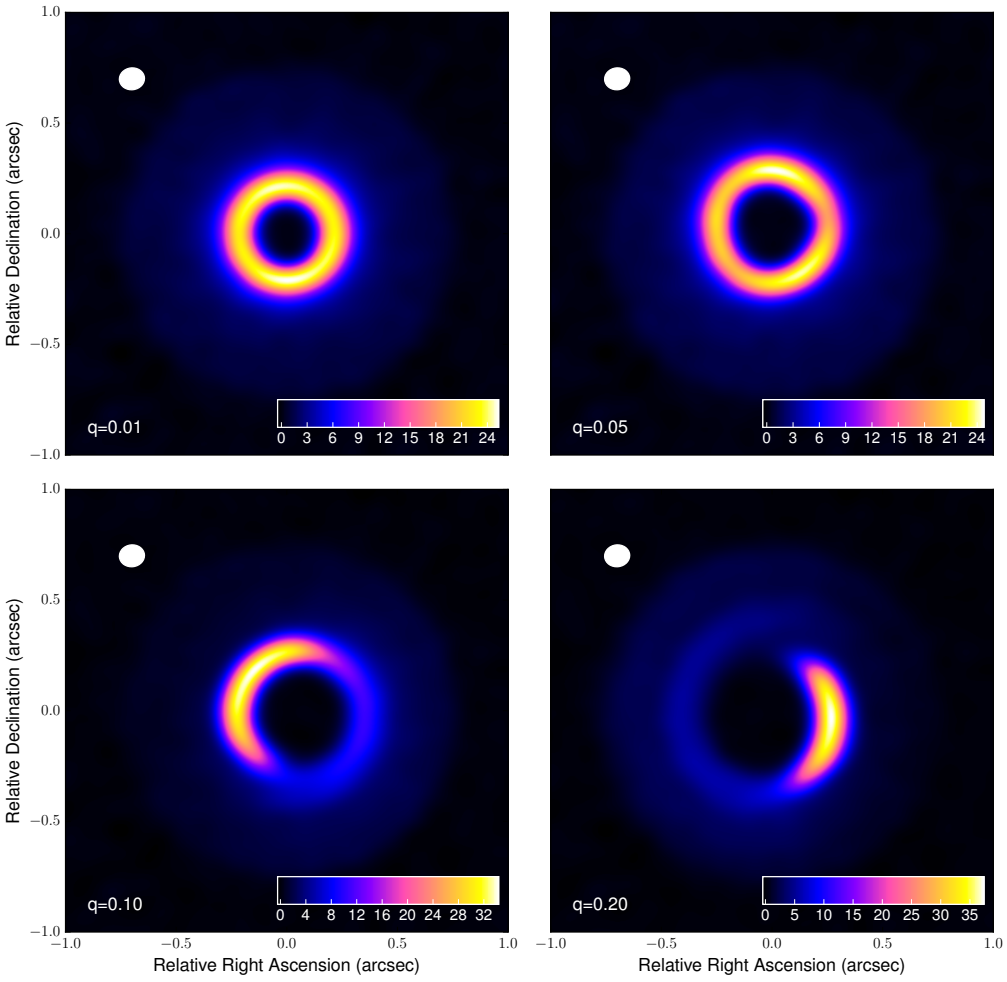


Figure 9.2: Comparison of ALMA simulated observations at 345 GHz of disc models with a mass ratio $q = 0.01$ (upper left), $q = 0.05$ (upper right), $q = 0.1$ (bottom left) and $q = 0.2$ (bottom right). Intensities are in mJy beam^{-1}. The white colour in the filled ellipse in the upper left corner indicates the size of the half-power contour of the synthesized beam: 0.12×0.1 arcsec ($\sim 16 \times 13$ au at 130 pc.).

of the gas on them, allowing the overall conservation of the binary-disc angular momentum, migration and eccentricity evolution of the binary. The sink particles are also allowed to accrete gas particles when they cross the sink radius and satisfy several dynamical conditions assuring that they are not able to escape from the gravitational field (Bate et al. 1995).

We model the gas-dust interaction using the one fluid model SPH formulation developed by Laibe & Price (2014a) and Price & Laibe (2015), assuming small grains (i.e. a Stokes number $\mathrm{St} \ll 1$) such that the terminal velocity approximation is valid. We set the dust grain size to be $s = 1\,\mathrm{mm}$ and vary the mass ratio of the binary. Pressure is computed using a locally isothermal equation of state assuming a radial power-law temperature profile. We do not consider the disc self-gravity.

We exploit the SPH artificial viscosity to model the physical processes responsible for the angular momentum transfer throughout the disc. With reference to the notation used in Lodato & Price (2010), we set the artificial viscosity parameter $\alpha^{\mathrm{AV}} = 0.1$ that corresponds, given our initial conditions, to a Shakura & Sunyaev (1973) α_{SS} parameter ranging between $0.01 \lesssim \alpha_{\mathrm{SS}} \lesssim 0.04$ across the disc. To prevent particle interpenetration we set the parameter $\beta = 2$ as prescribed in Price (2012).

SPH artificial viscosity provides also a natural way to reproduce turbulent diffusion of the gas (Arena & Gonzalez 2013) which is transmitted to the dust by the drag.

Each simulation is evolved for 140 binary orbits, corresponding to a physical time of $\sim 5600\,\mathrm{yr}$, which is long enough to allow the dust to settle from its initial displacement and to reach quasi-stationarity in the disc shape.

9.2.1 Initial conditions

Our initial setup consists of a sink particle binary surrounded by a disc of 2×10^6 SPH particles. The binary has a total mass $M_{\mathrm{tot}} = 2.2\,M_{\odot}$ (note, however, that the dynamics are only sensitive to the mass ratio, not the absolute mass), a binary separation $a = 15\,\mathrm{au}$ and an orbital eccentricity $e = 0$. We performed a set of four simulations varying the mass ratio $q = \{0.01; 0.05; 0.10; 0.20\}$.

The disc extends between an inner radius $R_{\mathrm{in}} = 18\,\mathrm{au}$ and an outer radius $R_{\mathrm{out}} = 100\,\mathrm{au}$, centred on the centre of mass of the binary. The surface density distribution is $\Sigma = \Sigma_0 R^{-p}$, where R is the radial coordinate in the disc, $p = 0.5$ and Σ_0 determines the total disc mass, $M_{\mathrm{disc}} = (1+\epsilon)M_{\mathrm{g,disc}}$, where $\epsilon = 10^{-3}$ is the millimetre dust-to-gas ratio (corresponding to a total dust-to-gas ratio of 0.01, for our assumed grain size distribution, see Section 9.2.2), and $M_{\mathrm{g,disc}} = 0.05\,M_{\odot}$ is the gas disc mass. Particles are distributed vertically according to a Gaussian distribution with thickness $H = c_{\mathrm{s}}/\Omega_{\mathrm{k}}$, where c_{s} is the gas sound speed and $\Omega_{\mathrm{k}} = \sqrt{GM_{\mathrm{tot}}/R^3}$. We assume that $H/R \propto R^{0.25}$ and that $H/R = 0.05$ at $R = 18$ au. The dust-to-gas ratio is $\epsilon = 10^{-3}$ throughout the entire disc, implying that the dust has initially the same vertical structure as the gas. After a few orbits of the secondary, the dust has settled from its initial displacement forming a layer with thickness $H_d = H\sqrt{\alpha_{\mathrm{SS}}/\mathrm{St}} \sim 0.7\,H$, consistently with the Dubrulle et al. (1995) model.

The velocity of each particle follows a Keplerian profile centered on the binary

centre of mass, with orbital velocities corrected to take account of the radial pressure gradient.

The average vertical resolution of this setup can be expressed as $\langle h/H \rangle \sim 0.2$, where h is the SPH smoothing length. Since, for our parameter choice, $H_{\rm d} \sim H$, the disc remains vertically well resolved both in the gas and in the dust.

9.2.2 Simulated ALMA observations

We performed mock ALMA observations of our models using the RADMC-3D Monte Carlo radiative transfer code (Dullemond 2012) together with the Common Astronomy Software Application (CASA) ALMA simulator (version 4.5.3), focusing on ALMA band 7 (continuum emission at 345 GHz). The source of radiation is assumed to be the central star, located at the centre of the coordinate system, with $M_\star = 2M_\odot$, $T_{\rm eff} = 5500$ K and $R_\star = 2R_\odot$. Dust opacities were produced using the routine[2] developed by Woitke et al. (2016) adopting the dust model from Min et al. (2016). Since ALMA band 7 images essentially trace millimetre particles with a maximum size of $\sim 3\lambda$ (Draine 2006), we assumed a dust population with a power-law grain size distribution given by $n(s) \propto s^{-m}$ between $s_{\rm min} = 0.1$ mm to $s_{\rm max} = 3$ mm, with $m = 3.5$. Starting from the 3D density distribution of millimetre grains of our model, we computed the spatial densities of grains in this size range by scaling the dust mass for each grain sizes according to the assumed size distribution, with a total dust mass in the size range $[0.1\,\mu{\rm m}, 10\,{\rm cm}]$ equal to 0.01 of the gas mass.

We computed full-resolution images using 10^8 photon packages. These images were then used as input sky models to simulate realistic ALMA observations taking into account the thermal noise from the receivers and the atmosphere and assuming a perfect calibration of the visibility measurements. We assumed that all the sources were located in Ophiuchus star-forming region ($d \sim$130 pc), observed with a transit duration of 3 minutes. We assumed Cycle 2 ALMA capabilities adopting an antenna configuration that provides a beam of 0.12 × 0.1 arcsec ($\sim 16 \times 13$ au).

9.3 Results

Fig. 9.1 shows the surface density after 140 binary orbits in the gas (top panels) and dust (middle panels) for four different disc models with increasing binary mass ratio ($q = 0.01, 0.05, 0.1$ and 0.2; left to right, respectively). As expected, the cavity size increases with the mass ratio (e.g. Artymowicz & Lubow 1994). The orbital eccentricity of the gas at the cavity edge also increases with q, reaching $e = 0.1$ for our highest mass ratio case ($q = 0.2$; last right column).

For lower mass ratios ($q = 0.01$ and $q = 0.05$; left two columns), the dust and gas density distribution are more axisymmetric, showing a ring-like overdensity at the cavity edge. After ≈ 100 orbits from the beginning of the simulation, an asymmetric crescent-shaped overdensity develops at the cavity edge for $q \gtrsim 0.05$

[2] `https://dianaproject.wp.st-andrews.ac.uk/data-results-downloads/fortran-package`

within a few binary orbital periods, with surface densities up to a factor ~ 10 denser than the surrounding gas, consistent with previous numerical simulations in the context of black hole binaries (Shi et al. 2012; Farris et al. 2014; Ragusa et al. 2016; Miranda et al. 2017). The overdensity is a Lagrangian feature that rotates with the local orbital frequency.

For fixed mass ratio, the level of contrast in the surface density across the crescent-shaped region is similar in both in the gas and in the dust. This is due to the fact that the high gas density in the lump produces a strong aerodynamical coupling between the gas and the dust in the disc. Interestingly, the sharpness of the region increases with increasing mass ratio.

Fig. 9.2 shows mock ALMA images of our disc models at band 7 for the four different mass ratios. The simulated ALMA images reflect the density structures observed in Fig. 9.1. In particular, a crescent or 'dust horseshoe' is evident for $q > 0.05$, with the contrast increasing with increasing mass ratio: for $q = 0.1$ the typical contrast is ≈ 5, while for $q = 0.2$ we obtain a contrast ≈ 7. For $q = 0.05$, the ALMA image shows a double-lobed feature with a low contrast ~ 1.5, similar to those observed in SR21 or DoAr 44 (van der Marel et al. 2016a). For $q = 0.01$ a ring-like structure can be observed, as observed e.g. in Sz 91 (Canovas et al. 2016).

The right panel of Fig. 9.3 shows a snapshot of the vorticity $\boldsymbol{\omega} = \nabla \times \mathbf{v}$, scaled to the Keplerian value $\boldsymbol{\omega}_{\rm K} = \nabla \times \mathbf{v}_{\rm K}$, where $v_{\rm K}$ is the Keplerian velocity field. The flow is close to Keplerian in the outer regions of the disc, while in the overdense region the value of the vorticity is $0 \lesssim \boldsymbol{\omega}/\boldsymbol{\omega}_{\rm K} < 1$. The extended region outside the overdense crescent where $\boldsymbol{\omega} > \boldsymbol{\omega}_{\rm K}$ is due to the steeper than Keplerian gradient of the azimuthal velocity. Vortices induced by the Rossby wave instability typically result in much higher vorticities, with anti-cyclonic vortices reaching $|\boldsymbol{\omega}/\boldsymbol{\omega}_{\rm K}| \sim 2$ (Owen & Kollmeier 2017).

9.4 Discussion

The idea that large scale asymmetries might be due to a planetary companion was explored by Ataiee et al. (2013), who concluded that planetary mass objects only produce ring-like features in the disc, in contrast to the observed horseshoe. However, we have shown the dynamics induced in the disc by low and high mass companions is markedly different. It is known that low-mass companions, with $q \sim 10^{-3}$ can produce eccentric cavities, that precess slowly around the star-planet system (Papaloizou et al. 2001; Kley & Dirksen 2006; Thun et al. 2017; Ragusa et al. 2018). In contrast, more massive companions, with $q \gtrsim 0.2$–0.3 (Shi et al. 2012; D'Orazio et al. 2016; Haiman, private communication) produce strong non-axisymmetric lumps that orbit at the local Keplerian frequency. We have explored the latter case in this paper.

We observe the formation of non-axisymmetric structures starting from mass ratios $q \gtrsim 0.05$. For sufficiently massive companions (binary mass ratio $q = 0.2$), we obtain an azimuthal contrast of the order of ~ 10 in the mm-wavelength map (Fig. 9.2), where the contrast is an increasing function of the binary mass ratio.

The mechanism causing the formation of the gas overdensity is thought to be

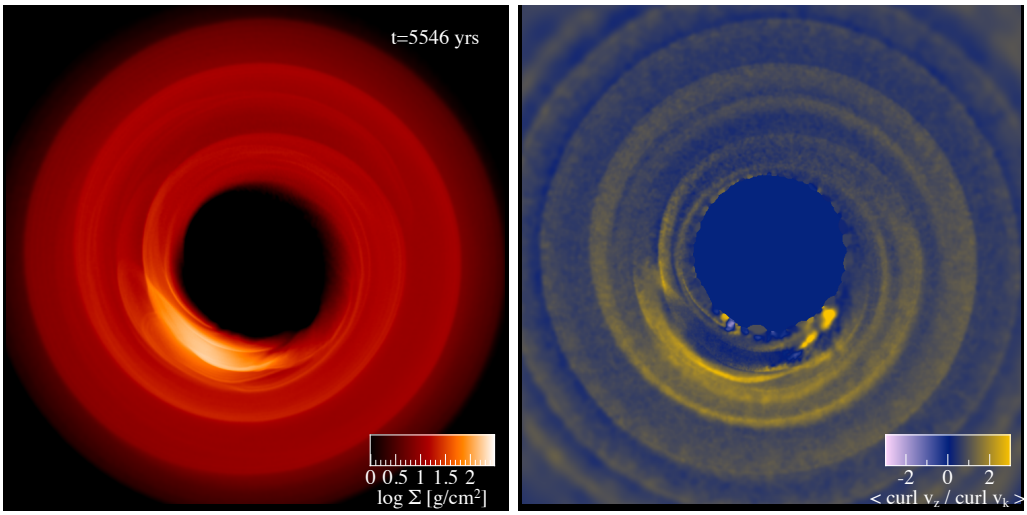


Figure 9.3: Map of the vorticity $\boldsymbol{\omega} = \nabla \times \mathbf{v}$ (right panel) in the region of the overdensity, scaled to the Keplerian value $\boldsymbol{\omega}_{\mathrm{K}} = \nabla \times \mathbf{v}_{\mathrm{K}}$, for $q = 0.2$. The left panel shows the gas density structure. There is no evidence of vortices associated with the overdense region (cf. Owen & Kollmeier 2017).

related to shocks in the gas at the cavity edge, arising from the intersection of gas flows within the cavity (Shi et al. 2012). D'Orazio et al. (2016) suggested that the formation of strong tidal streams has to be associated to the transition of the disc structure from an annular gap to a cavity, which take place for binary mass ratios $q \gtrsim 0.04$, due to lack of stable orbits around the Lagrange points L4 and L5. However, they also find that strong lopsidedness forms for slightly larger mass ratios $q = 0.2$–0.3 (Haiman, private communication). This appears to be in general agreement with our findings. However, we caution that further studies are required to investigate the evolution and survival of these features at longer timescales, and to determine whether the critical value of q depends on other parameters of the system, such as H/R and α. We also note that our sink prescription does not allow the formation of circum-individual discs that may or may not have an impact on the circum-binary disc structure, possibly causing the artificial accretion of some material in the tidal streams.

Due to the temperature change in shocking regions, it might be expected that the chemistry would be affected by shocks. Processes such as desorption of various chemical species from the surface of disc dust grains and gas-phase chemical reactions due to shocks occurring in the cavity wall, produce clear chemical signatures of the disc dynamics (see e.g. Ilee et al. 2011 in the case of shocks induced by gravitational instabilities) which may be detected using ALMA. Additionally, shocks might induce the emission of forbidden lines, the detection of which would confirm eccentric cavities as the origin of these structures.

As previously mentioned, horseshoes in transition discs are often assumed to be due to a vortex induced by a low-mass companion in the cavity. In this case, a variation of the azimuthal extent of the horseshoe at different wavelengths is expected. Indeed, models of dust trapping by a vortex predict that larger grains

would be more azimuthally concentrated in the centre of the vortex. However, in some cases (SR21 and HD135344B Pinilla et al. 2015) smaller grains appear to be more trapped than larger grains. Since we compute the dynamics of a single species of dust, we cannot predict if our model would reproduce this scenario.

9.4.1 Comparison with observed systems

Non-axisymmetric features have been observed in a handful of transitional discs (see Table 9.1). We report here below a brief summary about the cavity features and the current evidence for the presence of massive companions in these systems.

Regarding the upper-limits on the mass of putative companions, the most accurate results have been obtained applying the aperture masking interferometric observations and speckle imaging in the near-IR waveband. It is worth noticing that detecting planets through imaging is challenging due to the proximity of planets to the central star and their low contrast ratio in emission compared to the brightness of their host star. Additionally, massive companions might have eccentric or misaligned orbits with respect to the disc. This implies that the size of the cavity they are able to carve can be much larger than ~ 2 times the separation at which they are located in the imaging due to projection effects or orbital phase (a planet might not be resolved at the pericentre of its orbit, while resolved and thus detectable at the apocentre).

9.4.1.1 HD135344B

Recent observations of line and continuum emission from HD135344B evidenced the presence of a cavity both in the gas (~ 30 au) and in the dust (~ 40 au) (van der Marel et al. 2016a). The continuum emission shows also a well defined crescent shaped overdense feature at the cavity edge with a mild contrast (van der Marel et al. 2016a). A spiral structure has also been detected in the near-IR scattered light, constituting a strong indication of the presence of a massive companion (Garufi et al. 2013). Using the "locally optimized combination of images" (LOCI) technique, in order to be able to possibly resolve and locate the exact position of the companion, Vicente et al. (2011) put an upper-limit of $M_c \sim 230 M_\mathrm{J}$ at separations $a \lesssim 14$ au and $M_c \sim 85\, M_\mathrm{J}$ at $a \lesssim 37$ au.

Given the central star estimated mass $M_\star \sim 1.7 M_\odot$, the upper-limits on the secondary mass imply mass ratios $q \lesssim 0.05$ at $a \lesssim 37$ au and up to $q \sim 0.13$ for separations $a \lesssim 14$ au. This is consistent with our models, since the crescent-like feature with contrast $\lesssim 10$ observed in HD135344B (van der Marel et al. 2016c) is similar to what we obtain for our $q = 0.1$ case (bottom left panel of Fig. 9.2).

9.4.1.2 SR 21

The continuum emission from this system shows different asymmetric features at different wavelengths: a crescent shaped overdense feature at 690 GHz (Pérez et al. 2014), and a double-lobed structure at 345 GHz (van der Marel et al. 2016a). In both cases the contrast is mild ($\lesssim 10$). Modeling the dust emission van der Marel et al. (2016a) inferred a cavity edge in the dust at ~ 25 au, while the gas cavity appears to be much smaller (~ 7 au, Pontoppidan et al. 2008).

The presence of a warm companion surrounded by a cloud of accreting gas in this system was invoked by Eisner et al. (2009) to explain an excess in the near-IR and mid-IR SEDs, which could be explained by an additional warm ($\sim 700\,\mathrm{K}$) black body emission from an extended region of $40\,R_\odot$. The total luminosity produced by the companion in this framework appears to be consistent with a T-Tauri star with mass $M_c \sim 0.2\,M_\odot$ enveloped in a gaseous cloud (Follette et al. 2013). Using the angular differential imaging (ADI) technique, Follette et al. (2013) were able to rule out the presence of a secondary stellar object for separations $a \gtrsim 18$ au; based on the contrast sensitivity achieved by Follette et al. (2013), Wright et al. (2015) constrained the upper-limit on the companion mass to $\sim 40 - 60\,M_\mathrm{J}$ at separations $a \gtrsim 18$ au, implying that such a stellar source needs to be located at separations $a \lesssim 18$ au. This implies possible secondary-to-primary mass ratios of $q \sim 0.1$ for $a \lesssim 18$ au and $q \sim 0.03$ for $a \gtrsim 18$ au. The double-lobed structures detected in the continuum emission in SR 21 (e.g. Pinilla et al. 2015) at 345 GHz are consistent with our models with $q \sim 0.05$, which is in agreement with the detection limits reported in literature.

9.4.1.3 IRS 48

With an azimuthal contrast of $\gtrsim 130$ of the peak emission compared to the background disc, IRS 48 represents the source with the strongest crescent-shaped dust structure in our sample. This dramatic azimuthal range at the cavity wall observed in IRS 48 is best described by the segregation of millimetre grain sizes induced by an azimuthal bump in pressure (van der Marel et al. 2013). The dust cavity was found to be extended 60 au from the central star (Bruderer et al. 2014b). The continuum asymmetry has been modeled as a major dust trap, triggered by the presence of a substellar companion with a mass of $9M_\mathrm{J}$ (Zhu & Stone 2014). This would seem very unlikely to correspond to an equally sharp gas distribution. Alternatively, Wright et al. (2015), based on the detection limit reported in Ratzka et al. (2005), rule out a potential companion with a mass $\gtrsim 100\,M_\mathrm{J}$, which correspond to a mass ratio of $q \sim 0.05$ at a radial separation of 19 au. As expected, this result does not appear consistent with the mechanism presented in this paper.

9.4.1.4 DoAr 44

DoAr 44 (also known as ROX 44 and Haro 1-16) is the source with the mildest contrast in (sub-) mm continuum emission in our sample (van der Marel et al. 2016a). This source has been classified as a pre-transition disc with a dust cavity between 2 and 32 au (Espaillat et al. 2010; van der Marel et al. 2016a). Based on the companion detection limits reported in Ratzka et al. (2005), Wright et al. (2015) derived that a potential companion should have a mass $\lesssim 80\,M_\mathrm{J}$ at a separation $\lesssim 12.5$ au which correspond to a mass ratio of $\lesssim 0.06$, adopting the star properties reported in Espaillat et al. (2010). The values of the mass ratio inferred from observations appear to be consistent with our scenario: the double-lobed structures observed in DoAr 44 can be explained for $q \sim 0.05$ according to our model.

9.4.1.5 HD142527

HD142527 harbours a disc with a wide dust cavity extending from 10 au to 120 au. At the cavity inner wall, the (sub-) millimetre dust continuum emission show a bright horseshoe with contrast ∼30. The existence of a massive close companion with $q \approx 0.16$ has been established for HD142527 using the sparse masking aperture technique (Biller et al. 2012; Lacour et al. 2016). This is particularly interesting since the contrast in HD142527 of ~ 30 is within a factor of 3 of the contrast we find in our highest mass ratio. However, this companion is also inclined by $\sim 70°$ with respect to the disc (Lacour et al. 2016) and this case might be further complicated by a strong warp (Casassus et al. 2015a). Intriguingly, Casassus et al. (2015b) comment that "*the large sub-mm crescent [in HD142527] mostly reflects the gas background, with relatively inefficient trapping, so that the observed contrast ratio of* $\sim$ *30 is accounted for with a contrast of 20 in the gas*", consistent with our model. Seemingly this conflicts with Muto et al. (2015) who found variations of $\sim$ 10–30 in the dust-to-gas ratio. This difference may be explained by uncertainties in grain surface chemistry, in particular whether or not a fraction of CO is depleted on dust grains (Casassus 2016). Recently Price et al. (2018b) showed that all the main observational features of HD142527 can be explained by the interaction of the disc with the inclined sub-stellar companion hosted in this system.

9.4.1.6 Lk Hα 330

Lk Hα 330 is characterised by a millimetre dust cavity with a size of about 40 au and an azimuthal intensity variation of a factor of two. Recent observations performed by Willson et al. (2016), using the sparse aperture masking technique in the K' near infrared band, on Lk Hα 330 revealed the presence of a possible massive companion characterized by a value of $M_{\rm c}\dot{M}_{\rm c} \sim 10^{-3}\,{\rm M_J^2 yr^{-1}}$ orbiting at a separation $a \sim 37\,{\rm au}$ from the central star. Assuming an accretion rate on the secondary object of $\dot{M}_{\rm c} \lesssim 10^{-8}\,M_\odot {\rm yr}^{-1}$ implies an mass $M_{\rm c} \gtrsim 100\,{\rm M_J}$ and a mass ratio $q \gtrsim 0.05$. It should be noted that previous works (Brown et al. 2009; Andrews et al. 2011; Isella et al. 2013) had reported private communications that, based on near-IR observations, ruled out the presence of secondary objects with masses $M_c \gtrsim 50\,{\rm M_J}$ ($q \gtrsim 0.025$) at separations $M_c \gtrsim 10\,{\rm au}$, indicating how elusive these objects might be. In any event, these estimates for the companion mass appear to be consistent with the low sharpness of the crescent-shaped structure predicted by our model (see the upper panels of Fig. 9.2).

9.5 Conclusions

We performed 3D SPH gas and dust simulations of circumbinary discs surrounding a protostar and either a low mass stellar companion or massive protoplanet. We showed that the companion carves a wide, eccentric cavity in the disc, resulting in a non-axisymmetric gas overdensity at the cavity edge. For sufficiently large binary mass ratios this feature appears as a 'horseshoe' in millimetre wavelength dust continuum images, as observed in several transition discs.

Name	Contrast	Dust trapping	Companion	Consistency
HD135344B	$\lesssim 10$	No	Strong indication	Yes
SR 21	$\lesssim 10$	No	Indication	Yes
DoAr 44	$\lesssim 10$	?	?	Yes
IRS 48	$\gtrsim 100$	Yes	?	No
HD142527	~ 30	cm grains?	Yes	Yes
Lk Hα 330	$\lesssim 10$	?	Indication	Yes

Table 9.1: Summary of transition discs displaying horseshoe or other non-axisymmetric features. For each source, we indicate the observed contrast in mm images, whether there is evidence for dust trapping in the crescent, and whether the system is known to host a massive companion. The last column indicates whether the observed structures are consistent with our model, given the upper-limits on the companion mass as reported in the literature.

Our model makes testable predictions that can be used to observationally distinguish the eccentric cavity model from the more commonly assumed 'gap edge vortex' model. We identify the following main features differentiating the two processes, which can be used as the basis for observational tests of our hypothesis:

1. Dust and gas kinematics. In our model, the fluid velocity is close to Keplerian, and does not show the large vorticity expected in the vortex model (see Fig. 9.3).

2. Our mechanism applies both in high and in low viscosity discs, while vortices only arise for $\alpha \lesssim 10^{-4}$.

3. With our parameter choice, dust horseshoes arise from eccentric cavities only for relatively large mass ratios $q \gtrsim 0.05$, while in principle vortices can arise for lower mass planets. Clearly, stablishing whether a relatively massive companion is present within the cavity is a key observational test of our model.

4. The structures described in this paper only occur at the edge of the central cavity, while vortices can occur in principle at any location within the disc. Observing a non-axisymmetric feature at a location different from the edge of the cavity excludes the mechanism we propose to be responsible for its formation.

5. Less massive companions should produce more axisymmetric structures, potentially explaining the 'dust rings' seen in (e.g.) Sz 91 and DoAr 44. We predict that a higher degree of non-axisymmetry around larger central cavities.

6. Our model does not require azimuthal dust trapping and the observed contrast largely reflects the gas density contrast.

In summary, cavities opened by massive companions are a promising mechanism for explaining rings, lopsided features and horseshoes seen in transition discs.

Future works should investigate the validity of the threshold mass ratio $q \gtrsim 0.05$ for different values of viscosity, disc thickness and the evolution of the structures at longer timescales.

Summary and conclusions

"...And everything under the sun is in tune
But the sun is eclipsed by the moon."

Pink Floyd, Eclipse

In this thesis we discussed the role of the disc-satellite mutual interaction in determining the dynamics and electromagnetic output of the system. The presence of a secondary body embedded in a gaseous accretion disc significantly alters the dynamics of the material with respect to the predictions of classic accretion theory.

Accounting for the presence of secondary bodies embedded in a gaseous environment is of paramount importance in a large variety of astrophysical systems (star + star, planet + star, black hole + black hole, planet + moons, e.g. Saturn and its rings and moons). Despite the differences among all these systems, their dynamics and, as a consequence, the evolution of their geometry/structure can be modelled under one unique framework: the tidal disc-satellite interaction.

With the ambitious aim to make the thesis as self-consistent as possible, in **Part I** we provided a broad introduction of the fundamental concepts required for the understanding of the topics presented throughout this manuscript.

Part II and **Part III** are then dedicated to three aspects of disc-satellite interaction that are still not completely understood, in particular: the suppression of accretion rate driven by the tidal effects of the satellite, the mutual long term evolution of satellite/disc eccentricity and the stability of the disc with respect to azimuthal density perturbations producing non-axisymmetric structures.

Each of these problems can be studied from a purely theoretical point of view, studying the fluid dynamics equations under the effects of gravitational forces exerted by point masses, and vice versa. As a consequence, the results found in this thesis can be applied to any astrophysical system satisfying the appropriate

assumptions. However, in all cases, our studies focus on specific classes of objects where we believe the physical processes under investigation are relevant. We summarize below our main results:

- **Part II**, Chapter 6 (Ragusa et al. 2016): We investigate whether the tidal effects exerted on the disc by the binary can reduce the accretion rate of material on to it. The dam effect produced by the companion depends on the balance between two competing mechanisms: viscous/pressure processes promote the inward motion of the material, while tidal effects exert an opposite torque on the disc reducing the accretion. In particular, the accretion rate on to the binary $\dot{M}_{\rm bin}$ is expected to decrease with the the disc thickness faster than what would be expected for the rate on to a single central object $\dot{M}_0$ ($\dot{M}_0 \propto (H/R)^2$), so that possibly $\dot{M}_{\rm bin}/\dot{M}_0 < 1$.

  Some numerical simulations showed that $\dot{M}_{\rm bin}$ is not altered with respect to $\dot{M}_0$ for a disc thickness $H/R \sim 0.1$ (D'Orazio et al. 2013; Farris et al. 2014; Miranda et al. 2017). However, since the intensity of the viscous processes depends on the disc thickness, we expect that the tidal forces overcome the viscous ones for smaller values of H/R, resulting in $\dot{M}_{\rm bin}/\dot{M}_0 < 1$.

  The reduction of the accretion rate due to a binary companion has extremely important implications in the context of black hole binary systems. In particular, it affects the detectability and the dynamical evolution of these systems, both before and after the merger of the two black holes.

  In order to address this issue, we performed a set of numerical simulations of equal mass binaries surrounded by accretion discs for different values of disc thickness. We recover the previous result in the literature $\dot{M}_{\rm bin}/\dot{M}_0 \approx 1$ for $H/R = 0.1$, but, as expected, we find a linear decrease of $\dot{M}_{\rm bin}/\dot{M}_0$ for smaller values of H/R. Our results suggest that extrapolating the accretion rate for a disc thickness $H/R \approx 10^{-3}$, which is expected in discs surrounding black holes, one would get a reduction of a factor ~ 100 of the accretion rate.

  We caution that in our simulations we do not inject material steadily at large radii, as D'Orazio et al. (2013), Farris et al. (2014) and Miranda et al. (2017) did, we use instead a finite disc. Which one of these two implementations of the mass supply is the most appropriate to describe the physical behaviour of real systems likely depends on whether these systems receive a constant inflow of material for the entire length of their life or they undergo multiple isolated accretion events.

  In conclusion, this work suggests that the reduction of the luminosity of the system, associated with the suppression of the accretion rate, might affect our ability to identify these systems using the electromagnetic channel. Furthermore, low accretion rates have been associated with poor spin-orbit alignment of the black holes at the time of the merger, possibly resulting in high recoil velocities of the merger remnant (see Sec. 5.4).

- **Part III**, Chapter 7 (Ragusa et al. 2018): The interaction between a satellite and the disc causes the mutual evolution of their orbital properties, impacting the structure of the whole system. The role of the tidal torque in

driving the growth of the eccentricity of both the satellite and the disc is well documented in the literature. In particular, the growth (or decrease) of the eccentricity appears to be due to the relative intensity of the angular momentum exchange between the disc and the satellite at resonant locations, which intrinsically depends on the local disc structure and depletion of the co-orbital region (Artymowicz et al. 1991). This translates to a minimum mass (but possibly also initial minimum satellite eccentricity, D'Angelo et al. 2006,Duffell & Chiang 2015) for which eccentricity growth occurs (Ogilvie & Lubow 2003; Goldreich & Sari 2003). When this happens, the disc has been observed to achieve an eccentricity profile that decreases with radius. The mass of the disc has also been shown to play a role in this process, enabling a faster growth of the eccentricity of the system when the disc is more massive (Papaloizou et al. 2001; Dunhill et al. 2013).

This mechanism has possibly a deep impact in determining the orbital properties of planets during the protoplanetary disc phase. Motivated by the observation of eccentric hot Jupiters orbiting very close to their host stars, we investigated the long term evolution of a 13 MJ satellite and two different values of disc mass (we refer to them as the light and the massive case), by means of performing two long timescale 2D hydrodynamical simulations. Both the planet and the disc were left free to evolve under the effects of their mutual interaction for $\approx 3 \times 10^5$ planet orbits (for reference, the longest simulation in the literature covered 2×10^4 orbits).

The evolution of planet and disc eccentricity shows a very peculiar periodic exchange of eccentricity between the disc and the planet, superimposed to a long term trend of linearly growing planet eccentricity for the case where the less massive disc was used, and, in contrast a linearly decreasing planet eccentricity for the more massive disc case. The disc eccentricity is instead linearly decreasing in both cases at a similar rate. Both the planet and the disc eccentricity show very similar precession patterns characterized by aligned or antialigned pericentres configurations, rapidly or slowly precessing at two discrete frequencies.

We provided a simplified toy model to explain the periodic oscillations. The results can be qualitatively reproduced and modeled by the presence of two eigen-modes for the evolution of planet and disc eccentricity. One mode is growing, the other one is decreasing. The relative intensity of the eigenmodes, that appears to be related to the ratio between the planet and disc angular momenta, determines the evolution of the system.

The results obtained within this part of the thesis highlight the importance of the long term evolution for these systems to determine the final structure after the disc dispersal. Furthermore, our results also suggest that the expectation that more massive discs are more prone to produce higher values of planetary eccentricity can be completely reversed at very long timescales. However, a deeper study further exploring the parameter space is required to confirm this result.

- **Part III**, Chapter 9 (Ragusa et al. 2017): The presence of a secondary

object embedded in the disc alters significantly its structure and, for sufficiently large mass ratios, it results in the formation of a cavity. Furthermore, the disc-satellite interaction appears to promote the development of a certain degree of axial asymmetry in the disc that results in the formation of overdensities at the cavity edge.

Recent (sub)mm images taken by ALMA reveal a large variety of azimuthal features. The formation of such structures has been broadly associated with the formation of vortices in the flow induced by a planet (Lyra & Lin 2013; Zhu & Stone 2014) or other possible mechanisms (Regály et al. 2012; Ruge et al. 2016) triggering the RWI (rapidly evolving structures, orbital motion of the feature), that requires very low levels of disc viscosity ($\alpha_{\rm SS} \sim 10^{-4}$). These types of structures have also been associated with the growth of the disc eccentricity (Ataiee et al. 2013; Ragusa et al. 2018) induced by a planet (slowly evolving eccentric structures, secular precession of the feature). However, a number of simulations with large mass ratios ($q \approx 1$) in the field of black hole binaries, showed the formation of similar features with relatively high viscosity ($\alpha_{\rm SS} \sim 0.1$, Shi et al. 2012; Farris et al. 2014; Ragusa et al. 2016; Miranda et al. 2017).

With the aim to investigate this issue, we performed a set of hydro gas+dust simulations. Consistent with the results obtained in the black holes simulations, our results show that for $q \geq 0.1$ the disc develops an overdense non-axisymmetric feature (constrast ratio, up to a factor 10), that rapidly moves with Keplerian frequency, where the vorticity vanishes. This suggests the existence of an alternative mechanism for the formation of these structures that require larger masses than those expected to trigger RWI, but that is not suppressed by higher levels of disc viscosity ($\alpha_{\rm SS} = 0.01$–0.04). We created radiative transfer synthetic images to ensure that the overdensities we observe in the hydro simulations translates to the surface brightness observed in the ALMA images. The mechanism feeding the overdensities appears to be a form of instability related to the formation of strong tidal streams and an eccentric cavity, but still some work needs to be done in order to properly understand its physical origin.

In conclusion, horseshoe structures in protoplanetary systems might be due to the presence of a relatively massive substellar companion in the brown dwarf regime rather than a planet. The constraints in the literature about the detectability of such a massive companion appear to allow this mechanism as a possible explanation for the observed structures in a number of systems. However, we are not able to reproduce contrast ratios of the asymmetries $\gtrsim 10$, as for example the case of IRS48 (van der Marel et al. 2013)

We want here to note the paramount importance for scientists in the research field of disc-satellite interaction to keep updated about the results from both the high energy side, involving black holes, and from the protoplanetary discs side. Indeed, we remark that two of the three projects presented throughout this thesis were originally inspired by the results and advancements obtained in completely

different fields: the accretion rate on to binary objects has been long supposed to depend on the disc thickness in the field of protoplanetary systems, but the impact of this dependence on the luminosity and dynamical evolution of black hole binaries had not been considered before (Ragusa et al. 2016); analogously, the formation of non-axisymmetric structures in simulations of discs surrounding black hole binaries has been widely reported, but they have not been associated with the horseshoe structures observed in the resolved images of protoplanetary discs provided after the advent of ALMA (Ragusa et al. 2017).

We also note here how the results presented in Ragusa et al. (2018), obtained for planetary systems, might be relevant for the dynamics of extreme mass ratio inspirals (EMRI, i.e. gravitational wave events involving black holes with extremely small mass ratios) possibly setting the initial conditions of the gravitational wave phase during their way toward the merger.

Finally we want to remark how important performing very long timescale simulations is, in order to fully capture the dynamics of the system: secular effects might be effective over thousands of orbits, and can possibly reverse the dynamical behaviour at short timescales as shown in Ragusa et al. (2018).

Future developments

Starting from the work I performed in this thesis, I identified four issues that I will address in future projects

1. A criterion for the formation of non-axisymmetric structures

As broadly discussed, numerical simulations performed both by the protoplanetary disc community (Ataiee et al. 2013; Lyra & Lin 2013; Ragusa et al. 2017) and by the black hole one (Shi et al. 2012; Ragusa et al. 2016; Farris et al. 2014) have revealed that non-axisymmetric structures are divided mainly in two broad categories: slowly evolving features (Ataiee et al. 2013; Ragusa et al. 2018), apparently due to the clustering of eccentric fluid trajectories (slowly precessing over a secular timescale), and rapidly evolving ones (wih the Keplerian frequency at the lump location of the overdensity). The latter come in two different flavours: one showing vortical motion (Ataiee et al. 2013; Lyra & Lin 2013) requiring low viscosities $\alpha \sim 10^{-4}$ and the setting in of the Rossby Wave Instability (RWI), and the other one showing no vortices (Ragusa et al. 2017; Farris et al. 2014).

Deepening our understanding of the formation mechanisms of non-axisymmetric structures is of great importance both for compact objects binaries, since it provides a modulation of the accretion rate and thus of the EM luminosity, and for planets, since they favour the aggregation of dust grains into larger objects.

A possible development in this direction could be to determine a criterion in terms of mass ratio, viscosity and disc thickness (performing 3D SPH numerical simulations and providing an analytical criterion) to describe the transition from the slowly evolving to the rapidly evolving features and also understand whether the vortical and non-vortical configuration are produced by two separate mechanism or are just two faces of the same coin. The starting point is to investigate a poorly understood instability that depends on the binary mass ratio (D'Orazio

et al. 2016), and see whether it can be associated with the formation of tidal streams shocking at the cavity edge.

2. Distinguishing vortices, slowly evolving and rapidly evolving features

In the numerical simulations performed in Ragusa et al. (2017), where we observed rapidly evolving features, we did not observe dust trapping for 1 mm dust grains, but we are able to achieve a contrast ratio in the surface density $\Sigma/\Sigma_0 \sim 10$. Slow eccentric features, being simple "traffic jams" in the fluid trajectories should not show dust trapping of large particles; they can reach surface contrast ratios $\Sigma/\Sigma_0 \sim 3$ (Ataiee et al. 2013). Regarding vortices, numerical simulations showed that they present dust trapping of large dust grains, and produce surface densities contrast ratios reaching $\Sigma/\Sigma_0 \sim 6$ (Ataiee et al. 2013).

Pure imaging alone is not able to discern the difference between these three mechanisms since the contrast ratio is similar, therefore it would be interesting to predict the kinematical appearance of these structures in protoplanetary systems (or at least to put constraints on the technical requirements we need to distinguish them). This can be done by post processing the output of numerical simulations (from both grid and particle methods) using a radiative transfer code, such as RADMC-3D (Dullemond 2012), in order to obtain the predictions about velocity maps (1st moment) of the CO vibrational lines to be compared with ALMA real observations.

3. Secular disc-satellite interaction for long timescales

In Ragusa et al. (2018) we studied the long term evolution of the eccentricity in a protoplanetary system in presence of a massive planet (planet-star mass ratio $q = 0.013$). We performed two very long timescale (3×10^5 orbits) numerical simulations for two different disc masses. Both simulations show periodic eccentricity oscillations superimposed on a growing/decreasing trend (see Fig. 7.6) and the precession of pericentre of both planet and disc, and a reversal of the mass dependence of the eccentricity growth rate at late times, i.e.: the less massive disc pumps the eccentricity at late times, while the massive one damps it, contrary to what previously observed at short timescales (Papaloizou et al. 2001; Dunhill et al. 2013).

The complete reversal of the initial growth trends of the eccentricity and its periodic oscillations offer a clear evidence of the importance of carrying out very long timescale numerical simulations in order to also allow the secular interaction to play a role in the evolution.

In this context it might be useful to explore in greater detail the secular binary disc interaction and its effects on the dynamics at very late times. This includes an extensive survey of the parameter space in order to simulate systems of interest both for the protoplanetary community ($q \ll 1$, $H/R \approx 0.1$, $\alpha_{\rm SS} = 0.001$) and for the BHB one ($q \sim 1$, $H/R \approx 10^{-3}$, $\alpha_{\rm SS} = 0.1$).

4. Accretion rate: modulation and reduction

In (Ragusa et al. 2016) we investigated the dependence of the accretion rate on

the disc thickness, which is expected to be very small for discs surrounding black holes ($H/R \sim 10^{-3}$). Numerically, we found a linear relationship of the type $\dot{M}_{\rm bin} \propto \xi(H/R)\dot{M}_0$, where $\dot{M}_{\rm bin}$ is the accretion rate on to the binary, while $\dot{M}_0$ is the corresponding rate for a single central object; $\xi(H/R)$ is a roughly linear function of H/R, thus implying a reduction of the accretion rate as the disc becomes thinner. However, we were not able to simulate discs thinner than $H/R = 0.02$ due the poor vertical resolution below this threshold. In this context, the immediate development is to provide an analytical model to describe the dependence of the accretion rate in binary systems on the disc temperature: this would allow us to strengthen the numerical results with theoretical predictions, increasing the reliability of the extrapolation of the accretion rate at values of disc thickness that are prohibitively small to be investigated with numerical simulations.

Appendices

Appendix A

Disc thickness in discs surrounding supermassive black holes

A.1 The Shakura & Sunyaev solutions

In this section we aim to provide an analytical estimate of the disc thickness. With reference to the notation we adopted in Chap. 2, we provide here the closed set of equations determining the structure of a thin, isothermal disc (Frank et al. 2002); the further underlying assumptions are that the disc is optically thick and in steady state

$$\begin{cases} 1. \quad \rho = \dfrac{\Sigma}{H}, \\ 2. \quad H = c_{\rm s}\sqrt{\dfrac{R^3}{GM}}, \\ 3. \quad c_{\rm s}^2 = \dfrac{\partial P}{\partial \rho}, \\ 4. \quad p = \dfrac{k_b T_{\rm c}\rho}{\mu m_{\rm p}} + \dfrac{4\sigma_{\rm SB}}{3c}T_{\rm c}^4, \\ 5. \quad \dfrac{4\sigma_{\rm SB}}{3\tau}T_{\rm c}^4 = \dfrac{3}{8\pi}\dfrac{GM_p\dot{M}}{R^3}\left(1-\sqrt{\dfrac{R_{\rm in}}{R}}\right), \\ 6. \quad \tau(\Sigma,\rho,T_{\rm c}) = \kappa_R(\rho,T_{\rm c})\Sigma, \\ 7. \quad \nu\Sigma = \dfrac{\dot{M}}{3\pi}\left(1-\sqrt{\dfrac{R_{\rm in}}{R}}\right), \\ 8. \quad \nu = \nu(\rho,\Sigma,H,c_{\rm s},\ ...). \end{cases} \tag{A.1.1}$$

The second term on the r.h.s. of Eq. 4. is the radiation pressure, where $\sigma_{\rm SB} = 5.6704 \times 10^{-5}\mathrm{erg\,cm^{-2}\,s^{-1}\,K^{-4}}$ is the Stefan-Boltzmann constant, while c is the

speed of light. This set of equations consists of 8 equations in 8 unknowns ρ, Σ, H, c_s, p, T_c, τ and ν. The system is completely self-consistent, however, in order to close the equations, a prescription for the viscous parameter $\nu(c_\mathrm{s}, H)$ and the Rosseland mean opacity $\kappa_R(\rho, T_\mathrm{c})$ needs to be provided.

In their seminal paper, Shakura & Sunyaev (1973) found analytical solutions for the set of equations (A.1.1), prescribing $\nu(c_\mathrm{s}, H) = \alpha c_\mathrm{s} H$ (see Eq. 2.35), while Rosseland mean opacity is determined by to two competitive mechanism: free-free interaction ruled by the Kramers' law for $\kappa_\mathrm{R,ff}$:

$$\kappa_\mathrm{R,ff} = 6.4 \times 10^{-24} \left(\frac{\rho}{\mathrm{g\,cm^{-3}}}\right) \left(\frac{T_\mathrm{c}}{K}\right)^{-7/2} \mathrm{cm^2\,g^{-1}}; \tag{A.1.2}$$

the second is the Thomson scattering:

$$\kappa_\mathrm{R,Th} = \frac{\sigma_\mathrm{Th}}{\mu m_\mathrm{p}}. \tag{A.1.3}$$

μ is the average atomic number and $\sigma_\mathrm{Th} = 6.665 \times 10^{-25}\mathrm{cm}^2$ is the Thomson scattering cross section; the dominance of one process on the other depends on the temperature and on the density of the gas. The solutions will then depend on the free parameters α, $\dot{M}$, M and R.

Shakura & Sunyaev (1973) discuss the solutions in three distinct regions of the disc (a, b, c). Each region assumes a different combinations of pressure and opacity as follows:

a. It is the region closest to the central object. The radiation pressure dominates over the gas one and the mean opacity is mostly due to Thomson scattering of photons on free electrons.

b. Gas pressure dominates over the radiation one, but Thomson scattering still dominates the opacity, Eq. (A.1.3).

c. Gas pressure still dominates with respect to radiation but now the dominant process for opacity is free-free absorption and Eq.(A.1.2) is used.

Solving the system of equations (A.1.1) in the c. region, the predicted H/R aspect ratio reads

$$\frac{H}{R} = 0.001 \left(\frac{R}{GM_\star/c^2}\right)^{1/20} \left(\frac{M_\star}{10^7\,\mathrm{M_\odot}}\right)^{-1/10} \left(\frac{\dot{M}/\dot{M}_\mathrm{Edd}}{0.1}\right)^{1/5} \left(\frac{\alpha}{0.2}\right)^{-1/10} \tag{A.1.4}$$

where $\dot{M}_\mathrm{Edd} = 4\pi G M_\star m_p / c\sigma_\mathrm{T}$ is the Eddington limit for the accretion, Eq. (1.33).

Appendix

B Solutions at Lindblad resonant locations

In the following sections we will discuss in detail the analytic form of the density perturbations produced by barred potentials. With reference to the harmonic decomposition in Eq. 3.36 of the satellite potential, we will not specify the dependence of ω on l, in order to not lose generality: the only effect of the l variable is to produce a secondary set of harmonics with pattern frequencies $\Omega_{\rm p}^{lm}$ given in Eq. (3.37). For a given ω one can always recover the result for a specific l by substituting $\omega = l\Omega_{\rm s}$.

B.1 The WKB and tight-winding approximation

Analytical solutions to the equations (3.31,3.32,3.34) presented in Sec. 3.3 can be found under two main simplifying assumptions. We firstly assume that the solutions are characterized by radial oscillations that can be captured expressing the generic perturbed quantity X_1 as follows

$$X_1(R,\varphi,t) = \underbrace{\chi_1(R) e^{i\int^R k_R(u)du}}_{x_1(R)} e^{i(m\varphi-\omega t)}. \tag{B.1.1}$$

Here the quantity $x_1(R)$ has been divided into a purely real, slowly varying, component $\chi_1(R)$ and into a radially varying complex phase $e^{i\int^R k_R(u)du}$.

This form of the perturbations intrinsically assume that the perturbations are spiral shaped waves. Indeed, it can be easily verified[1], since $e^{i\int^R k_R(u)du}$ represents a phase that progressively grows (or decreases) with the radius, so that at different radii it provides a different amount of azimuthal deviation from the the pure bar-like perturbation $e^{i(m\varphi-\omega t)}$. This causes the winding of the bar perturbation into

[1] We note that, if $k_R = \text{const}$, the complex exponent in Eq. (B.1.1) would read $e^{i(k_R R + m\varphi - \omega t)}$.

a spiral. We will refer to those waves that wind backward with respect to the direction of motion of the gas as "trailing" ($k_R(R) > 0$); in contrast, those that wind in the opposite direction ($k_R(R) < 0$) as "leading".

We introduce now the so called "tight-winding" approximation. The "tight-wound" nature of the perturbation implies that the radial wave number $k_R(R)$ is much greater than the azimuthal one $k_R \gg k_\varphi = m/R$ (this is satisfied obviously for sufficiently small m values). Note that in accretion discs, perturbations with lengthscales lower than the disc vertical displacement H are stable; this intrinsically sets $k_R \lesssim 1/H$. Thus, the tight-winding approximation is valid only for $m \ll (H/R)^{-1}$.

Throughout the next sections, we will also largely use the WKB approximation, which requires that radial oscillations occurs on a lengthscale that is shorter than the variation lengthscale of $\chi(R)$, i.e. characterized by a radial wavenumber $k_R \gg R$. This assumption implies, when computing the radial derivative of the perturbed quantities X_1, that we can neglect $e^{i\int k_R dR}\partial_R \chi_1$ with respect to $\chi_1 \partial_R e^{i\int k_R dR}$ so that

$$\frac{\partial}{\partial R} X_1 \approx i \frac{\partial}{\partial R}\left(\int^R k_R(u)du\right) X_1 = i k_R(R) X_1. \tag{B.1.2}$$

Furthermore, given the fast radial oscillations of X_1, $R^{-1} \ll \partial_R$ so that we are allowed to write $\partial_R(RX_1) \approx R\partial_R X_1$. More generally, this result can be applied to any $f(R)$ that is assumed to satisfy $\partial_R f(R) \ll \partial_R X_1$, implying as a consequence that $f(R)$ can be brought outside and inside the derivative operator as a constant without affecting the final result as follows

$$\frac{\partial}{\partial R}[f(R)X_1] \approx f(R)\frac{\partial}{\partial R} X_1. \tag{B.1.3}$$

We note that assuming that we can safely apply the WKB approximation does not guarantee that we can apply safely the tight-winding one: e.g., when m is sufficiently large that $m/R \gg k_R$ the tigh-winding approximation fails, even though the condition $k_R \gg R^{-1}$ for the applicability of the WKB approximation is satisfied. vice versa, being able to safely apply the tight-winding approximation does not necessarily imply that we can safely apply also the WKB one: e.g., $m = 0$ always satisfy the tight-winding approximation but, nothing prevents to have $k_R \ll R$.

Under these assumptions, radial derivatives of the quantities X_1 can be substituted as

$$\frac{\partial}{\partial R} \to i k_R(R). \tag{B.1.4}$$

In the light of the tight winding assumption[2], we drop the term containing v_φ term on the r.h.s of Eq. 3.30, so that we can re-write the perturbed continuity equation (3.30) as

$$i[m\Omega(R) - \omega]\Sigma_1 = \frac{\partial}{\partial R}(\Sigma_0 v_{R,1}), \tag{B.1.5}$$

[2]Since we know from Eq.s (3.31) and (3.32) that $v_{1,\varphi} \sim v_{1,R}$, assuming that perturbations are thight wound ($k_\varphi \ll k_R$) also implies that $im/R v_{1,\varphi} \ll k_R v_R$.

Finally, using the tight-winding prescription $k_\varphi \ll k_R$, the substitution in Eq. B.1.3, and assuming a WKB form of the perturbation Φ_{Σ_1}, we can rewrite the Poisson's equation (3.14) as

$$-k_R^2(R)\Phi_{\Sigma_1} + \frac{\partial^2 \Phi_{\Sigma_1}}{\partial z^2} = 4\pi G \Sigma_1 \delta(z). \tag{B.1.6}$$

B.2 Solving the Poisson's equation

This section is dedicated to the solution of the Poisson's equation in the simplified form provided in Eq. (B.1.6). The following standard approach to its solution can be found in Armitage (2010). The WKB form of the self-gravitating potential is

$$\Phi_{\Sigma_1} = \underbrace{V(R) e^{i\int^R k_R(u)du}}_{\phi_{\Sigma_1}(R)} e^{i(m\varphi - \omega t)}, \tag{B.2.1}$$

Since $\delta(z \neq 0) = 0$, the solution of Φ_{Σ_1} for $z \neq 0$ satisfies[3]

$$\frac{\partial^2 \Phi_{\Sigma_1}}{\partial z^2} = k_R^2 \Phi_{\Sigma_1}, \tag{B.2.2}$$

that implies[4]

$$\Phi_{\Sigma_1} = V(R) e^{-|k_R(R)z|} e^{i\int^R k_R(u)du} e^{i(m\varphi - \omega t)}. \tag{B.2.3}$$

We find the dependence of $\Phi_{\Sigma_1}(R)$ on Σ_1 by integrating Eq. (B.1.6) along z across the infinitesimally thin disc layer:

$$\lim_{\epsilon \to 0} \int_{-\epsilon}^{+\epsilon} \frac{\partial^2 \Phi_{\Sigma_1}}{\partial z^2} dz = \lim_{\epsilon \to 0} \int_{-\epsilon}^{+\epsilon} \left[k_R^2(R)\Phi_{\Sigma_1} + 4\pi G \Sigma_1 \delta(z) \right] dz. \tag{B.2.4}$$

Assuming the functional form of Φ_{Σ_1} in Eq. (B.2.3), we find the solution in the $z = 0$ plane, by further developing Eq. (B.2.4) as

$$\lim_{\epsilon \to 0} \frac{\partial \Phi_{\Sigma_1}}{\partial z}\bigg|_{-\epsilon}^{+\epsilon} \equiv \lim_{\epsilon \to 0} -2|k_R(R)| \Phi_{\Sigma_1} e^{-|k_R \epsilon|} = 4\pi G \Sigma_1, \tag{B.2.5}$$

that finally becomes

$$\Phi_{\Sigma_1} = -\frac{2\pi G \Sigma_1}{|k_R(R)|}. \tag{B.2.6}$$

Where Σ_1 has the same WKB form of Φ_{Σ_1}. Keeping in mind that the substitution in Eq. (B.1.4) can be applied in both directions, we can state Eq. (B.2.6) also as

$$i \operatorname{sgn}[k_R(R)] \frac{\partial \Phi_{\Sigma_1}}{\partial R} = 2\pi G \Sigma_1. \tag{B.2.7}$$

[3]Even though the problem is 2D and the domain is the plane $z = 0$, it is required add a spurious vertical dependence of Φ_{Σ_1} in order deal with the presence of the $\delta(z)$ function while solving the equation.

[4]The sign of the z-exponent in $e^{-|k_R(R)z|}$ is set in order to have that Φ_{Σ_1} vanishes for $z \to \pm\infty$.

Eq.s (B.2.6) and (B.2.7) set the relation between Σ_1 and Φ_{Σ_1}.We introduce here an additional constraint on the relation between Φ_{Σ_1} and $k_R(R)$ that will be useful below:

$$\frac{\partial}{\partial R}\left[R\Phi_{\Sigma_1}^2\left(\frac{c_{s,0}^2|k_R(R)|}{\pi G\Sigma_0}-1\right)\right]=0. \tag{B.2.8}$$

This relation can be obtained[5] by considering a more accurate version of Eq. (B.2.6), which instead is only $O(|k_R R|^{-1})$ accurate (see Eq. 17 in Goldreich & Tremaine (1979), which is $O(|k_R R|^{-2})$ accurate).

B.3 Density waves equation

Our approximations explicitly assume the dependence of the density perturbation Σ_1 on φ and t. As a consequence, solving Eq. (B.1.5) implies determining the radial behaviour of the perturbations.

We rewrite Eq. (B.1.5) as a function Φ_{Σ_1} exploiting Eq. (B.2.7). Under the assumptions in Sec. B.1, substituting Eq. (3.28) in Eq. (B.1.5), with the aid also of Eq. (B.2.7) in order to rewrite h_1 as a function of Φ_{Σ_1}, we get[6]

$$\underbrace{\frac{i\,\mathrm{sgn}\,k_R(R)}{2\pi G}\frac{\partial\Phi_{\Sigma_1}}{\partial R}}_{\Sigma_1}=\frac{\partial}{\partial R}\left\{\overbrace{\frac{\Sigma_0}{D(R)}\left[\frac{\partial}{\partial R}+\frac{2m\Omega}{R(m\Omega(R)-\omega)}\right]\left(\Phi_{\Sigma_1}+\underbrace{\frac{i\,\mathrm{sgn}\,k_R(R)c_s^2}{2\pi G\Sigma_0}\frac{\partial\Phi_{\Sigma_1}}{\partial R}}_{h_1}+\Phi_1^m\right)}^{v_{R,1}}\right\}. \tag{B.3.1}$$

We define the quantity Ψ_m with the aim to shorten the expression of the differential operators acting on Φ_1^m

$$\frac{1}{R}\left[R\frac{\partial\Phi_1^m}{\partial R}+\frac{2m\Omega}{m\Omega(R)-\omega}\Phi_1^m\right]\equiv\frac{\Psi_m(R)}{R}, \tag{B.3.2}$$

Far from the corotation resonance, Eq. (B.3.1) can be further simplified, using the tight-winding assumption: since $\Phi_{\Sigma_1}+h_1\ll\partial_R(\Phi_{\Sigma_1}+h_1)$, we can drop the remaining terms[7] containing $2m\Omega(\Phi_{\Sigma_1}+h_1)/[R(m\Omega(R)-\omega)]$ (Goldreich & Tremaine 1979; Shu 1984; Ward 1986).

In the light of these considerations, by equating the arguments of the radial derivative on the l.h.s. with the one on the r.h.s., Eq. (B.3.1) can be approxi-

[5] See Appendix J of Binney & Tremaine 2008 for further details.

[6] For a matter of clarity, the property in Eq. (B.1.3) allows us to bring $m\Omega(R)-\omega$ out from the derivative operator. Anyway we do not make any assumption on $\partial_R\Sigma_0/D$. We also remind the reader that the term containing $v_{\varphi,1}$ as been dropped as a consequence of our assumptions when deriving Eq. (B.1.5).

[7] Artymowicz (1993a) showed then that this approximation is perfectly reasonable for low m values, but it ends up with neglecting some important terms when the value of m is sufficiently large.

mated[8] as

$$\left(\frac{\partial^2}{\partial R^2} - i\,\mathrm{sgn}\,k_R\frac{2\pi G\Sigma_0}{c_\mathrm{s}^2}\frac{\partial}{\partial R} - \frac{D(R)}{c_\mathrm{s}^2}\right)\Phi_{\Sigma_1} = i\,\mathrm{sgn}\,k_R\frac{2\pi G\Sigma_0}{c_\mathrm{s}^2}\frac{\Psi_m(R)}{R}. \quad \text{(B.3.3)}$$

It is very easy to recognize that Eq. (B.3.3) is solved by a density wave (Φ_{Σ_1} directly relates with the perturbed density Σ_1).

B.4 Solutions far from resonances

Insights about the asymptotic behaviour of the solutions to Eq. (B.3.3) at large distance from resonances can be successfully gained by studying separately the wave-like solutions of the homogeneous equation ($\Phi_1^m = 0 \rightarrow \Psi_m = 0$) and the non-wave solutions of the non-homogeneous one ($\Psi_m \neq 0$).

Regarding the non-wave solution, it can be shown that, at distances $|R - R_L|/R_L \gg (c_\mathrm{s}/\Omega R)^{1/2}$ and $|R - R_c|/R_c \gg c_\mathrm{s}/\Omega R$, the non-wave solution can be found solving Eq. B.3.3 neglecting the first two terms in the l.h.s.: i.e., shutting down the effects of pressure (h_1) and self-gravity (Φ_{Σ_1}), conserving only the effects introduced by the perturbing gravitational potential (Φ_1^m) Goldreich & Tremaine (1979). This part of the solution has a radial power-law decrease at large distance from the resonances (Goldreich & Tremaine 1979; Shu 1984; Ward 1986). In the non-wave solution Σ_1 and Φ_1^m are in phase, as a consequence, no angular momentum flux is associated with this solution (Goldreich & Tremaine 1979). Overall, this condition produces a vanishing net torque for a matter of symmetry (see also Sec. 3.6).

Regarding wave-like solutions to the homogeneous equation, we can gain some insights about the dispersion relation of the spiral waves by applying substitution in Eq. (B.1.4) to Eq. (B.3.3). Rearranging the terms we obtain

$$[m\Omega(R) - \omega]^2 = \kappa^2 - 2\pi G\Sigma_0|k_R(R)| + c_{\mathrm{s},0}^2 k_R^2(R). \quad \text{(B.4.1)}$$

This equation represents the dispersion relation for spiral density waves (Eq. B.1.1) in a self-gravitating disc. With $m = 0$ the disc is stable with respect to the gravitational instability if

$$Q \equiv \frac{c_\mathrm{s}\kappa}{\pi G\Sigma_0} > 1, \quad \text{(B.4.2)}$$

obtained requiring the r.h.s. of Eq. (B.4.1) to be positive for all the values of k_R.

In the left panel of Fig. B.1 we plot the quantity $[m\Omega(R) - \omega]/\kappa$ as a function of k_R/k_Q where $k_Q = \pi \mathrm{G}\Sigma_0/c_\mathrm{s}^2$. The plot shows three main regimes of propagation of the density perturbation:

- For values of $Q > 1$, all values of k_R are allowed, but the waves do not propagate throughout a prohibited region. This region covers the entire area between the ILR and OLR for values of $Q \gg 1$. Furthermore, the left panel of Fig. B.1 also shows that long-waves ($k_R/k_Q \ll 1$) do not propagate outside the region between the ILR and the OLR.

[8] Further simplifications have been done by assuming that the perturbation is tightly wound, which allows us to move slowly varying quantities in and out the radial derivative sign.

- For values of $Q = 1$, waves can propagate up to the corotation resonance, and no prohibited region is present.

- For values of $Q < 1$, some wavelenghts k_R provide $(m\Omega - \omega)^2 < 0$, implying exponential growth of the perturbation, causing the disc material to clump setting in the gravitational collapse of the perturbations at those wavelengths.

We can study the group velocity $c_{\rm g}$ of the wave (Goldreich & Tremaine 1979):

$$c_{\rm g} \equiv \frac{\partial \omega}{\partial k_R} = \operatorname{sgn} k_R \frac{\pi G \Sigma_0 - |k_R(R)| c_{\rm s}^2}{m\Omega(R) - \omega}. \tag{B.4.3}$$

The group velocity sets the direction of propagation of the perturbation. From Eq. (B.4.3) we see that it depends on three main properties of the wave and of the disc. Firstly, the sign of $k_R(R)$, i.e. whether the spiral wave is leading or trailing. Secondly, it depends on whether the wave is a gravity or a pressure wave: i.e., whether pressure effects ($c_{\rm s}^2$) are dominant with respect to gravity ones ($\pi G \Sigma_0$). Finally, on the wavelength of the perturbation: $k_R \gg k_Q$, are short waves $k_R \ll k_Q$ are long waves.

Looking at the power of k_R terms in Eq. (B.4.1) it is easy to understand that for small k_R (long waves), the dispersion relation is dominated by the self-gravity term, while for large k_R (short waves) the dispersion relation is dominated by the pressure term.

It is interesting to note that the solutions to the homogeneous equation provide predictions that do not depend on the mechanisms that have excited the wave, describing only how both leading and trailing perturbations propagate throughout the disc. However, studying in more detail the solutions at resonant locations, it can be shown (see Sec. B.5.1 and B.5.2) that no leading perturbations are excited by the perturbing potential (Goldreich & Tremaine 1979; Meyer-Vernet & Sicardy 1987). Therefore, density perturbations are excited as trailing waves with long wavelengths ($k_R \sim 0$) around $R \approx R_L$. After the excitation, they propagate toward the CR and are reflected at the boundary with the forbidden zone where $c_{\rm g}$ changes sign. Then, they propagate away from the CR at shorter wavelengths as pressure waves. In the right panel of Fig. B.1 a sketch of the wave direction of propagation we just discussed can be found.

B.5 Density waves

The simplified approach used to study the properties of wave and non-wave solutions cannot be applied in domains containing resonant regions, where the coupling between the disc and the perturbing satellite is stronger. In these specific regions, the amplitude of both wave and non-wave like solutions to Eq. (B.3.3) is set by the effects of the perturbing potential Φ_1^m.

The solutions to Eq. (B.3.3) is given in integral form by Goldreich & Tremaine (1979). Nevertheless, we find more instructive to separate long gravity waves (i.e. $k_R \ll \pi G \Sigma_0 / c_{\rm s}^2$) and short pressure waves ($k_R \gg \pi G \Sigma_0 / c_{\rm s}^2$) following the

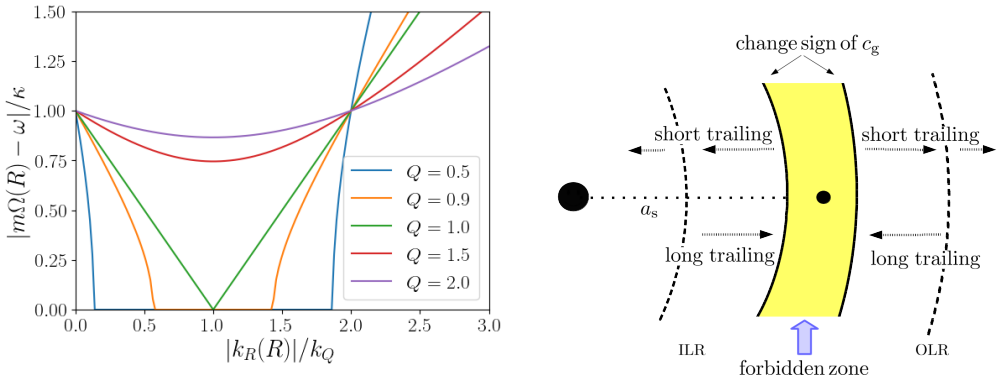


Figure B.1: Left panel: $|m\Omega(R)-\omega|/\kappa$ as a function of $k_R(R)/k_Q$ where $k_Q = \pi G\Sigma_0/c_s^2$ for different values of Q. Right panel: sketch of the direction of long and short trailing waves; the yellow shaded area represents the region where waves cannot propagate because the wave group velocity changes sign. The forbidden region has nothing to do with the condition $\kappa^2 + c_s^2 k_R^2(R) - 2\pi G\Sigma_0|k_R(R)| < 0$, that instead sets the wavelengths for which the perturbations are gravitationally unstable.

discussion given by Shu (1984) and Ward (1986), respectively, but using a unified notation[9].

B.5.1 Long gravity waves

We want to find the solution to Eq. (B.3.3) neglecting the pressure term, i.e. the term $\partial^2/\partial R^2$. The equation we want to solve is

$$\frac{\partial \Phi_{\Sigma_1}}{\partial R} - i\frac{\operatorname{sgn} k_R D(R)}{2\pi G\Sigma_0}\Phi_{\Sigma_1} = -\frac{\Psi_m(R)}{R}. \tag{B.5.1}$$

We now study the solution at Lindblad resonances. To do so, we substitute R with the variable x that satisfies

$$x = \frac{R - R_L}{R_L}. \tag{B.5.2}$$

We approximate D in the region of the resonance R_L as

$$D(R) = (R - R_L)\frac{\partial}{\partial R}D(R_L) = xR_L\frac{\partial}{\partial R}D(R_L). \tag{B.5.3}$$

We define the quantity

$$\mathcal{D} \equiv R_L\left|\frac{\partial}{\partial R}D(R_L)\right|, \tag{B.5.4}$$

so that we can write $D(R) = \epsilon\mathcal{D}x$, where $\epsilon = \pm 1$ at ILRs and at OLRs respectively (the sign of ϵ depends on the sign of $\partial_R D$). Eq. (B.5.1) becomes[10]

$$\left(\frac{\partial}{\partial x} - i\frac{\operatorname{sgn} k_R R_L \epsilon\mathcal{D}}{2\pi G\Sigma_0}x\right)\Phi_{\Sigma_1} = -\Psi_m(R_L). \tag{B.5.5}$$

[9] See also the textbook Hahn (2009).

[10] In the r.h.s. $\Psi_m(R_L)/(1+x) \approx \Psi_m(R_L)$ since $1 + x \approx 1$.

This equation can be solved using the integrating factor technique (Shu 1984). Indeed, it can be shown that for a differential equation of the type

$$\frac{dy}{dx} + p(x)y = \mathcal{Q}(x), \tag{B.5.6}$$

its solution $y(x)$ is given by

$$y(x) = e^{-\mathcal{P}(x)} \int_{C_0}^{x} e^{\mathcal{P}(t)} \mathcal{Q}(t) dt, \tag{B.5.7}$$

where $\mathcal{P}(x) = \int^x p(t)dt$, C_0 is an integration constant. Applying this technique to Eq. (B.5.5), we firstly identify the integrating factor $\mathcal{P}(x)$ as

$$\mathcal{P}(x) \equiv \int^x -i \frac{\operatorname{sgn} k_R R_L \epsilon \mathcal{D}}{2\pi G \Sigma_0} t \, dt = -i \frac{\operatorname{sgn} k_R R_L \epsilon \mathcal{D}}{2\pi G \Sigma_0} \frac{x^2}{2}. \tag{B.5.8}$$

We simplify the notation making the substitution

$$u = \xi \sqrt{\frac{R_L \mathcal{D}}{2\pi G \Sigma_0}} \frac{x}{\sqrt{2}}, \tag{B.5.9}$$

where $\xi = \operatorname{sgn}(\epsilon k_R)$. The general solution to Eq. (B.5.5) is

$$\Phi_{\Sigma_1}(u) = -\left[\Psi_m(R_L) \xi \left(\sqrt{\frac{R_L \mathcal{D}}{4\pi G \Sigma_0}} \right)^{-1} e^{i\xi u^2} \int_{C_0}^{u} e^{-i\xi t^2} dt \right] e^{i(m\phi - \omega t)}. \tag{B.5.10}$$

We define $H_\xi(u)$

$$H_\xi(u) \equiv \frac{1}{\sqrt{\pi}} e^{i\xi u^2} \int_{C_0}^{u} e^{-i\xi t^2} dt. \tag{B.5.11}$$

H_ξ can be expressed as a function of the Fresnel type integrals (Meyer-Vernet & Sicardy 1987). In Fig. (B.2) a plot of the function H_+ can be found. H_- can be obtained exploiting the complex conjugate operation as $H_- = H_+^\star$. Imposing radiative boundary conditions at Lindblad resonances, so that waves can only radiate away from the excitation region, it can be shown that no leading waves emerge from Lindblad resonances. This intrinsically sets $\xi = \epsilon$. As previously shown, no waves can be found in the region of $u < 0$. It can be shown that requiring the real part in Eq. (B.5.11) to vanish for $u \to -\infty$ implies $C_0 = -\infty$.

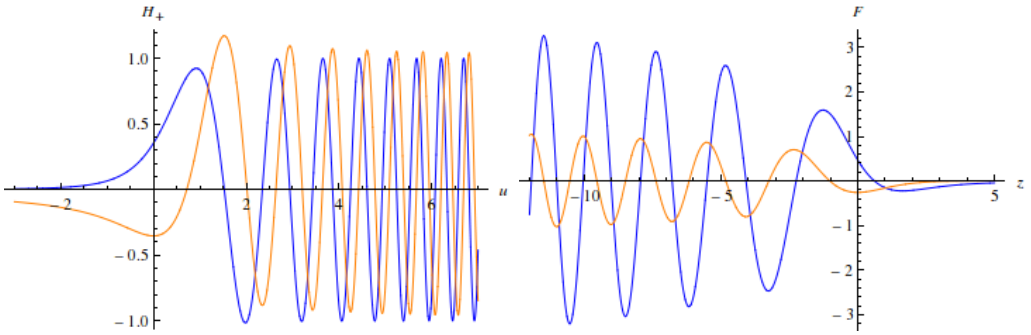
The solution to Eq. (B.5.5) can be now written as

$$\Phi_{\Sigma_1} = -2\pi \sqrt{\frac{G\Sigma_0}{R_L \mathcal{D}}} \xi \Psi_m(R_L) H_\xi(u) e^{i(m\phi - \omega t)}, \tag{B.5.12}$$

Eq. (B.5.12) is a wave for $u > 0$, while the non-wave part vanishes for $u \to -\infty$ (see the left panel of B.2).

Far from the resonance, when $u \to \infty$ it can be shown that the asymptotic behaviour of Eq. (B.5.12) is (Shu 1984):

$$\Phi_{\Sigma_1} \sim -\xi 2\pi \sqrt{\frac{G\Sigma_0}{R_L \mathcal{D}}} \Psi_m(R_L) \exp\left[i\xi \left(u^2 + \frac{\pi}{4} \right) \right] e^{i(m\phi - \omega t)}, \quad \text{for } u \to \infty. \tag{B.5.13}$$



.

Figure B.2: Left panel: Plot of the function H_+ function in Eq. (B.5.11). The blue line represents $\mathrm{Re}(H_+)$, the orange one $\mathrm{Im}(H_+)$. H_- can be obtained inverting the sign of $\mathrm{Im}(H_+)$, keeping in mind that $u \propto \xi x$, so that the wavy part is be in the negative side of the physical spatial coordinate x. Waves propagate away from Lindblad resonances $x = 0$ toward the corotation region. Right Panel: function F in Eq. (B.5.22) for $\epsilon = +1$ (i.e., ILR). Again, the blue curve represents $\mathrm{Re}(F)$ and the orange one $\mathrm{Im}\, F$. In this case, the waves propagate away from the resonance at $z = 0$ in the opposite direction with respect to the corotation resonance.

A comparison with Eq. (B.2.1) allows us to finally define the asymptotic amplitude of the self-gravitating potential wave far from the resonance $\mathcal{V}_\mathrm{g} \equiv |\Phi_{\Sigma_1}(\infty)|$, so that $\mathcal{V}_\mathrm{g}$ reads

$$\mathcal{V}_\mathrm{g} = 2\pi\sqrt{\frac{G\Sigma_0}{R_L\mathcal{D}}}\Psi_m(R_L). \tag{B.5.14}$$

It should be noted that $\mathcal{V}_\mathrm{g} \equiv |\Phi_{\Sigma_1}|$ represents the slowly varying component of the radial part in the WKB form of Φ_{Σ_1} (see Eq. B.1.1)

B.5.2 Short pressure waves

This time we neglect the self-gravitating part in Eq. (B.3.1), i.e. the $\partial/\partial R$ term in Eq. (B.3.3). In this context, Φ_{Σ_1} has no physical meaning, since we are here neglecting the effects of self gravity. We restate Eq. (B.3.3) using the substitution $(c_\mathrm{s}^2/\Sigma_0)\Sigma_1 = R_L\partial_R\zeta$, where ζ is a generic function of R.

$$\frac{\Sigma_0}{c_\mathrm{s}^2}R_L\frac{\partial\zeta}{\partial R} = \frac{\partial}{\partial R}\left[\frac{\Sigma_0}{D(R)}\frac{\partial}{\partial R}\left(R_L\frac{\partial\zeta}{\partial R} + \frac{\Psi_m(R)}{R}\right)\right]. \tag{B.5.15}$$

Equating the arguments of the radial derivative on the l.h.s. with the one on the r.h.s. we end up with the following equation[8]

$$R_L\frac{\partial^2\zeta}{\partial R^2} - \frac{D(R)}{c_\mathrm{s}^2}R_L\zeta = -\frac{\Psi_m(R)}{R}. \tag{B.5.16}$$

As in the previous section, we make the substitution $R \to x$ as prescribed in Eq. (B.5.2). Using then Eq. (B.5.4), we get

$$\frac{\partial^2\zeta}{\partial x^2} - \frac{\epsilon R_L^2\mathcal{D}x}{c_\mathrm{s}^2}\zeta = -\Psi_m(R_L). \tag{B.5.17}$$

We substitute $z \equiv \epsilon(R_L^2\mathcal{D}/c_\mathrm{s}^2)^{1/3}x$. Eq. (B.5.17) becomes

$$\frac{\partial^2\zeta}{\partial z^2} - z\zeta = -\Psi_m(R_L)\left(\frac{c_\mathrm{s}^2}{R_L^2\mathcal{D}}\right)^{2/3}. \tag{B.5.18}$$

The general solution to this equation is

$$\zeta(z) = \pi\Psi_m(R_L)\left(\frac{c_\mathrm{s}^2}{R_L^2\mathcal{D}}\right)^{2/3}\left[\mathrm{Ai}(z)\left(a + \int_0^z \mathrm{Bi}(t)\,dt\right) - \mathrm{Bi}(z)\left(b + \int_0^z \mathrm{Ai}(t)\,dt\right)\right] e^{i(m\phi-\omega t)}, \tag{B.5.19}$$

where $\mathrm{Ai}(z)$ and $\mathrm{Bi}(z)$ are Airy functions. We apply then the radiative boundary condition (again no leading waves satisfy this condition), so that waves can only radiate away from the excitation region and we prescribe that the solution vanishes for $z \to \infty$ due to the considerations on the direction of propagation of pressure waves we did in Sec. B.4. It can be shown that this procedure leads to $a = i\epsilon$ and $b = -1/3$ (Ward 1986). The solution to Eq. (B.5.18) is

$$\zeta(z) = \pi\Psi_m(R_L)\left(\frac{c_\mathrm{s}^2}{R_L^2\mathcal{D}}\right)^{2/3}\left[i\epsilon\mathrm{Ai}(z) + \mathrm{Gi}(z)\right] e^{i(m\phi-\omega t)}, \tag{B.5.20}$$

where $\mathrm{Gi}(z)$ is the Scorer function. We can now express Σ_1 as

$$c_\mathrm{s}^2\frac{\Sigma_1}{\Sigma_0} \triangleq \frac{\partial\zeta}{\partial z}\frac{\partial z}{\partial x} = \pi\Psi_m(R_L)\left(\frac{c_\mathrm{s}^2}{R_L^2\mathcal{D}}\right)^{1/3} F(z)e^{i(m\phi-\omega t)}, \tag{B.5.21}$$

where $F(z)$ is given by (see also Fig. B.2)

$$F(z) = i\frac{\partial}{\partial z}\mathrm{Ai}(z) + \epsilon\frac{\partial}{\partial z}\mathrm{Gi}(z). \tag{B.5.22}$$

Using the properties of $\partial_z\mathrm{Ai}(z)$ and $\partial_z\mathrm{Gi}(z)$ (Abramowitz & Stegun 1968) it can be shown that for $z \to -\infty$

$$\begin{aligned} F(z) &\sim \frac{1}{\sqrt{\pi}}|z|^{1/4}\left\{i\cos\left[\frac{2}{3}z^{3/2} + \frac{\pi}{4}\right] + \epsilon\sin\left[\frac{2}{3}z^{3/2} + \frac{\pi}{4}\right]\right\} = \\ &= i\frac{1}{\sqrt{\pi}}|z|^{1/4}\exp\left[-i\epsilon\left(\frac{2}{3}z^{3/2} + \frac{\pi}{4}\right)\right]. \end{aligned} \tag{B.5.23}$$

It can be shown that the asymptotic behaviour of Σ_1 far away from the resonance is (Ward 1986):

$$\Sigma_1 \sim i\frac{\sqrt{\pi}\Sigma_0\Psi_m}{c_\mathrm{s}^2}\left(\frac{c_\mathrm{s}^2}{R_L^2\mathcal{D}}\right)^{1/3}|z|^{1/4}\exp\left[-i\epsilon\left(\frac{2}{3}z^{3/2} + \frac{\pi}{4}\right)\right] e^{i(m\phi-\omega t)}, \text{ for } z \to -\infty, \tag{B.5.24}$$

which is a wave with an amplitude $\mathcal{A}_\mathrm{p} \equiv |\Sigma_1(-\infty)|$ that, recasted as a function of x, reads

$$\mathcal{A}_\mathrm{p} = \sqrt{\pi}\frac{\Sigma_0\Psi_m}{c_\mathrm{s}^2}\left(\frac{c_\mathrm{s}^2}{R_L^2\mathcal{D}}\right)^{1/4}|x|^{1/4}. \tag{B.5.25}$$

It should be noticed that $\mathcal{A}_\mathrm{p} \equiv |\Sigma_1|$ represents the slowly varying component of the radial part in the WKB form of Σ_1 (see Eq. B.1.1).

B.6 Angular momentum transport by density waves

The theoretical framework regarding the angular momentum transport provided by spiral waves propagating in discs was originally developed by Lynden-Bell & Kalnajs (1972) for spiral galaxies. Exploiting their formalism, Goldreich & Tremaine (1979) studied the transport of angular momentum in discs where a perturbing satellite potential is responsible for the excitation of the density spiral waves. The angular momentum flux due to the disc self-gravity at a specific radius R carried by a spiral wave is given by

$$\mathcal{F}_g \equiv \frac{1}{4\pi G}\int_0^{2\pi}\int_{-\infty}^{\infty} \mathrm{Re}\left(\frac{\partial \Phi_{\Sigma_1}}{\partial \varphi}\right)\mathrm{Re}\left(\frac{\partial \Phi_{\Sigma_1}}{\partial R}\right) r dz\, d\varphi = \mathrm{sgn}\, k_R \frac{mR|\Phi_{\Sigma_1}|^2}{4G}. \tag{B.6.1}$$

The advective flux of angular momentum due to the radial motion of material across a non-zero radial gradient of angular momentum carried by a spiral density wave is given instead by[11]

$$\mathcal{F}_a \equiv R^2\Sigma_0 \int_0^{2\pi} \mathrm{Re}(v_{R,1})\,\mathrm{Re}(v_{\varphi,1})\, d\varphi = -\frac{\pi m R \Sigma_0 k_R}{D}\left|\Phi_{\Sigma_1} + \frac{c_\mathrm{s}^2}{\Sigma_0}\Sigma_1\right|^2. \tag{B.6.2}$$

For a self-gravitating disc, exploiting the WKB form of the quantities involved in these equations, and using $D = -c_\mathrm{s}^2 k_R^2 + 2\pi G\Sigma_0|k_R|$ from the dispersion relation in Eq. (B.4.1) to eliminate D in Eq. (B.6.2), it is possible to reduce the total angular momentum flux to

$$\mathcal{F}_\mathrm{tot} \equiv \mathcal{F}_g + \mathcal{F}_a = -\,\mathrm{sgn}\, k_R \frac{mR|\Phi_{\Sigma_1}|^2}{4G}\left(1 - \frac{c_\mathrm{s}^2|k_R|}{\pi G\Sigma_0}\right). \tag{B.6.3}$$

Comparing Eq. (B.2.8) and Eq. (B.6.3) we see that the angular momentum flux carried by a free wave is conserved throughout the entire disc.

As previously mentioned, density waves are excited at Lindblad resonances where the exchange of angular momentum between the satellite and the disc is mostly effective. The angular momentum is then transported away by the density waves. Due to the conservation of the angular momentum during the disc-satellite interaction, in absence of dissipation, the angular momentum flux the wave carries away from its excitation region must equal the tidal torque the planet exerts on the disc.

We can compute the angular momentum flux carried by the wave excited by an m-barred perturbing potential by deriving the asymptotic amplitudes of pure gravity and pressure waves $\mathcal{V}_g$ (Eq. B.5.14) and $\mathcal{A}_p$ (Eq. B.5.25), respectively, and substituting them into Eq.s (B.6.1) and (B.6.2).

We rewrite the term in the bracket of Eq. (B.6.3) as $1-|k_R|/k_Q$. For $|k_R| \ll k_Q$ we are in the regime of long gravity waves, therefore we can neglect the term $|k_R|/k_Q$ and substitute $|\Phi_{\Sigma_1}|$ with the asymptotic amplitude of gravity wave $|\mathcal{V}_\mathrm{g}|$ in Eq. (B.5.14). The asymptotic amplitude of the gravity wave reads:

$$|\Phi_{\Sigma_1}| \equiv |\mathcal{V}_\mathrm{g}| = 2\pi\sqrt{\frac{G\Sigma_0}{R_L\mathcal{D}}}|\Psi_m(R_L)| \tag{B.6.4}$$

[11] A detailed derivation of both Eq.s (B.6.1) and (B.6.2) can be found in Hahn (2009).

Substituting Eq. (B.6.4) into Eq. (B.6.3), neglecting $|k_R|/k_Q \ll 1$ the angular momentum flux across R_L reads:

$$\mathcal{F}_{\rm tot,g} = -\operatorname{sgn} k_R \frac{\pi^2 m \Sigma_0}{\mathcal{D}} |\Psi_m(R_L)|^2, \tag{B.6.5}$$

where $\mathcal{D}$ is defined in Eq. (3.47) (see also Eq. (B.5.4).

On the other hand, when $|k_R|/k_Q \gg 1$, i.e. in the regime of short pressure waves, or in general when self-gravity is neglected, the asymptotic amplitude of the density wave is given by (Eq. B.5.25):

$$|\Sigma_1| \equiv \mathcal{A}_{\rm p} = \sqrt{\pi} \frac{\Sigma_0 |\Psi_m|}{c_{\rm s}^2} \left(\frac{c_{\rm s}^2}{R_L^2 \mathcal{D}} \right)^{1/4} \left| \frac{R - R_L}{R_L} \right|^{1/4}. \tag{B.6.6}$$

Substituting $|\Sigma_1|$ in Eq. (B.6.3) and neglecting the first term in the bracket since $|k_R|/k_Q \ll 1$, we obtain[12] the angular momentum flux for short pressure waves at R_L

$$\mathcal{F}_{\rm tot,p} = \frac{m \pi^2 \Sigma_0}{\mathcal{D}} |\Psi_m(R_L)|^2, \tag{B.6.7}$$

It should be noticed that the total flux for short-pressure waves is purely advective $\mathcal{F}_{\rm tot,p} = \mathcal{F}_a$. Eq.s (B.6.5) and (B.6.7) interestingly have the same magnitude but different sign. Indeed, the angular momentum is transferred at resonant locations by the perturbing potential Φ_1^m. This excites a low k_R wave that transports the angular momentum away from the excitation region; the direction of propagation of the wave depends on the magnitude of k_R through its group velocity (Eq. B.4.3), as well as the angular momentum flux to it associated.

[12] To do so, we note that $\epsilon \mathcal{D}(R - R_L)/R_L = -c_{\rm s}^2 k_R^2$, and recall the relationship between Φ_{Σ_1} and Σ_1 derived in Eq. B.2.6.

Appendix C

Secular eccentricity evolution during disc-planet interaction

With reference to Chapter 7 we report here two further developments of the model provided in Sec. 7.4 concerning the limiting case $q = 0$ and analytical approximations of the pericentre phase evolution in three limiting cases.

C.1 The limiting case $q = 0$

To understand the physical meaning of the two eigen-modes, it is instructive to consider the limit $q = 0$. This case is particularly interesting from a physical point of view because is the case in which the second outer planet has negligible mass. Thus it constitutes the reference situation for all those simulations in which the binary is kept fixed on its initial orbit with constant orbital eccentricity and semi-major axis (D'Angelo et al. 2006; Müller & Kley 2013; Duffell & Dong 2015; Miranda et al. 2017; Thun et al. 2017). The problem becomes in fact the classical restricted three body problem in which a test particle orbits a binary object.

To further simplify the equations we assume also $\alpha \ll 1$, so that $\beta \approx 5/4\alpha$ (Murray & Dermott 1999). In this limit the eigen-frequencies reads

$$\omega_{s0} = 0, \tag{C.1.1}$$

$$\omega_{f0} = \Omega_{\rm sec}\sqrt{\alpha}, \tag{C.1.2}$$

while the non-unit component of the eigenvectors η reads

$$\eta_{s0} = \frac{4}{5\alpha}, \tag{C.1.3}$$

$$\eta_{f0} = 0. \tag{C.1.4}$$

With reference to Eq. (7.18), these values for eigen-frequencies and eigen-vectors imply the following two limiting cases: first, the system is completely in the fast

f mode, the inner planet (M_p) has $e_\mathrm{p} = 0$ while the outer planet (for this case $M_\mathrm{d} = 0$) orbits with arbitrary eccentricity with a pericentre precession frequency given by Eq. C.1.2, no oscillations of the outer-planet eccentricity are observed. Second, the system is completely in the slow s mode, the inner planet has an eccentricity $e_\mathrm{p} = C_1\eta_{s0} \neq 0$ while the outer one orbiting with an eccentricity $e_\mathrm{d} = e_\mathrm{p}/\eta_{s0} = C_1$, its pericentre phase does not experience any form of precession, no oscillations of the outer-planet eccentricity are observed. In both cases, obviously the inner planet pericentre phase do not experience any precession since in this limit it does not feel the the presence of the outer planet at all.

In a mixed situation (the system is both in the slow and in the fast mode, i.e. $e_\mathrm{p} \neq 0,\ e_\mathrm{d} \neq e_\mathrm{p}/\eta_{s0}$) if $e_\mathrm{d} > e_\mathrm{p}/\eta_\mathrm{s}$ the outer planet will experience a complete precession (the pericentre phase of the outer planet will complete a revolution of 360° around the central star) with a precession rate given by C.1.2. In contrast if $e_\mathrm{d} < e_\mathrm{p}/\eta_{s0}$, the outer planet librates around the pericentre phase of the inner planet spanning a range of phases that becomes progressively smaller as $e_\mathrm{d} \to e_\mathrm{p}/\eta_{s0}$.

We can further expand ω_{f0} in Eq. C.1.2 expliciting Ω_sec

$$\omega_{f0} = \frac{3}{4}\alpha^{7/2}\Omega_\mathrm{p}\frac{M_\mathrm{p}}{M_\star}, \tag{C.1.5}$$

associated with a precession period t_prec of the outer planet pericentre phase given by

$$t_\mathrm{prec} = \frac{4}{3}\alpha^{-7/2}\left(\frac{M_\mathrm{p}}{M_\star}\right)^{-1} t_\mathrm{orb}, \tag{C.1.6}$$

which is perfectly consistent with the precession rate predicted by Moriwaki & Nakagawa (2004) for the restricted three body problem apart from higher order corrections in e in the expansion of the perturbing potential and with the interpretation of the precession frequency in the fixed planet simulations in Miranda et al. (2017) and Thun et al. (2017) (equivalent to the $q = 0$ case). Furthermore, e_p/η_{s0} has the equivalent role of e_forced in Moriwaki & Nakagawa (2004) and also in this case the two expressions are perfectly consistent apart from higher order corrections in e in the expansion of the perturbing potential.

C.2 Phase evolution

In this section we will present some analytical approximations of Eq. (7.26) and (7.27) through which it will be possible to obtain Eq. (7.28). In order to simplify the notation, we will refer to the modulus part of the eigen-modes as $\mathcal{A}$ and $\mathcal{B}$, implying that $\mathcal{A} = \eta_\mathrm{s}C_1(t)$ and $B = \eta_\mathrm{f}C_2(t)$ for the planet equations, and that $\mathcal{A} = C_1(t)$ and $\mathcal{B} = C_2(t)$ for the disc ones. With this simplifying substitution Eq. (7.21) reads

$$|E|e^{i\Phi} = \mathcal{A}e^{i\omega_\mathrm{s}t} + \mathcal{B}e^{i\omega_\mathrm{f}t}, \tag{C.2.1}$$

which can be restated also as:

$$|E|e^{i\Phi} = e^{i\frac{\omega_\mathrm{s}+\omega_\mathrm{f}}{2}t}\underbrace{(\mathcal{A}+\mathcal{B})\left[\cos\left(\frac{\omega_\mathrm{f}-\omega_\mathrm{s}}{2}t\right) + i\frac{\mathcal{B}-\mathcal{A}}{\mathcal{A}+\mathcal{B}}\sin\left(\frac{\omega_\mathrm{f}-\omega_\mathrm{s}}{2}t\right)\right]}_{(\mathcal{A}+\mathcal{B})\cos\left(\frac{\Delta\omega}{2}t\right)+i(\mathcal{B}-\mathcal{A})\sin\left(\frac{\Delta\omega}{2}t\right) = \mathcal{A}e^{-i\frac{\Delta\omega}{2}t}+\mathcal{B}e^{i\frac{\Delta\omega}{2}t}}. \tag{C.2.2}$$

We can obtain the pericentre phase Φ applying the logarithm on both sides and taking only the imaginary part:

$$\Phi = \arg(|E|e^{i\Phi}) = \mathrm{Im}\left[\log\left(\mathcal{A}e^{i\omega_\mathrm{s}t} + \mathcal{B}e^{i\omega_\mathrm{f}t}\right)\right] \tag{C.2.3}$$

which gives[1]:

$$\Phi = \frac{\omega_\mathrm{s}}{2}t + \frac{\omega_\mathrm{f}}{2}t + \mathrm{Im}\left\{\log\left[\cos\left(\frac{\omega_\mathrm{f}-\omega_\mathrm{s}}{2}t\right) + i\frac{\mathcal{B}-\mathcal{A}}{\mathcal{A}+\mathcal{B}}\sin\left(\frac{\omega_\mathrm{f}-\omega_\mathrm{s}}{2}t\right)\right]\right\}. \tag{C.2.4}$$

Expanding Eq. (C.2.4) to the first order in $\mathcal{A}/\mathcal{B} \to \{0, 1, \infty\}$ one then gets

$$\Phi \approx \begin{cases} \omega_\mathrm{s}t + \overbrace{\mathrm{Im}\left[\log\left(1+\frac{\mathcal{B}}{\mathcal{A}}e^{i(\omega_\mathrm{f}-\omega_\mathrm{s})t}\right)\right]}^{\approx\frac{\mathcal{B}}{\mathcal{A}}\sin(\Delta\omega t)}, & \text{if } \mathcal{B} \ll \mathcal{A} \\ \frac{\omega_\mathrm{s}}{2}t + \frac{\omega_\mathrm{f}}{2}t + \frac{\mathcal{B}^2-\mathcal{A}^2}{4\mathcal{A}\mathcal{B}}\tan\left(\frac{\Delta\omega t}{2}\right), & \text{if } \mathcal{A} \sim \mathcal{B} \\ \omega_\mathrm{f}t + \underbrace{\mathrm{Im}\left[\log\left(1+\frac{\mathcal{A}}{\mathcal{B}}e^{-i(\omega_\mathrm{f}-\omega_\mathrm{s})t}\right)\right]}_{\approx-\frac{\mathcal{A}}{\mathcal{B}}\sin(\Delta\omega t)}, & \text{if } \mathcal{A} \ll \mathcal{B} \end{cases}, \tag{C.2.5}$$

where $\Delta\omega = \omega_\mathrm{f} - \omega_\mathrm{s}$. In the limiting cases $\mathcal{A} \ll \mathcal{B}$ and $\mathcal{B} \ll \mathcal{A}$, Eq. (C.2.5) tells us that the pericentre precesses linearly with the rate of the dominant mode; however, small periodic oscillations with the beat frequency of the two eigen-modes are superimosed to the linear trend.

[1] Obtained exploiting $\log x = 1/2\log x + 1/2\log x$.

Appendix

D Smoothed Particle Hydrodynamics

The acronym SPH (**S**moothed **P**article **H**ydrodynamics) was firstly used in Gingold & Monaghan (1977) and, independently, in Lucy (1977) as a numerical approach for non-axisymmetric problems in astrophysics involving fluid dynamics.

The algorithm discretises the fluid using a set of moving fluid particles storing information about the fluid properties (position, velocity, density, internal energy). Fluid particles update their properties using specific discretised equations at each timestep and are allowed to move under the action of fluid forces (e.g. pressure, viscosity or fluid self-gravity) and external ones (like the gravity of point masses).

Fluid particles perfectly comove with the fluid. This causes the advection terms to vanish, making in fact the description of the fluid properties Lagrangian. The Lagrangian formalism also enables the conservation of linear momentum, energy and other physical invariants, that instead is not always easy to be guaranteed in grid codes.

Furthermore, other "attractive features" of SPH are (Monaghan 2005):

1. SPH allows the exact treatment of pure advection, which for grid methods needs a moving mesh with a specific geometry.
2. Interface problems can be treated easily in SPH using different sets of particles, which is not the case for finite difference methods (e.g. like dust + gas mixtures).
3. Particle methods like SPH are the best way to treat fractures and fragmentations of solids.
4. The fact that resolution can be made dependent from position and time: SPH invests computational power only where particles are present.
5. The inclusion of complex microscopical physics is made quite simple by the close similarity of SPH to molecular dynamics.

Even though SPH has been originally developed to study the fluid dynamics in astrophysical environment, it has gained through the years a lot of popularity also outside the field. Indeed, it finds many applications also in the film and videogame industry, engineering and geophysics. We note that for these characteristics, the SPH algorithm represents a good example of a tool, developed for pure research, that later has been broadly employed for commercial purposes beyond the scope it was originally designed for.

In this Appendix, we rapidly review SPH fundamental concepts and equations. For a deeper discussion we refer to Monaghan (2005) and Price (2012) as thorough reviews on the subject. The details about the implementation of the equations in the code PHANTOM, that we used as an SPH code throughout this thesis, can be found in Price et al. (2018a).

D.1 Computing continous quantities from point particles: interpolation theory

We can "smooth" the value of a quantity $\mathcal{A}$ over a kernel of width h as follows

$$A(\boldsymbol{r}) = \int \mathcal{A}(\boldsymbol{r}')W(\boldsymbol{r}-\boldsymbol{r}',h)d\boldsymbol{r}', \tag{D.1.1}$$

where $W(\boldsymbol{r}-\boldsymbol{r}',h)$ is the so called "Smoothing kernel", which acts as a sort of weight function over a lengthscale h (called smoothing length) and has the dimension of the inverse of a volume. This function must have two main properties:

$$\int W(\boldsymbol{r}-\boldsymbol{r}',h)d\boldsymbol{r}' = 1, \tag{D.1.2}$$

$$\lim_{h\to 0} W(\boldsymbol{r}-\boldsymbol{r}',h) = \delta(\boldsymbol{r}-\boldsymbol{r}'). \tag{D.1.3}$$

The operation of "smoothing" the quantity $\mathcal{A}$ is equivalent to compute a weighted average of $\mathcal{A}(\boldsymbol{r})$ in the surroundings of $\boldsymbol{r}$.

If we do not know the value of $\mathcal{A}$ over a continuum domain of points $\boldsymbol{r}'$, but only its value at a discrete set of points, we can anyway compute $A(\boldsymbol{r})$ discretizing Eq. (D.1.1) as follows

$$A(\boldsymbol{r}) = \int \frac{\mathcal{A}(\boldsymbol{r}')}{\rho(\boldsymbol{r}')}W(\boldsymbol{r}-\boldsymbol{r}',h)\rho(\boldsymbol{r}')d\boldsymbol{r}' = \tag{D.1.4}$$

$$\approx \sum_{b=1}^{N} m_b \frac{A(\boldsymbol{r}_b)}{\rho(\boldsymbol{r}_b)}W(\boldsymbol{r}-\boldsymbol{r}_b,h) \tag{D.1.5}$$

where we used $m_b = \rho(\boldsymbol{r}')d\boldsymbol{r}'$ and it represents the mass of the b-th particle. Using equation (D.1.5) and substituting A with ρ we obtain the following SPH density estimate for the a-th particle:

$$\rho_a = \sum_{b=1}^{N} m_b W(\boldsymbol{r}_a-\boldsymbol{r}_b,h_a). \tag{D.1.6}$$

The smoothing kernel $W(\boldsymbol{r}-\boldsymbol{r}_b,h)$ used in SPH are generally compact support functions. This implies that only a limited number of neighbours contributes to the density estimate at a given location[1]. As a consequence, the choice of h plays a fundamental role in determining the accuracy of the density estimate. Indeed, if we used a constant h, when particles are in a sufficiently sparse configuration, some of them might have no neighbours at all. One possible solution to the issue is to provide a self-consistent estimate of h and ρ, which also guarantees the algorithm to be fully enforce the conservation of linear momentum and energy, and that the number of particles within the kernel is approximately the same for each particle. In 3D SPH, this can be achieved solving simultaneously Eq. (D.1.6) and (see Sec. 2.1.4 of Price et al. 2018a for an implementation of the numerical method)

$$h(\boldsymbol{r}_a)=\eta\left(\frac{m_a}{\rho_a}\right)^{1/3} \tag{D.1.7}$$

where η is an arbitary constant. For compact support kernels, η qualitatively sets the number of neighbours involved in the density estimate for a given particle (see Price 2012 for further details).

Note that the density estimate only relies on the location of particles. As a consequence, there is no need to study the time evolution of the density: as far as we know the position of particles we can always compute the instantaneous density.

D.2 Equations of motion

The motion of the a-th particle is determined by pressure, viscous and gravitational forces, the evolution of its velocity $\boldsymbol{v}_a$ reads

$$\frac{dv_a^i}{dt}=-\sum_b m_b\left[\frac{\Theta_a^{ij}}{\Omega_a\rho_a^2}\nabla_a^j W_{ab}(h_a)+\frac{\Theta_b^{ij}}{\Omega_b\rho_b^2}\nabla_a^j W_{ab}(h_b)\right]+f_{pot}^i, \tag{D.2.1}$$

where W_{ab} is a shortcoming for $W_{ab}=W(\boldsymbol{r}_a-\boldsymbol{r}_b,h_a)$, f_i accounts for the contribution of gravity forces and other external forces, Θ^{ij} is the pressure-viscous stress tensor:

$$\Theta_a^{ij}=-(P_a+q_a^{\mathrm{AV}})\delta^{ij}+\underbrace{\left[\left(\zeta_a-\frac{2}{3}\eta_a\right)\frac{\partial v_a^k}{\partial x_a^k}\delta^{ij}+\eta_a\left(\frac{\partial v_a^i}{\partial x_a^j}+\frac{\partial v_a^j}{\partial x_a^i}\right)\right]}_{\mathrm{Navier-Stokes}}. \tag{D.2.2}$$

The term Ω_a is related to the gradient of the smoothing length, it is given by:

$$\Omega_a=1-\sum_c m_c\frac{\partial W_{ac}(h_a)}{\partial h_a}\frac{\partial h_a}{\partial \rho_a}, \tag{D.2.3}$$

[1]We note that, in order to prevent misleading interpretations of the physical interpretation of the algorithm, one should always keep in mind the first Golden Rule of SPH:

"However, if you want to find a physical interpretation of an SPH equation, it is always best to assume the kernel is a Gaussian." Monaghan (1992)

it accounts for the variation of h with the density and becomes $\Omega = 1$ if a constant h prescription is adopted. Equation (D.2.1) can be obtained using variational principles and the Lagrangian for hydrodynamics introduced by Eckart (1960). The expression of Ω_a follows from the imposition of the constraint $\rho_a h_a^3 - \eta^3 m_a = 0$ in the Lagrangian, that is equivalent to require that the mass entangled by the kernel is constant (Springel & Hernquist 2002)[2].

The term q_a^{AV} in Eq. (D.2.1) is the "artificial viscosity" term, that is required in SPH to properly resolve discontinuities in the velocity field (i.e. shocks), it reads:

$$q_a^{AV} = \begin{cases} \frac{1}{2}\rho_a v_{\mathrm{sig},a} |\boldsymbol{v}_{ab} \cdot \hat{\boldsymbol{r}}_{ab}|, & \boldsymbol{v}_{ab} \cdot \hat{\boldsymbol{r}}_{ab} < 0 \\ 0, & \boldsymbol{v}_{ab} \cdot \hat{\boldsymbol{r}}_{ab} \geq 0 \end{cases}, \tag{D.2.4}$$

where $\hat{\boldsymbol{r}}_{ab} = (\boldsymbol{r}_b - \boldsymbol{r}_a)/|\boldsymbol{r}_b - \boldsymbol{r}_a|$ is the unit vector in the direction joining the particles, $\boldsymbol{v}_{ab} = \boldsymbol{v}_b - \boldsymbol{v}_a$ and v_{sig} represents the signal speed in the fluid, it is given by

$$v_{\mathrm{sig},a} = \alpha_a^{\mathrm{AV}} c_{\mathrm{S},a} + \beta^{\mathrm{AV}} |\boldsymbol{v}_{ab} \cdot \hat{\boldsymbol{r}}_{ab}|. \tag{D.2.5}$$

The term $\alpha_a^{\mathrm{AV}} \in [0, 1]$ is the artificial viscosity parameter and $\beta^{\mathrm{AV}} = 2$ is introduced to avoid particle interpenetration (imagine two fluid fronts shocking: a shock interface between the two must form, particles cannot interpenetrate and freely move on the other side of the shock). We note that artificial viscosity is active only when particles are approaching ($\boldsymbol{v}_{ab} \cdot \hat{\boldsymbol{r}}_{ab} < 0$), while it vanishes when particles are receiding. Since artificial viscosity is a numerical tool introduced ad hoc to resolve shocks, it is generally desirable to keep it as low as possible. For this reason, often a switch that regulates the value of α_a^{AV} is introduced, in order to increase its value in the regions where the fluid is shocking and keep it as low as possible elsewhere (see Morris & Monaghan 1997 and Cullen & Dehnen 2010 for two possible implementations of the switch).

Finally P_a is the fluid pressure at the location of the a-th particle. It can be prescribed through an equation of state $P_a(\rho_a, \boldsymbol{x}_a)$ (where $\boldsymbol{x}_a$ is a spatial coordinate), or it can be selfconsitently obtained evolving the specific internal energy u of the fluid (energy per unit mass)

$$\frac{du_a}{dt} = \frac{P_a}{\Omega_a \rho_a^2} \sum_b m_b (\boldsymbol{v}_a - \boldsymbol{v}_b) \cdot \nabla_a W_{ab}(h_a) + \mathcal{L}, \tag{D.2.6}$$

where $\mathcal{L}$ represents the contribution of all the internal and external sources of heating and cooling. The system of equations is closed in this case prescribing $P(\rho, u)$

$$P_a = (\gamma - 1)\rho_a u_a, \tag{D.2.7}$$

where γ is the adiabatic index.

[2]Indeed, fixed a number of particle N_{part} of mass m_{p} the mass entangled by the kernel is

$$\frac{4}{3}\pi\rho h^3 = m_{\mathrm{p}} N_{\mathrm{part}}$$

The position of the a-th particle is then updated at each step integrating the following equation:

$$\frac{d\boldsymbol{x}_a}{dt} = \boldsymbol{v}_a. \tag{D.2.8}$$

We note however that in this thesis we never evolved the internal energy through Eq. (D.2.6), we always relied on a locally isothermal equation of state.

D.3 Dust evolution: the one fluid algorithm

There are essentially two ways to simulate the coupled evolution of a dust + gas mixture using SPH: two fluid and one fluid approach.

The first consists in the addition of a second fluid, whose evolution is ruled by an indipendent set of equations (where pressure and viscous terms do not appear), which is coupled to the gas equations through a drag term (Laibe & Price 2012a,b). However, the two fluid algorithm suffers an important slowdown when the dust is in the regime of strong coupling with the gas, since the small stopping time constrains the length of the timestep in the simulation. We did not use this algorithm throughout this thesis, its description and implementation can be found in Price et al. (2018a).

Concerning the second, Laibe & Price (2014a) re-formulated the two fluid hydrodynamical formalism showing that the evolution of the a gas + dust mixture could be effectively described as a single fluid composed by a fraction $1-\epsilon$ by gas and a fraction ϵ of dust. This formulation prescribes the equations for the evolution of:

1. The total density of dust and gas $\rho_{\rm tot} = \rho_{\rm g} + \rho_{\rm d}$.

2. The barycentric velocity of the mixture $\boldsymbol{v}_{\rm bar} = (\boldsymbol{v}_{\rm g}\rho_{\rm g} + \boldsymbol{v}_{\rm d}\rho_{\rm d})/\rho_{\rm tot}$ (see also Eq. 2.57).

3. The dust fraction $\epsilon = \rho_{\rm d}/\rho_{\rm tot}$ (so that $\rho_{\rm g} = (1-\epsilon)\rho_{\rm tot}$ and $\rho_{\rm d} = \epsilon\rho_{\rm tot}$).

4. The velocity difference $\Delta\boldsymbol{v}$ between gas and dust.

The most interesting aspect of this formalism is that the case of strong coupling ${\rm St} \ll 1$ do not cause the simulation to slow down, instead it allows to further simplify the equations: $|\Delta\boldsymbol{v}|/|\boldsymbol{v}_{\rm bar}| \ll 1$ and $\Delta\boldsymbol{v}$ is given by the terminal velocity approximation, introduced in Eq. (2.58).

Laibe & Price (2014b) described the SPH equations for the one fluid approach. The one fluid formulation is in principle exact, but the numerical implementation has some problems in reproducing the streaming of dust particles in the weak coupling limit. For this reason in Price & Laibe (2015) the numerical SPH implementation of the terminal velocity approximation is provided, being much less computationally expensive. This implies that the numerical SPH implementation of the one fluid formalism is adequate only for the description of strongly coupled dust grains (${\rm St} \ll 1$, i.e. $a_{\rm dst} \lesssim 1$ mm for typical disc parameters).

Within this approximation, the equations ruling the evolution of the dust gas mixture are exactly the same as those presented in the previous section (substituting $\rho = \rho_{\rm tot}$ with the total density and $\boldsymbol{v} = \boldsymbol{v}_{\rm bar}$ with the baricentric velocity). The fluid moves as if it was composed only by gas, with the only difference that the effective sound speed is slightly reduced by the presence of the dust[3] $\bar{c}_{\rm s}^2 = c_{\rm s}^2/(1+\rho_{\rm d}/\rho_{\rm g})$ since dust grains are not contributing to the pressure (Laibe & Price 2014a). The difference between dust and gas velocities is obtained from the pressure gradient, as prescribed in Eq. (2.58) $\Delta\boldsymbol{v} = t_{\rm s}\nabla p/\rho_{\rm g}$, while the evolution of the dust fraction satisfies the following equation[4]

$$\frac{d\epsilon}{dt} = -\frac{1}{\rho}\nabla\cdot(\epsilon t_{\rm s}\nabla p)\,. \tag{D.3.1}$$

The previous equation is discretised as follows

$$\frac{{\rm d}\epsilon_a}{{\rm d}t} = -\sum_b m_b\left[\frac{\epsilon_a(1-\epsilon_a)t_{{\rm s},a}}{\Omega_a\rho_a}\Delta\boldsymbol{a}_a\cdot\nabla_a W_{ab}(h_a) + \frac{\epsilon_b(1-\epsilon_b)t_{{\rm s},b}}{\Omega_b\rho_b}\Delta\boldsymbol{a}_b\cdot\nabla_a W_{ab}(h_b)\right], \tag{D.3.2}$$

where $\Delta\boldsymbol{a}_i = -\boldsymbol{a}_{\rm g}^i$ represents the difference of the accelerations of dust and gas (i.e. the pressure gradient and viscous terms) so that $\boldsymbol{a}_{\rm g}^i$ reads

$$\boldsymbol{a}_{\rm g}^i = -\frac{1}{1-\epsilon_a}\sum_b m_b\left[\frac{P_a+q_a^{AV}}{\Omega_a\rho_a^2}\nabla_a W_{ab}(h_a) + \frac{P_b+q_b^{AV}}{\Omega_b\rho_b^2}\nabla_a W_{ab}(h_b)\right], \tag{D.3.3}$$

where the factor $1-\epsilon_a$ comes from the fact that this acceleration acts only on the gas contributing by an amount $(1-\epsilon)\boldsymbol{a}$ to the overall acceleration of the dust + gas mixture.

Recently Ballabio et al. (2018) provided a new formulation of the dust fraction in order to avoid that numerical fluctuations lead to $\epsilon > 1$.

D.4 Code Units

The simulations presented in chapter 6 are completely scalable. We have chosen the parameters in order to allow the easy conversion to physical units. Our setup has indeed $M_{\rm tot} = 1$, $a = 1$ and $G = 1$; this gives $t_{\rm bin} = 2\pi$, where $t_{\rm bin} = 2\pi(GM/a^3)^{-1/2}$. Once the physical total mass of the binary $M_{\rm tot}$ and the binary distance a have been fixed, lengths $l_{\rm code}$, time $t_{\rm code}$, velocity $v_{\rm code}$, density $\rho_{\rm code}$, surface density $\Sigma_{\rm code}$ and accretion rate $\dot{M}_{\rm code}$ in code units have the following

[3] We note indeed that

$$\frac{\nabla P}{\rho_{\rm tot}} = \frac{1}{1+\rho_{\rm d}/\rho_{\rm g}}\frac{\nabla P}{\rho_{\rm g}}.$$

[4] Note that using a globally isothermal equation of state $P = c_{\rm s}\rho_{\rm g} = c_{\rm s}(1-\epsilon)\rho_{\rm tot}$ and constant density, Eq. (D.3.1) becomes a diffusion equation

$$\frac{d\epsilon}{dt} = -\nabla\cdot\left(\epsilon t_{\rm s}c_{\rm s}^2\nabla\epsilon\right)$$

rescaling to physical units:

$$l_{\rm phys} = 1.496 \times 10^{13}\, a_{\rm au} \cdot l_{\rm code}\ {\rm cm} = 3.085 \times 10^{18}\, a_{\rm pc} \cdot l_{\rm code}\ {\rm cm}, \tag{D.4.1}$$

$$t_{\rm phys} = \frac{1}{2\pi}\sqrt{\frac{a_{\rm au}^3}{M_{\rm tot,M_\odot}}} \cdot t_{\rm code}\ {\rm yr}, \tag{D.4.2}$$

$$= 1.487 \times 10^3 \sqrt{\frac{a_{\rm pc}^3}{M_{\rm tot,10^8\cdot M_\odot}}} \cdot t_{\rm code}\ {\rm yr}, \tag{D.4.3}$$

$$v_{\rm phys} = 29.78\sqrt{\frac{M_{\rm tot,M_\odot}}{a_{\rm au}}} \cdot v_{\rm code}\ {\rm km\,s^{-1}} \tag{D.4.4}$$

$$= 6.561 \times 10^2 \sqrt{\frac{M_{\rm tot,10^8\cdot M_\odot}}{a_{\rm pc}}} \cdot v_{\rm code}\ {\rm km\,s^{-1}}, \tag{D.4.5}$$

$$\rho_{\rm phys} = 5.941 \times 10^{-6} \frac{M_{\rm tot,M_\odot}}{a_{\rm au}^3} \cdot \rho_{\rm code}\ {\rm g\,cm^{-3}} \tag{D.4.6}$$

$$= 6.435 \times 10^{-14} \frac{M_{\rm tot,10^8\cdot M_\odot}}{a_{\rm pc}^3} \cdot \rho_{\rm code}\ {\rm g\,cm^{-3}}, \tag{D.4.7}$$

$$\Sigma_{\rm phys} = 8.887 \times 10^7 \frac{M_{\rm tot,M_\odot}}{a_{\rm au}^2} \cdot \Sigma_{\rm code}\ {\rm g\,cm^{-2}} \tag{D.4.8}$$

$$= 1.969 \times 10^5 \frac{M_{\rm tot,10^8\cdot M_\odot}}{a_{\rm pc}^2} \cdot \Sigma_{\rm code}\ {\rm g\,cm^{-2}}, \tag{D.4.9}$$

$$\dot{M}_{\rm phys} = 2\pi \cdot \sqrt{\frac{M_{\rm tot,M_\odot}^3}{a_{\rm au}^3}} \cdot \dot{M}_{\rm code}\ {\rm M_\odot\,yr^{-1}}, \tag{D.4.10}$$

$$= 6.724 \times 10^4 \sqrt{\frac{M_{\rm tot,10^8\cdot M_\odot}^3}{a_{\rm pc}^3}} \cdot \dot{M}_{\rm code}\ {\rm M_\odot\,yr^{-1}}, \tag{D.4.11}$$

where $l_{\rm phys}$ is the physical length, $t_{\rm phys}$ is the physical time, $v_{\rm phys}$ is the physical speed, $\rho_{\rm phys}$ the physical density, $\Sigma_{\rm phys}$ the physical surface densisty of the disc and $M_{\rm phys}$ the physical accretion rate; $a_{\rm au}$ and $a_{\rm pc}$ are the binary separation in au and pc units, respectively; $M_{\rm tot,M_\odot}$ and $M_{\rm tot,10^8\cdot M_\odot}$ are the total mass of the binary in solar masses, $\rm M_\odot$, and $10^8 \rm M_\odot$ units, respectively.

Bibliography

ALMA Partnership et al., 2015, ApJL, 808, L3

Abbott B. P., et al., 2016a, Physical Review X, 6, 041015

Abbott B. P., et al., 2016b, Physical Review Letters, 116, 061102

Abbott B. P., et al., 2016c, Physical Review Letters, 116, 241102

Abbott B. P., et al., 2016d, Physical Review Letters, 116, 241103

Abbott B. P., et al., 2016e, ApJ, 818, L22

Abbott B. P., et al., 2017a, Physical Review Letters, 118, 221101

Abbott B. P., et al., 2017b, Physical Review Letters, 119, 141101

Abbott B. P., et al., 2017c, Physical Review Letters, 119, 161101

Abbott B. P., et al., 2017d, ApJ, 851, L35

Abramowitz M., Stegun I. A., 1968, Handbook of mathematical functions with formulas, graphs and mathematical tables

Alexander R. D., Clarke C. J., Pringle J. E., 2006, MNRAS, 369, 229

Alexander R., Pascucci I., Andrews S., Armitage P., Cieza L., 2014, Protostars and Planets VI, pp 475–496

Aly H., Dehnen W., Nixon C., King A., 2015, MNRAS, 449, 65

André P., Ward-Thompson D., Barsony M., 2000, Protostars and Planets IV, p. 59

Andrews S. M., Williams J. P., 2007, ApJ, 659, 705

Andrews S. M., Wilner D. J., Espaillat C., Hughes A. M., Dullemond C. P., McClure M. K., Qi C., Brown J. M., 2011, ApJ, 732, 42

Andrews S. M., et al., 2016, ApJ, 820, L40

Arena S. E., Gonzalez J.-F., 2013, MNRAS, 433, 98

Armitage P. J., 2007, ApJ, 665, 1381

Armitage P. J., 2010, Astrophysics of Planet Formation

Armitage P. J., 2018, preprint, (`arXiv:1803.10526`)

Armitage P. J., Natarajan P., 2002, ApJ, 567, L9
Armitage P. J., Natarajan P., 2005, ApJ, 634, 921
Artymowicz P., 1993a, ApJ, 419, 155
Artymowicz P., 1993b, ApJ, 419, 166
Artymowicz P., Lubow S. H., 1994, ApJ, 421, 651
Artymowicz P., Lubow S. H., 1996, ApJ, 467, L77
Artymowicz P., Clarke C. J., Lubow S. H., Pringle J. E., 1991, ApJ, 370, L35
Arzoumanian Z., et al., 2016, ApJ, 821, 13
Ataiee S., et al., 2013, A&A, 553, L3
Avenhaus H., Quanz S. P., Schmid H. M., Meyer M. R., Garufi A., Wolf S., Dominik C., 2014, ApJ, 781, 87
Avenhaus H., et al., 2017, AJ, 154, 33
Bañados E., et al., 2018, ApJ, 856, L25
Bae J., Pinilla P., Birnstiel T., 2018, ApJ, 864, L26
Balbus S. A., Hawley J. F., 1998, Reviews of Modern Physics, 70, 1
Balbus S. A., Papaloizou J. C. B., 1999, ApJ, 521, 650
Ballabio G., Dipierro G., Veronesi B., Lodato G., Hutchison M., Laibe G., Price D. J., 2018, MNRAS, 477, 2766
Bardeen J. M., 1970, Nature, 226, 64
Bardeen J. M., Petterson J. A., 1975, ApJ, 195, L65
Barge P., Sommeria J., 1995, A&A, 295, L1
Barnes J. E., 2002, MNRAS, 333, 481
Bartos I., Kocsis B., Haiman Z., Márka S., 2017, ApJ, 835, 165
Baruteau C., Masset F., 2008, ApJ, 672, 1054
Baruteau C., Masset F., 2013, in Souchay J., Mathis S., Tokieda T., eds, Lecture Notes in Physics, Berlin Springer Verlag Vol. 861, Lecture Notes in Physics, Berlin Springer Verlag. p. 201 (arXiv:1203.3294), doi:10.1007/978-3-642-32961-6˙6
Baruteau C., Fromang S., Nelson R. P., Masset F., 2011, A&A, 533, A84
Baruteau C., Ramirez-Ruiz E., Masset F., 2012, MNRAS, 423, L65
Baruteau C., et al., 2014, Protostars and Planets VI, pp 667–689
Bate M. R., Bonnell I. A., 1997, MNRAS, 285, 33
Bate M. R., Bonnell I. A., Price N. M., 1995, MNRAS, 277, 362
Beck T. L., Bary J. S., Dutrey A., Piétu V., Guilloteau S., Lubow S. H., Simon M., 2012, ApJ, 754, 72
Begelman M. C., 1979, MNRAS, 187, 237
Begelman M. C., 2010, MNRAS, 402, 673
Begelman M. C., Volonteri M., 2017, MNRAS, 464, 1102
Begelman M. C., Blandford R. D., Rees M. J., 1980, Nature, 287, 307
Begelman M. C., Volonteri M., Rees M. J., 2006, MNRAS, 370, 289
Belczynski K., Holz D. E., Bulik T., O'Shaughnessy R., 2016, Nature, 534, 512
Belczynski K., et al., 2017, preprint, (arXiv:1706.07053)

Benacquista M. J., Downing J. M. B., 2013, Living Reviews in Relativity, 16, 4
Benisty M., et al., 2015, A&A, 578, L6
Benisty M., et al., 2017, A&A, 597, A42
Benítez-Llambay P., Masset F. S., 2016, ApJS, 223, 11
Benítez-Llambay P., Masset F., Koenigsberger G., Szulágyi J., 2015, Nature, 520, 63
Biller B., et al., 2012, ApJ, 753, L38
Biller B. A., et al., 2014, ApJ, 792, L22
Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition. Princeton University Press
Birnstiel T., Andrews S. M., Ercolano B., 2012, A&A, 544, A79
Birnstiel T., Dullemond C. P., Pinilla P., 2013, A&A, 550, L8
Bitsch B., Kley W., 2010, A&A, 523, A30
Bitsch B., Crida A., Morbidelli A., Kley W., Dobbs-Dixon I., 2013a, A&A, 549, A124
Bitsch B., Crida A., Libert A.-S., Lega E., 2013b, A&A, 555, A124
Blecha L., et al., 2016, MNRAS, 456, 961
Blum J., Wurm G., 2008, ARA&A, 46, 21
Boehler Y., Weaver E., Isella A., Ricci L., Grady C., Carpenter J., Perez L., 2017, ApJ, 840, 60
Bogdanović T., 2015, in Sopuerta C. F., ed., Astrophysics and Space Science Proceedings Vol. 40, Gravitational Wave Astrophysics. p. 103 (arXiv:1406.5193), doi:10.1007/978-3-319-10488-1˙9
Bogdanović T., Reynolds C. S., Miller M. C., 2007, ApJ, 661, L147
Bonetti M., Sesana A., Barausse E., Haardt F., 2018a, MNRAS, 477, 2599
Bonetti M., Haardt F., Sesana A., Barausse E., 2018b, MNRAS, 477, 3910
Bromm V., Loeb A., 2003, ApJ, 596, 34
Brown J. M., Blake G. A., Qi C., Dullemond C. P., Wilner D. J., Williams J. P., 2009, ApJ, 704, 496
Bruderer S., van der Marel N., van Dishoeck E. F., van Kempen T. A., 2014a, A&A, 562, A26
Bruderer S., van der Marel N., van Dishoeck E. F., van Kempen T. A., 2014b, A&A, 562, A26
Bryden G., Chen X., Lin D. N. C., Nelson R. P., Papaloizou J. C. B., 1999, ApJ, 514, 344
Butler R. P., et al., 2006, ApJ, 646, 505
Calvet N., et al., 2005, ApJ, 630, L185
Campanelli M., Lousto C., Zlochower Y., Merritt D., 2007, ApJ, 659, L5
Canovas H., et al., 2016, MNRAS, 458, L29
Casassus S., 2016, PASA, 33, e013
Casassus S., et al., 2013, Nature, 493, 191
Casassus S., et al., 2015a, ApJ, 811, 92
Casassus S., et al., 2015b, ApJ, 812, 126

Casassus S., et al., 2018, MNRAS, 477, 5104

Casoli J., Masset F. S., 2009, ApJ, 703, 845

Celoria M., Oliveri R., Sesana A., Mapelli M., 2018, preprint, (`arXiv:1807.11489`)

Cerioli A., Lodato G., Price D. J., 2016, MNRAS, 457, 939

Chandrasekhar S., 1943, ApJ, 97, 255

Chandrasekhar S., 1964, ApJ, 140, 417

Chang P., Strubbe L. E., Menou K., Quataert E., 2010, MNRAS, 407, 2007

Charisi M., Bartos I., Haiman Z., Price-Whelan A. M., Graham M. J., Bellm E. C., Laher R. R., Márka S., 2016, MNRAS, 463, 2145

Chiang E. I., Goldreich P., 1997, ApJ, 490, 368

Cielo S., Bieri R., Volonteri M., Wagner A. Y., Dubois Y., 2018, MNRAS, 477, 1336

Clarke C., Carswell B., 2014, Principles of Astrophysical Fluid Dynamics

Clarke C. J., Gendrin A., Sotomayor M., 2001, MNRAS, 328, 485

Collin-Souffrin S., Dumont A. M., 1990, A&A, 229, 292

Colpi M., Sesana A., 2017, Gravitational Wave Sources in the Era of Multi-Band Gravitational Wave Astronomy. World Scientific Publishing Co, pp 43–140, doi:10.1142/9789813141766˙0002

Comerford J. M., Pooley D., Barrows R. S., Greene J. E., Zakamska N. L., Madejski G. M., Cooper M. C., 2015, ApJ, 806, 219

Connaughton V., et al., 2016, ApJ, 826, L6

Connaughton V., et al., 2018, ApJ, 853, L9

Cossins P., Lodato G., Clarke C. J., 2009, MNRAS, 393, 1157

Cresswell P., Dirksen G., Kley W., Nelson R. P., 2007, A&A, 473, 329

Crida A., Morbidelli A., Masset F., 2006, Icarus, 181, 587

Cuadra J., Armitage P. J., Alexander R. D., Begelman M. C., 2009, MNRAS, 393, 1423

Cullen L., Dehnen W., 2010, MNRAS, 408, 669

Currie T., Cloutier R., Brittain S., Grady C., Burrows A., Muto T., Kenyon S. J., Kuchner M. J., 2015, ApJ, 814, L27

D'Alessio P., Calvet N., Hartmann L., Lizano S., Cantó J., 1999, ApJ, 527, 893

D'Alessio P., Calvet N., Hartmann L., Franco-Hernández R., Servín H., 2006, ApJ, 638, 314

D'Angelo G., Lubow S. H., 2010, ApJ, 724, 730

D'Angelo G., Lubow S. H., Bate M. R., 2006, ApJ, 652, 1698

D'Orazio D. J., Loeb A., 2018a, Phys. Rev. D, 97, 083008

D'Orazio D. J., Loeb A., 2018b, ApJ, 863, 185

D'Orazio D. J., Haiman Z., MacFadyen A., 2013, MNRAS, 436, 2997

D'Orazio D. J., Haiman Z., Schiminovich D., 2015, Nature, 525, 351

D'Orazio D. J., et al., 2016, MNRAS, 459, 2379

Davies M. B., Miller M. C., Bellovary J. M., 2011, ApJ, 740, L42

Devecchi B., Volonteri M., 2009, ApJ, 694, 302

Dipierro G., et al., 2018a, MNRAS, 475, 5296
Dipierro G., Laibe G., Alexander R., Hutchison M., 2018b, MNRAS, 479, 4187
Dodson-Robinson S. E., Salyk C., 2011, ApJ, 738, 131
Dominik C., Tielens A. G. G. M., 1997, ApJ, 480, 647
Dong R., Dawson R., 2016, ApJ, 825, 77
Dong R., Zhu Z., Whitney B., 2015, ApJ, 809, 93
Dong R., et al., 2018, ApJ, 860, 124
Dotti M., Colpi M., Haardt F., 2006, MNRAS, 367, 103
Dotti M., Colpi M., Haardt F., Mayer L., 2007, MNRAS, 379, 956
Dotti M., Ruszkowski M., Paredi L., Colpi M., Volonteri M., Haardt F., 2009, MNRAS, 396, 1640
Dotti M., Volonteri M., Perego A., Colpi M., Ruszkowski M., Haardt F., 2010, MNRAS, 402, 682
Dotti M., Sesana A., Decarli R., 2012, Advances in Astronomy, 2012, 940568
Draine B. T., 2006, ApJ, 636, 1114
Drake A. J., et al., 2009, ApJ, 696, 870
Dubrulle B., Morfill G., Sterzik M., 1995, Icarus, 114, 237
Duffell P. C., 2015, ApJ, 807, L11
Duffell P. C., Chiang E., 2015, ApJ, 812, 94
Duffell P. C., Dong R., 2015, ApJ, 802, 42
Duffell P. C., MacFadyen A. I., 2013, ApJ, 769, 41
Duffell P. C., Haiman Z., MacFadyen A. I., D'Orazio D. J., Farris B. D., 2014, ApJ, 792, L10
Dullemond C. P., 2012, RADMC-3D: A multi-purpose radiative transfer tool, Astrophysics Source Code Library (ascl:1202.015)
Dullemond C. P., Dominik C., 2005, A&A, 434, 971
Dunhill A. C., Alexander R. D., Armitage P. J., 2013, MNRAS, 428, 3072
Dunhill A. C., Alexander R. D., Nixon C. J., King A. R., 2014, MNRAS, 445, 2285
Dunhill A. C., Cuadra J., Dougados C., 2015, MNRAS, 448, 3545
Durisen R. H., Boss A. P., Mayer L., Nelson A. F., Quinn T., Rice W. K. M., 2007, Protostars and Planets V, pp 607–622
Dürmann C., Kley W., 2015, A&A, 574, A52
Dutrey A., et al., 2008, A&A, 490, L15
Eckart C., 1960, Physics of Fluids, 3, 421
Edgar R. G., 2007, ApJ, 663, 1325
Eisner J. A., Hillenbrand L. A., White R. J., Akeson R. L., Sargent A. I., 2005, ApJ, 623, 952
Eisner J. A., Monnier J. D., Tuthill P., Lacour S., 2009, ApJ, 698, L169
Epstein P. S., 1924, Phys. Rev., 23, 710
Ercolano B., Pascucci I., 2017, Royal Society Open Science, 4, 170114
Ercolano B., Clarke C. J., Hall A. C., 2011, MNRAS, 410, 671
Escala A., Larson R. B., Coppi P. S., Mardones D., 2005, ApJ, 630, 152

Espaillat C., Calvet N., D'Alessio P., Hernández J., Qi C., Hartmann L., Furlan E., Watson D. M., 2007, ApJ, 670, L135

Espaillat C., et al., 2010, ApJ, 717, 441

Espaillat C., et al., 2014, Protostars and Planets VI, pp 497–520

Farr W. M., Stevenson S., Miller M. C., Mandel I., Farr B., Vecchio A., 2017, Nature, 548, 426

Farris B. D., Duffell P., MacFadyen A. I., Haiman Z., 2014, ApJ, 783, 134

Farris B. D., Duffell P., MacFadyen A. I., Haiman Z., 2015a, MNRAS, 446, L36

Farris B. D., Duffell P., MacFadyen A. I., Haiman Z., 2015b, MNRAS, 447, L80

Fedele D., et al., 2017, A&A, 600, A72

Fedele D., et al., 2018, A&A, 610, A24

Fendyke S. M., Nelson R. P., 2014, MNRAS, 437, 96

Fishbach M., Holz D. E., 2017, ApJ, 851, L25

Fitchett M. J., 1983, MNRAS, 203, 1049

Flaherty K. M., Hughes A. M., Rosenfeld K. A., Andrews S. M., Chiang E., Simon J. B., Kerzner S., Wilner D. J., 2015, ApJ, 813, 99

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

Flebbe O., Muenzel S., Herold H., Riffert H., Ruder H., 1994, ApJ, 431, 754

Follette K. B., et al., 2013, ApJ, 767, 10

Follette K. B., et al., 2017, AJ, 153, 264

Fontecilla C., Chen X., Cuadra J., 2017, MNRAS, 468, L50

Ford E. B., Rasio F. A., 2008, ApJ, 686, 621

Forrest W. J., et al., 2004, ApJS, 154, 443

Fragos T., McClintock J. E., 2015, ApJ, 800, 17

Frank J., King A., Raine D. J., 2002, Accretion Power in Astrophysics: Third Edition

Fromang S., Terquem C., Nelson R. P., 2005, MNRAS, 363, 943

Fu W., Li H., Lubow S., Li S., 2014a, ApJ, 788, L41

Fu W., Li H., Lubow S., Li S., Liang E., 2014b, ApJ, 795, L39

Fu W., Li H., Lubow S., Li S., Liang E., 2014c, ApJ, 795, L39

Fuente A., et al., 2017, ApJ, 846, L3

Fung J., Artymowicz P., Wu Y., 2015, ApJ, 811, 101

Gammie C. F., 2001, ApJ, 553, 174

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

Garufi A., Quanz S. P., Schmid H. M., Avenhaus H., Buenzli E., Wolf S., 2014, A&A, 568, A40

Gerosa D., 2018, in Journal of Physics Conference Series. p. 012014 (`arXiv:1711.10038`), doi:10.1088/1742-6596/957/1/012014

Gerosa D., Berti E., 2017, Phys. Rev. D, 95, 124046

Gerosa D., Veronesi B., Lodato G., Rosotti G., 2015, MNRAS, 451, 3941

Giacobbo N., Mapelli M., Spera M., 2018, MNRAS, 474, 2959

Giacomazzo B., Perna R., Rezzolla L., Troja E., Lazzati D., 2013, ApJ, 762, L18

Gingold R. A., Monaghan J. J., 1977, MNRAS, 181, 375
Goicovic F. G., Cuadra J., Sesana A., Stasyszyn F., Amaro-Seoane P., Tanaka T. L., 2016, MNRAS, 455, 1989
Gold R., Paschalidis V., Etienne Z. B., Shapiro S. L., Pfeiffer H. P., 2014, Phys. Rev. D, 89, 064060
Goldreich P., Sari R., 2003, ApJ, 585, 1024
Goldreich P., Tremaine S., 1979, ApJ, 233, 857
Goldreich P., Tremaine S., 1980, ApJ, 241, 425
Goldreich P., Tremaine S., 1981, ApJ, 243, 1062
Goldreich P., Tremaine S., 1982, ARA&A, 20, 249
Goldreich P., Ward W. R., 1973, ApJ, 183, 1051
Goldstein A., et al., 2017, ApJ, 848, L14
Goodman J., 2003, MNRAS, 339, 937
Goodman J., Rafikov R. R., 2001, ApJ, 552, 793
Gorti U., Dullemond C. P., Hollenbach D., 2009, ApJ, 705, 1237
Grady C. A., et al., 2013, ApJ, 762, 48
Graham A. W., 2008, PASA, 25, 167
Graham M. J., et al., 2015, MNRAS, 453, 1562
Gravity Collaboration et al., 2018, A&A, 615, L15
Greif T. H., Springel V., White S. D. M., Glover S. C. O., Clark P. C., Smith R. J., Klessen R. S., Bromm V., 2011, ApJ, 737, 75
Greiner J., Burgess J. M., Savchenko V., Yu H.-F., 2016, ApJ, 827, L38
Günther R., Kley W., 2002, A&A, 387, 550
Haemmerlé L., Woods T. E., Klessen R. S., Heger A., Whalen D. J., 2018, MNRAS, 474, 2757
Hahn J., 2009, The Dynamics of Planetary Systems and Astrophysical Disks. WILEY-VCH
Haiman Z., Rees M. J., Loeb A., 1997, ApJ, 476, 458
Haiman Z., Kocsis B., Menou K., Lippai Z., Frei Z., 2009a, Classical and Quantum Gravity, 26, 094032
Haiman Z., Kocsis B., Menou K., 2009b, ApJ, 700, 1952
Hammer M., Kratter K. M., Lin M.-K., 2017, MNRAS, 466, 3533
Hammer M., Pinilla P., Kratter K. M., Lin M.-K., 2018, MNRAS,
Hanawa T., Ochi Y., Ando K., 2010, ApJ, 708, 485
Hayasaki K., Mineshige S., Sudou H., 2007, PASJ, 59, 427
Hayashi C., 1981, Progress of Theoretical Physics Supplement, 70, 35
Heger A., Fryer C. L., Woosley S. E., Langer N., Hartmann D. H., 2003, ApJ, 591, 288
Hendler N. P., et al., 2018, MNRAS, 475, L62
Hosokawa T., Yorke H. W., Inayoshi K., Omukai K., Yoshida N., 2013, ApJ, 778, 178
Hunter J. D., 2007, Computing in Science and Engineering, 9, 90
Ilee J. D., et al., 2011, MNRAS, 417, 2950

Inayoshi K., Haiman Z., 2016, ApJ, 828, 110
Inayoshi K., Haiman Z., Ostriker J. P., 2016, MNRAS, 459, 3738
Ireland M. J., Kraus A. L., 2008, ApJ, 678, L59
Isella A., Pérez L. M., Carpenter J. M., Ricci L., Andrews S., Rosenfeld K., 2013, ApJ, 775, 30
Ivanov P. B., Papaloizou J. C. B., Polnarev A. G., 1999, MNRAS, 307, 79
Jacquet E., Balbus S., Latter H., 2011, MNRAS, 415, 3591
Jennings J., Ercolano B., Rosotti G. P., 2018, MNRAS, 477, 4131
Jiang Y.-F., Stone J., Davis S. W., 2017, preprint, (arXiv:1709.02845)
Johansen A., Andersen A. C., Brandenburg A., 2004, A&A, 417, 361
Johns-Krull C. M., et al., 2016, ApJ, 826, 206
Jurić M., Tremaine S., 2008, ApJ, 686, 603
Kanagawa K. D., Ueda T., Muto T., Okuzumi S., 2017, ApJ, 844, 142
Kanagawa K. D., Tanaka H., Szuszkiewicz E., 2018, preprint, (arXiv:1805.11101)
Kastner J. H., et al., 2018, preprint, (arXiv:1806.10553)
Katz H., Sijacki D., Haehnelt M. G., 2015, MNRAS, 451, 2352
Kelley L. Z., Haiman Z., Sesana A., Hernquist L., 2018, preprint, (arXiv:1809.02138)
Kenyon S. J., Hartmann L., 1987, ApJ, 323, 714
Keppler M., et al., 2018, preprint, (arXiv:1806.11568)
Kerr R. P., 1963, Physical Review Letters, 11, 237
Khan F. M., Fiacconi D., Mayer L., Berczik P., Just A., 2016, ApJ, 828, 73
Kharb P., Lal D. V., Merritt D., 2017, Nature Astronomy, 1, 727
Kidder L. E., 1995, Phys. Rev. D, 52, 821
Kimura S. S., Takahashi S. Z., Toma K., 2017, MNRAS, 465, 4406
King A., 2016, MNRAS, 456, L109
King A. R., Pringle J. E., 2006, MNRAS, 373, L90
King A. R., Pringle J. E., Livio M., 2007, MNRAS, 376, 1740
Kinugawa T., Inayoshi K., Hotokezaka K., Nakauchi D., Nakamura T., 2014, MNRAS, 442, 2963
Kley W., Dirksen G., 2006, A&A, 447, 369
Kley W., Nelson R. P., 2012, ARA&A, 50, 211
Kolmogorov A., 1941, Akademiia Nauk SSSR Doklady, 30, 301
Komossa S., 2012, Advances in Astronomy, 2012, 364973
Kormendy J., Ho L. C., 2013, ARA&A, 51, 511
Korycansky D. G., Papaloizou J. C. B., 1996, ApJS, 105, 181
Kowalska-Leszczynska I., Regimbau T., Bulik T., Dominik M., Belczynski K., 2015, A&A, 574, A58
Kratter K., Lodato G., 2016, ARA&A, 54, 271
Kraus A. L., Ireland M. J., 2012, ApJ, 745, 5
Krause M. G. H., et al., 2018, preprint, (arXiv:1809.04050)

Kruskal M. D., 1960, Physical Review, 119, 1743
Lacour S., et al., 2016, A&A, 590, A90
Laibe G., Price D. J., 2012a, MNRAS, 420, 2345
Laibe G., Price D. J., 2012b, MNRAS, 420, 2365
Laibe G., Price D. J., 2014a, MNRAS, 440, 2136
Laibe G., Price D. J., 2014b, MNRAS, 440, 2147
Laibe G., Gonzalez J.-F., Maddison S. T., 2012, A&A, 537, A61
Lamberts A., et al., 2018, preprint, (arXiv:1801.03099)
Latif M. A., Schleicher D. R. G., Schmidt W., Niemeyer J. C., 2013, MNRAS, 436, 2989
Lega E., Crida A., Bitsch B., Morbidelli A., 2014, MNRAS, 440, 683
Lentati L., et al., 2015, MNRAS, 453, 2576
Ligi R., et al., 2018, MNRAS, 473, 1774
Lin D. N. C., Papaloizou J., 1979, MNRAS, 186, 799
Lin D. N. C., Papaloizou J., 1986, ApJ, 309, 846
Lin D. N. C., Papaloizou J. C. B., 1993, in Levy E. H., Lunine J. I., eds, Protostars and Planets III. pp 749–835
Lippai Z., Frei Z., Haiman Z., 2008, ApJ, 676, L5
Liu Y. T., Shapiro S. L., 2010, Phys. Rev. D, 82, 123011
Liu Y., et al., 2018, preprint, (arXiv:1811.04074)
Lodato G., 2008, New A Rev., 52, 21
Lodato G., 2012, Advances in Astronomy, 2012, 846875
Lodato G., Gerosa D., 2013, MNRAS, 429, L30
Lodato G., Natarajan P., 2006, MNRAS, 371, 1813
Lodato G., Price D. J., 2010, MNRAS, 405, 1212
Lodato G., Nayakshin S., King A. R., Pringle J. E., 2009, MNRAS, 398, 1392
Lodato G., Scardoni C. E., Manara C. F., Testi L., 2017, MNRAS, 472, 4700
Loeb A., 2016, ApJ, 819, L21
Long F., et al., 2018, preprint, (arXiv:1810.06044)
Lousto C. O., Zlochower Y., 2008, Phys. Rev. D, 77, 044028
Lousto C. O., Zlochower Y., 2011, Physical Review Letters, 107, 231102
Lovelace R. V. E., Li H., Colgate S. A., Nelson A. F., 1999, ApJ, 513, 805
Lubow S. H., 1991a, ApJ, 381, 259
Lubow S. H., 1991b, ApJ, 381, 268
Lubow S. H., Artymowicz P., 1997, in Wickramasinghe D. T., Bicknell G. V., Ferrario L., eds, Astronomical Society of the Pacific Conference Series Vol. 121, IAU Colloq. 163: Accretion Phenomena and Related Outflows. p. 505
Lubow S. H., D'Angelo G., 2006, ApJ, 641, 526
Lubow S. H., Martin R. G., Nixon C., 2015, ApJ, 800, 96
Lucy L. B., 1977, AJ, 82, 1013
Luhman K. L., Allen P. R., Espaillat C., Hartmann L., Calvet N., 2010, ApJS, 186, 111

Lynden-Bell D., 1969, Nature, 223, 690
Lynden-Bell D., Kalnajs A. J., 1972, MNRAS, 157, 1
Lyra W., Lin M.-K., 2013, ApJ, 775, 17
Lyra W., Johansen A., Klahr H., Piskunov N., 2009, A&A, 493, 1125
MacFadyen A. I., Milosavljević M., 2008, ApJ, 672, 83
Madau P., Rees M. J., 2001, ApJ, 551, L27
Makino J., Funato Y., 2004, ApJ, 602, 93
Manara C. F., et al., 2016, A&A, 591, L3
Marshall M. D., Avara M. J., McKinney J. C., 2018, MNRAS, 478, 1837
Martin R. G., Nixon C., Xie F.-G., King A., 2018, MNRAS, 480, 4732
Masset F. S., 2008, in Goupil M.-J., Zahn J.-P., eds, EAS Publications Series Vol. 29, EAS Publications Series. pp 165–244, doi:10.1051/eas:0829006
Mayer L., 2013, Classical and Quantum Gravity, 30, 244008
Mayer L., Bonoli S., 2019, Reports on Progress in Physics, 82, 016901
Mayer L., Kazantzidis S., Madau P., Colpi M., Quinn T., Wadsley J., 2007, Science, 316, 1874
McKee C. F., Ostriker E. C., 2007, ARA&A, 45, 565
Mendigutía I., et al., 2017, A&A, 608, A104
Menou K., Perna R., Hernquist L., 2001, ApJ, 559, 1032
Merritt D., Poon M. Y., 2004, ApJ, 606, 788
Meyer-Vernet N., Sicardy B., 1987, Icarus, 69, 157
Miller M. C., Krolik J. H., 2013, ApJ, 774, 43
Milosavljević M., Merritt D., 2001, ApJ, 563, 34
Milosavljević M., Merritt D., 2003, in Centrella J. M., ed., American Institute of Physics Conference Series Vol. 686, The Astrophysics of Gravitational Wave Sources. pp 201–210 (`arXiv:astro-ph/0212270`), doi:10.1063/1.1629432
Milosavljević M., Phinney E. S., 2005, ApJ, 622, L93
Min M., Rab C., Woitke P., Dominik C., Ménard F., 2016, A&A, 585, A13
Miranda R., Muñoz D. J., Lai D., 2017, MNRAS, 466, 1170
Miyoshi M., Moran J., Herrnstein J., Greenhill L., Nakai N., Diamond P., Inoue M., 1995, Nature, 373, 127
Monaghan J. J., 1992, ARA&A, 30, 543
Monaghan J. J., 2005, Reports on Progress in Physics, 68, 1703
Moriwaki K., Nakagawa Y., 2004, ApJ, 609, 1065
Morris J. P., Monaghan J. J., 1997, Journal of Computational Physics, 136, 41
Mortlock D. J., et al., 2011, Nature, 474, 616
Muñoz D. J., Miranda R., Lai D., 2018, preprint, (`arXiv:1810.04676`)
Müller T. W. A., Kley W., 2013, A&A, 560, A40
Müller A., et al., 2018, A&A, 617, L2
Murase K., Kashiyama K., Mészáros P., Shoemaker I., Senno N., 2016, ApJ, 822, L9
Murray C. D., Dermott S. F., 1999, Solar system dynamics

Mustill A. J., Davies M. B., Johansen A., 2017, MNRAS, 468, 3000
Muto T., et al., 2012, ApJ, 748, L22
Muto T., et al., 2015, PASJ, 67, 122
Nagel E., D'Alessio P., Calvet N., Espaillat C., Sargent B., Hernández J., Forrest W. J., 2010, ApJ, 708, 38
Najita J. R., Bergin E. A., 2018, preprint, (arXiv:1808.05618)
Najita J. R., Crockett N., Carr J. S., 2008, ApJ, 687, 1168
Najita J. R., et al., 2009, ApJ, 697, 957
Naoz S., 2016, ARA&A, 54, 441
Narayan R., 2005, New Journal of Physics, 7, 199
Narayan R., Igumenshchev I. V., Abramowicz M. A., 2003, PASJ, 55, L69
Natarajan P., Pringle J. E., 1998, ApJ, 506, L97
Nixon C., Lubow S. H., 2015, MNRAS, 448, 3472
Nixon C. J., Cossins P. J., King A. R., Pringle J. E., 2011, Monthly Notices of the Royal Astronomical Society, 412, 1591
Nixon C., King A., Price D., 2013, MNRAS, 434, 1946
O'Shaughnessy R., Gerosa D., Wysocki D., 2017, Physical Review Letters, 119, 011101
Ochi Y., Sugimoto K., Hanawa T., 2005, ApJ, 623, 922
Ogilvie G. I., 2001, MNRAS, 325, 231
Ogilvie G. I., Lubow S. H., 2003, ApJ, 587, 398
Oh S. P., Haiman Z., 2002, ApJ, 569, 558
Omukai K., Schneider R., Haiman Z., 2008, ApJ, 686, 801
Oppenheimer J. R., Volkoff G. M., 1939, Physical Review, 55, 374
Ormel C. W., 2017, in Pessah M., Gressel O., eds, Astrophysics and Space Science Library Vol. 445, Astrophysics and Space Science Library. p. 197, doi:10.1007/978-3-319-60609-5˙7
Osorio M., et al., 2014, ApJ, 791, L36
Owen J. E., 2016, PASA, 33, e005
Owen J. E., Clarke C. J., 2012, MNRAS, 426, L96
Owen J. E., Kollmeier J. A., 2017, MNRAS, 467, 3379
Owen J. E., Ercolano B., Clarke C. J., 2011, MNRAS, 412, 13
Özel F., Freire P., 2016, ARA&A, 54, 401
Paardekooper S.-J., Johansen A., 2018, Space Science Reviews, 214, 38
Paardekooper S.-J., Mellema G., 2006, A&A, 459, L17
Paardekooper S.-J., Papaloizou J. C. B., 2008, A&A, 485, 877
Paardekooper S.-J., Papaloizou J. C. B., 2009a, MNRAS, 394, 2283
Paardekooper S.-J., Papaloizou J. C. B., 2009b, MNRAS, 394, 2297
Paardekooper S.-J., Baruteau C., Crida A., Kley W., 2010, MNRAS, 401, 1950
Paardekooper S.-J., Baruteau C., Kley W., 2011, MNRAS, 410, 293
Papaloizou J. C. B., 2002, A&A, 388, 615
Papaloizou J. C. B., Larwood J. D., 2000, MNRAS, 315, 823

Papaloizou J. C. B., Terquem C., 2001, MNRAS, 325, 221
Papaloizou J. C. B., Nelson R. P., Masset F., 2001, A&A, 366, 263
Papaloizou J. C. B., Nelson R. P., Kley W., Masset F. S., Artymowicz P., 2007, Protostars and Planets V, pp 655–668
Park K., Ricotti M., 2011, ApJ, 739, 2
Park K., Ricotti M., 2012, ApJ, 747, 9
Perets H. B., Alexander T., 2008, ApJ, 677, 146
Pérez L. M., Isella A., Carpenter J. M., Chandler C. J., 2014, ApJ, 783, L13
Perez S., et al., 2015, ApJ, 798, 85
Pérez L. M., et al., 2016, Science, 353, 1519
Perley R. A., Dreher J. W., Cowan J. J., 1984, ApJ, 285, L35
Perna R., Lazzati D., Giacomazzo B., 2016, ApJ, 821, L18
Perna R., Chruslinska M., Corsi A., Belczynski K., 2018, MNRAS, 477, 4228
Peters P. C., 1964, Physical Review, 136, 1224
Pezzulli E., Volonteri M., Schneider R., Valiante R., 2017, MNRAS, 471, 589
Pfister H., Lupi A., Capelo P. R., Volonteri M., Bellovary J. M., Dotti M., 2017, MNRAS, 471, 3646
Pinilla P., et al., 2015, A&A, 584, A16
Pinilla P., et al., 2017, ApJ, 839, 99
Pontoppidan K. M., Blake G. A., van Dishoeck E. F., Smette A., Ireland M. J., Brown J., 2008, ApJ, 684, 1323
Postnov K. A., Yungelson L. R., 2014, Living Reviews in Relativity, 17, 3
Price D. J., 2007, Publications of the Astronomical Society of Australia, 24, 159
Price D. J., 2012, Journal of Computational Physics, 231, 759
Price D. J., Federrath C., 2010, MNRAS, 406, 1659
Price D. J., Laibe G., 2015, MNRAS, 451, 813
Price D. J., et al., 2018a, PASA, 35, e031
Price D. J., et al., 2018b, MNRAS, 477, 1270
Pringle J. E., 1991, MNRAS, 248, 754
Pringle J. E., King A., 2007, Astrophysical Flows
Quanz S. P., Amara A., Meyer M. R., Kenworthy M. A., Kasper M., Girard J. H., 2013, ApJ, 766, L1
Quinlan G. D., 1996, New Astronomy, 1, 35
Rafikov R. R., 2002, ApJ, 569, 997
Rafikov R. R., 2013, ApJ, 774, 144
Ragusa E., Lodato G., Price D. J., 2016, MNRAS, 460, 1243
Ragusa E., Dipierro G., Lodato G., Laibe G., Price D. J., 2017, MNRAS, 464, 1449
Ragusa E., Rosotti G., Teyssandier J., Booth R., Clarke C. J., Lodato G., 2018, MNRAS, 474, 4460
Rameau J., et al., 2017, AJ, 153, 244
Rasio F. A., Ford E. B., 1996, Science, 274, 954

Ratzka T., Köhler R., Leinert C., 2005, A&A, 437, 611
Rau A., et al., 2009, PASP, 121, 1334
Regály Z., Juhász A., Sándor Z., Dullemond C. P., 2012, MNRAS, 419, 1701
Regály Z., Juhász A., Nehéz D., 2017, ApJ, 851, 89
Reggiani M., et al., 2018, A&A, 611, A74
Rezzolla L., 2009, Classical and Quantum Gravity, 26, 094023
Rodriguez C., Taylor G. B., Zavala R. T., Peck A. B., Pollack L. K., Romani R. W., 2006, ApJ, 646, 49
Rodriguez C. L., Haster C.-J., Chatterjee S., Kalogera V., Rasio F. A., 2016, ApJ, 824, L8
Roedig C., Sesana A., 2014, Monthly Notices of the Royal Astronomical Society, 439, 3476
Roedig C., Sesana A., Dotti M., Cuadra J., Amaro-Seoane P., Haardt F., 2012, A&A, 545, A127
Rosotti G. P., Lodato G., Price D. J., 2012, MNRAS, 425, 1958
Rosotti G. P., Ercolano B., Owen J. E., Armitage P. J., 2013, MNRAS, 430, 1392
Rosotti G. P., Booth R. A., Clarke C. J., Teyssandier J., Facchini S., Mustill A. J., 2017, MNRAS, 464, L114
Rossi E. M., Lodato G., Armitage P. J., Pringle J. E., King A. R., 2010, MNRAS, 401, 2021
Ruge J. P., et al., 2016, preprint, (arXiv:1603.05179)
Ruíz-Rodríguez D., Ireland M., Cieza L., Kraus A., 2016, MNRAS, 463, 3829
Ryu T., Perna R., Haiman Z., Ostriker J. P., Stone N. C., 2018, MNRAS, 473, 3410
Sakurai Y., Inayoshi K., Haiman Z., 2016, MNRAS, 461, 4496
Salinas V. N., Hogerheijde M. R., Mathews G. S., Öberg K. I., Qi C., Williams J. P., Wilner D. J., 2017, A&A, 606, A125
Sallum S., et al., 2015, Nature, 527, 342
Salyk C., Blake G. A., Boogert A. C. A., Brown J. M., 2007, ApJ, 655, L105
Savchenko V., et al., 2017a, ApJ, 846, L23
Savchenko V., et al., 2017b, ApJ, 848, L15
Schmidt M., 1963, Nature, 197, 1040
Schwarzschild K., 1916, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), 1916, Seite 189-196, pp 189–196
Schwarzschild K., 1999, ArXiv Physics e-prints,
Sesana A., 2012, Advances in Astronomy, 2012, 805402
Sesana A., Haiman Z., Kocsis B., Kelley L. Z., 2018, ApJ, 856, 42
Seyfert C. K., 1943, ApJ, 97, 28
Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337
Shannon R. M., et al., 2015, Science, 349, 1522
Shi J.-M., Krolik J. H., 2015, ApJ, 807, 131
Shi J.-M., Krolik J. H., Lubow S. H., Hawley J. F., 2012, ApJ, 749, 118
Shu F. H., 1984, in Greenberg R., Brahic A., eds, IAU Colloq. 75: Planetary

Rings. pp 513–561
Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, MNRAS, 380, 877
Skrutskie M. F., Dutkevitch D., Strom S. E., Edwards S., Strom K. M., Shure M. A., 1990, AJ, 99, 1187
Springel V., Hernquist L., 2002, MNRAS, 333, 649
Stepinski T. F., Valageas P., 1996, A&A, 309, 301
Stolker T., et al., 2016, A&A, 595, A113
Stone N. C., Metzger B. D., Haiman Z., 2017, MNRAS, 464, 946
Strom K. M., Strom S. E., Edwards S., Cabrit S., Skrutskie M. F., 1989, AJ, 97, 1451
Syer D., Clarke C. J., 1995, MNRAS, 277, 758
Szulágyi J., Plas G. v. d., Meyer M. R., Pohl A., Quanz S. P., Mayer L., Daemgen S., Tamburello V., 2018, MNRAS, 473, 3573
Tanaka H., Ward W. R., 2004, ApJ, 602, 388
Tanaka H., Takeuchi T., Ward W. R., 2002, ApJ, 565, 1257
Tang Y., MacFadyen A., Haiman Z., 2017a, MNRAS, 469, 4258
Tang Y.-W., et al., 2017b, ApJ, 840, 32
Tang Y., Haiman Z., MacFadyen A., 2018, MNRAS, 476, 2249
Taniguchi Y., Shioya Y., Tsuru T. G., Ikeuchi S., 2000, PASJ, 52, 533
Tazzari M., Lodato G., 2015, MNRAS, 449, 1118
Terquem C. E. J. M. L. J., 2003, MNRAS, 341, 1157
Testi L., et al., 2014, Protostars and Planets VI, pp 339–361
Teukolsky S. A., 2015, Classical and Quantum Gravity, 32, 124006
Teyssandier J., Ogilvie G. I., 2016, MNRAS, 458, 3221
Teyssandier J., Ogilvie G. I., 2017, MNRAS, 467, 4577
Thun D., Kley W., Picogna G., 2017, A&A, 604, A102
Tolman R. C., 1939, Physical Review, 55, 364
Turner N. J., Fromang S., Gammie C., Klahr H., Lesur G., Wardle M., Bai X.-N., 2014, Protostars and Planets VI, pp 411–432
Valtonen M. J., et al., 2008, Nature, 452, 851
Varnière P., Quillen A. C., Frank A., 2004, ApJ, 612, 1152
Vasiliev E., Antonini F., Merritt D., 2015, ApJ, 810, 49
Verrecchia F., et al., 2017a, ApJ, 847, L20
Verrecchia F., et al., 2017b, ApJ, 850, L27
Vicente S., et al., 2011, A&A, 533, A135
Volonteri M., 2012, Science, 337, 544
Wang L., Greene J. E., Ju W., Rafikov R. R., Ruan J. J., Schneider D. P., 2017, ApJ, 834, 129
Ward W. R., 1986, Icarus, 67, 164
Ward W. R., 1988, Icarus, 73, 330
Ward W. R., 1991, in Lunar and Planetary Science Conference.
Ward W. R., 1997, Icarus, 126, 261

Ward W. R., Hourigan K., 1989, ApJ, 347, 490
Weidenschilling S. J., 1977, Ap&SS, 51, 153
Willson M., et al., 2016, preprint, (arXiv:1608.03629)
Windmark F., Birnstiel T., Ormel C. W., Dullemond C. P., 2012, A&A, 544, L16
Wise J. H., 2018, preprint, (arXiv:1807.06080)
Woitke P., et al., 2016, A&A, 586, A103
Woosley S. E., 2016, ApJ, 824, L10
Woosley S. E., 2017, ApJ, 836, 244
Woosley S. E., Blinnikov S., Heger A., 2007, Nature, 450, 390
Wright C. M., et al., 2015, MNRAS, 453, 414
Wysocki D., Gerosa D., O'Shaughnessy R., Belczynski K., Gladysz W., Berti E., Kesden M., Holz D. E., 2018, Phys. Rev. D, 97, 043014
Yang Y., et al., 2017, AJ, 153, 7
Youdin A. N., 2010, in Montmerle T., Ehrenreich D., Lagrange A.-M., eds, EAS Publications Series Vol. 41, EAS Publications Series. pp 187–207 (arXiv:0807.1114), doi:10.1051/eas/1041016
Youdin A. N., Goodman J., 2005, ApJ, 620, 459
Young M. D., Clarke C. J., 2015, MNRAS, 452, 3085
Young M. D., Baird J. T., Clarke C. J., 2015, MNRAS, 447, 2907
Yu Q., 2002, MNRAS, 331, 935
Zhang K., Hamilton D. P., Matsumura S., 2013, ApJ, 778, 6
Zhang K., Isella A., Carpenter J. M., Blake G. A., 2014, ApJ, 791, 42
Zhu Z., Stone J. M., 2014, ApJ, 795, 53
Zhu Z., Nelson R. P., Hartmann L., Espaillat C., Calvet N., 2011, ApJ, 729, 47
Zhu Z., Hartmann L., Nelson R. P., Gammie C. F., 2012a, ApJ, 746, 110
Zhu Z., Nelson R. P., Dong R., Espaillat C., Hartmann L., 2012b, ApJ, 755, 6
Zhu Z., Stone J. M., Rafikov R. R., Bai X.-n., 2014, ApJ, 785, 122
Zhu Z., Andrews S. M., Isella A., 2018, MNRAS, 479, 1850
de Mink S. E., King A., 2017, ApJ, 839, L7
de Val-Borro M., et al., 2007, A&A, 471, 1043
de Val-Borro M., Gahm G. F., Stempels H. C., Pepliński A., 2011, MNRAS, 413, 2679
del Valle L., Volonteri M., 2018, preprint, (arXiv:1807.03844)
van der Marel N., et al., 2013, Science, 340, 1199
van der Marel N., et al., 2016a, A&A, 585, A58
van der Marel N., Verhaar B. W., van Terwisga S., Merín B., Herczeg G., Ligterink N. F. W., van Dishoeck E. F., 2016b, A&A, 592, A126
van der Marel N., Cazzoletti P., Pinilla P., Garufi A., 2016c, ApJ, 832, 178
van der Marel N., et al., 2018, ApJ, 854, 177
van der Plas G., et al., 2017, A&A, 597, A32

Acknowledgements

"Silent gratitude isn't very much to anyone."

Gertrude Stein

Below, I report for completeness the formal acknowledgements sections of the individual projects presented in this thesis

Chapter 6: We thank Cathie Clarke for useful discussions. We thank Davide Gerosa for a careful reading of the manuscript and precious comments. We acknowledge the referee for constructive comments and useful suggestions that improved the manuscript. ER and GL acknowledge financial support from PRIN MIUR 2010-2011, project "The Chemical and Dynamical Evolution of the Milky Way and Local Group Galaxies", prot. 2010LY5N2T. DJP is supported by a Future Fellowship from the Australian Research Council (FT130100034). Some of the plots in this thesis were produced using SPLASH (Price 2007).

Chapter 7: We thank Gordon Ogilvie for useful comments and fruitful discussions. We thank the referee, Gennaro D'Angelo, for his constructive comments and useful suggestions that improved the manuscript. This work has been supported by the DISCSIM project, grant agreement 341137 funded by the European Research Council under ERC-2013-ADG. JT acknowledges support from STFC through grant ST/L000636/1. This work used the Wilkes GPU cluster at the University of Cambridge High Performance Computing Service (http://www.hpc.cam.ac.uk/), provided by Dell Inc., NVIDIA and Mellanox, and part funded by STFC with industrial sponsorship from Rolls Royce and Mitsubishi Heavy Industries. We also thank the MIAPP for hosting us for the "Protoplanetary Disks and Planet Formation and Evolution" topical workshop held in Munich during June 2017.

All the figures were generated with the python-based package matplotlib Hunter (2007).

Chapter 9: We acknowledge the anonymous referee for constructive comments and useful suggestions that improved the paper. ER, GD and G. Lodato acknowledge funding via PRIN MIUR prot. 2010LY5N2T. G. Laibe is funded by ERC FP7 grant ECOGAL. DJP acknowledges funding via an Australian Research Council Future Fellowship FT130100034 and Discovery Project DP130102078. We used SPLASH (Price 2007). We thank S. Casassus, A. Sicilia-Aguilar, N. Van der Marel and J. Cuadra for useful discussions.

More personally, I feel really indebted to a number of people that I have been working with through these years, that I want to explicitly thank for their contribution to my scientific and personal growth.

I would like to offer my special thanks to my supervisor, Prof. Giuseppe Lodato, for advising me during these Ph.D. years. I believe that working together really shaped my scientific criticism and taught me how relevant details are in order to produce good science (this might seem obvious, but you should try at least once to present a paper in front of him for a journal club and go beyond page three in one hour discussion: you will understand how this is not obvious at all). Putting aside scientific teachings, which were a lot, to me you represent an example of how passion and knowledge improves us as people and not only as scientists. I will really treasure your scientific and non-scientific lessons.

I am deeply grateful to Prof. Daniel Price: this thesis would definitely not have been possible without your help and without PHANTOM. Thank you for your contribution to my research and for allowing me to participate to yours. I am really grateful for your teachings: computation, astronomy, paper writing, catchy titles editing and talks, you really gave me good tips and lessons in any ability good astronomers should have in their pockets.

I want also to warmly thank Prof. Cathie Clarke for accepting to host me as a visitor at the Institute of Astronomy in Cambridge to start together the massive project that would have later become Chapter 7 of this thesis. I am so grateful of this opportunity, you have been so motivating and inspiring for me. You are a scientist with such an immense scientific knowledge but also with great human qualities. The best wish I can give to any student is to be lucky as I was: having the chance to meet a professor like you during my early scientific career.

I want to sincerely thank my co-authors in the projects presented in this thesis: Dr. Giovanni Dipierro, Prof. Guillaume Laibe, Dr. Giovanni Rosotti, Dr. Jean Teyssandier, Dr. Richard Booth. Thank you for the energy and enthusiasm you put in these projects!

I want to thank Prof. Gordon Ogilvie and Prof. Zoltán Haiman, who refereed this thesis, for their valuable comments and suggestions that improved this manuscript. I want also to thank Prof. Wilhelm Kley and Dr. Chris Nixon for accepting to be the examiners of my thesis defense.

I want to thank Davide Gerosa for the fruitful discussion we had in July about stellar black hole binaries, which really helped me during the writing up of Sec.

5.1.

I want to thank my collegues and friends Dr. Hossam Aly and Dr. Claudia Toci for scientific discussion and for a careful proofreading of this thesis.

More generally, I want to thank the incredible people I had the privilege to work with throughout this Ph.D. experience. I want to express my gratitude to Gregory Herczeg for involving me in the large observational ALMA cycle 4 survey in the Taurus, of which he is the PI, and all the members of the ALMATaurus collaboration.

I am particularly grateful for the assistance I received by the computational centre of the University of Milan. In particular, I want to explicitly thank Dr. Michele Bensi, Dr. David Rebatto, Dr. Miguel Villaplana, Dr. Franco Leveraro and Prof. Francesco Prelz for the technical support and help they provided me with.

I want to thank Andrea Zanzani from the Ph.D. secretariat for his help in dealing with reimbursements, deadlines, forms and any other bureaucratic issue I had to deal with throughout these years.

I want to thank Francesca Tucci for creating the wonderful drawing on the cover of the thesis. Every time I look at it I still wonder how you were able to create such a terrific piece of art from the terrible pencil sketch I provided you with. Unbelievable, that's talent indeed (for those who do not have the paper version with cover, see Fig. D.1)!

I want to thank my past and present office mates, Giovanni, Alessia, Giulia, Fabricia, Quentin, Benedetta, Maria Giulia, Martina, Chiara, Andrea, Sheida and lunch partners, Federico and Francesco, for making our working environment very relaxed, supportive and friendly. The pleasure of going to work every day is something that really characterized my Ph.D. experience and positively contributed to the realization of this thesis. I am really grateful to all of you for this.

Last, but absolutely not least, I want to thank my friends and family for the love and support they always showed me. In particular, I want to explicitly thank my brother Guido to whom I want to dedicate this thesis. Despite your being the younger brother, you now always wisely act as the older one. Obviously, this is a way to thank you for making somehow this thesis possible: saving me from certain death by starvation cooking wonderful meals for the two of us during our "studio matto e disperatissimo" last August in Milan, while writing up this thesis. What a team! You cook, I watch!

Thank you indeed, everyone!

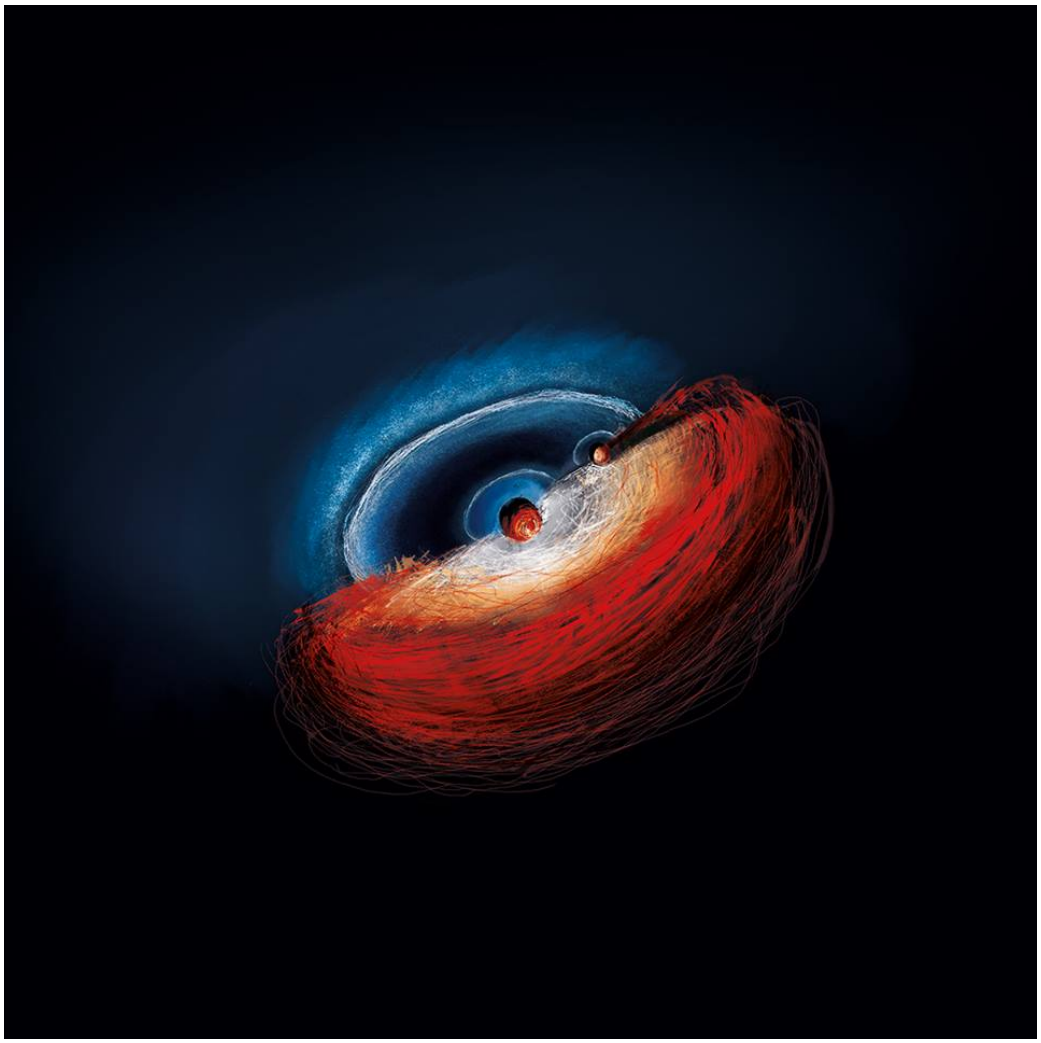


Figure D.1: Cover image of this thesis: artistic impression of a protoplanetary system and a black hole binary one (...in false colours). Credits and gratitude to Francesca Tucci.