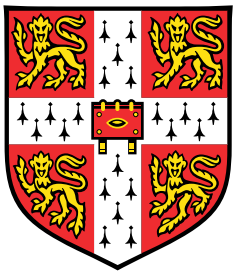

Secular Dynamics of Self-gravitating Debris Discs

Antranik Awedis Sefilian

Supervisor: Prof. Roman Rafikov

Department of Applied Mathematics and Theoretical Physics
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

Darwin College

January 2022

DECLARATION OF ORIGINALITY

I, Antranik Awedis Sefilian, hereby declare that this dissertation, entitled *"Secular dynamics of self-gravitating debris discs"*, is the result of my own work and includes nothing which is the outcome of work done in collaboration, except as declared below and clearly indicated in the text. This dissertation is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution. This dissertation does not exceed the prescribed word limit for the relevant Degree Committee.

Chapters 1, 2, and 3 constitute an introduction and review of background material related to the topic of this dissertation and as such, contain little or no original work. Figures are included from the work of other authors in these Chapters, which are clearly referenced in the associated captions. The remaining chapters represent original work, from which the following have been published in, submitted to, or in preparation for submission to, academic journals:

- **Chapter** 4 is adapted from the published work below which was completed in collaboration with the named authors.
  "Potential softening and eccentricity dynamics in razor-thin, nearly Keplerian discs"
  Sefilian, A. A., & Rafikov, R. R. 2019, MNRAS, 489, 4176
  The analytic calculation in Appendix A.4 has been made by Prof. Roman Rafikov.

- **Chapter** 5 is adapted from the published work below which was completed in collaboration with the named authors.
  "Formation of gaps in self-gravitating debris disks by secular resonance in a single-planet system. I. A simplified model"
  Sefilian, A. A., Rafikov, R. R., & Wyatt M. C. 2021, ApJ, 910, 13

- **Chapter** 6 describes unpublished research conducted in collaboration with Prof. Roman Rafikov and Prof. Mark Wyatt. This work is currently in preparation for submission as a peer-reviewed publication.

Except as described above, the co-authors of the listed publications only contributed indirectly to this dissertation, by providing useful feedback on the analysis, results, and writing during the drafting process of the above manuscripts.

Antranik Awedis Sefilian
January 2022

SUMMARY

Secular Dynamics of Self-gravitating Debris Discs

Antranik Awedis Sefilian

Debris discs are the leftovers from the epoch of planetary formation, akin to the Solar System's asteroid and Kuiper belts. They preserve a record of the physical processes at work throughout the system's age, not least planet–disc interactions. Deciphering this record has become a central focus of modern celestial mechanics, as it may provide unique insights into the formation and evolution of exoplanetary systems, in addition to indirectly detecting planets based on the structures (e.g., gaps and spiral arms) they imprint on debris discs. However, the problem is compounded by two interrelated issues. First, most existing studies of planet—debris disc interactions do not account for the (self-)gravitational effects of debris discs. That is, debris discs are usually treated as a collection of massless particles, subject only to the gravity of external massive companions. This is partly related to the second issue pertaining to the theoretical and computational challenges involved in modelling the long-term (secular) gravitational effects of discs in general. This dissertation aims to partly address these two problems.

In the first part of the dissertation, I provide a comparative analysis of the two existing methods for the analytical computation of the secular disturbing function due to a flat disc: softened and non-softened. I first show that the various methods which are based on softening the Newtonian point-mass potential, in the appropriate limit, may converge to the expected non-softened behaviour exactly, approximately, or not converge at all. I then explain this variety of outcomes by formulating a generalised Laplace–Lagrange theory for flat discs which is applicable for a wide variety of softening prescriptions. Finally, I study the conditions that must be obeyed by any discretised numerical treatment of self-gravitating discs to accurately reproduce the expected gravitational potential, finding that a fine numerical sampling is required.

In the second part of the dissertation, I explore the secular dynamical interaction between an eccentric planet and a massive, external debris disc. I first employ a simplified model accounting only for the axisymmetric component of the disc gravity to probe the ensuing basic dynamical effects. I find that even when the disc is less massive than the planet, the system may feature secular resonances which in turn can sculpt a gap within the debris disc through excitation of planetesimal eccentricities. I then characterise the properties of the secular resonances and illustrate how the planet–disc parameters can be constrained for a given system hosting a gapped disc. Then, using

a numerical method developed based on the first part of the dissertation, I examine the case where the full gravity of the debris disc is accounted for, including both axi- and non-axisymmetric components. In this case, I find new dynamical phenomena whereby the secular resonances cause e.g. the planetary orbit to circularise in time, as well as to launch a one-armed spiral arm beyond the gap. As an example, I apply these results to known gapped debris discs, and show that they could be sculpted by a single interior planet/companion, provided the discs have enough mass.

In summary, then, the contributions of this dissertation are three-fold: to better understand the methods for quantifying the gravitational potential of discs in general; to propose a new mechanism for sculpting gaps in debris discs; and more generally, to pave the way for advancing the theory of debris disc modelling by including the effects of disc gravity.

ACKNOWLEDGEMENTS

"We are not going in circles, we are going upwards. The path is a spiral; we have already climbed many steps."

-Hermann Hesse, *Siddhartha*

It is so strange to write these words from my office on the shores of the Mediterranean, realising that my time as a student at Cambridge – which has been of virtual nature over the last six months – is on the verge of running out. I find myself reflecting back on the journey that has led me to where I am today, and feeling extremely lucky and grateful for the opportunity to have pursued my studies at a university that I never dreamt I could reach. This would surely have not been possible without the guidance and support of many people whom I would like to thank sincerely.

I would first like to thank my supervisor, Roman Rafikov, for having believed in me and supporting my application to pursue a PhD at Cambridge, and for being willing and enthusiastic to assist in any way he could throughout my research. Undertaking this PhD under your supervision has been a truly life-changing experience for me, to say the least, teaching me lessons that I will carry throughout my academic journey.

I am also grateful to Henrik Latter, both for having his door always open to discuss just about anything, whether academic related or not, and for keeping an eye out for me, especially during my darker moments. Your advice at critical junctures along the way made a big difference, and I deeply appreciate it.

I would also like to sincerely express my gratitude to Mark Wyatt, not only for his patience and time to explain various concepts and theories related to debris discs, but also for his direct input to the dissertation through our work on gaps in debris discs. Thank you for being there for me as a secondary advisor.

Another person to whom I am indebted is Jihad Touma, for being a mentor and a friend throughout my journey in academia and life, and for continually instilling motivation and confidence in me, to say the least. I also highly appreciate his support in securing an office in the Center for Advanced Mathematical Sciences at the American University of Beirut, which provided me the right environment over the last six months to complete writing up my thesis amidst the chaotic situation Lebanon is in.

I also wish to thank my examiners, Gordon Ogilvie and Rebekah Dawson, for taking the time to read my work, as well as their perceptive and constructive comments.

I also express my sincere gratitude to the Department of Applied Mathematics and Theoretical Physics and Darwin College for all the administrative support they have provided over the last four years. I also acknowledge the financial support provided by the Gates Cambridge Trust, without which I would not have been here writing these words and this dissertation. It was a great honour to be a Gates Cambridge scholar.

I would also like to thank the new and amazing friends I made in Cambridge: William Bethune, Chris Hamilton, Luis Welbanks, Hayk Saribekyan, Loren Held, Roger Dufresne, Janosz Dewberry, Cleo Loi, Philip Leung, among many others. They have each enriched my academic and social life in multitude of ways, and I am sincerely thankful for their friendship. A special thanks goes out to William Bethune, without whom this whole endeavour would have been much more challenging, to say the least.

I also extend my thanks to my colleagues at Mark's research group: Sebastian Marino, Nicole Pawellek, Jessica Rigley, Joshua Lovell, Marija Jankovic, Steven Young, and Yinuo Han. I have learned a great deal about debris discs from each and every one of you. A special thanks goes out to Sebastian who supported me both as a colleague and a friend.

It goes without saying, but I would be remiss if I were not to thank my friends abroad, who have all contributed to my general well-being, sharpening my sanity when needed, knowingly or unknowingly. In particular, I thank Hrag Karakachian, Rodrique Badr, and Lama Tannoury for being there for me no matter what.

I must also acknowledge that part of this work was carried out during the COVID-19 pandemic. This would not have been possible without the support and kindness of countless people, many of whom are unaware of their contributions.

Finally, I would like to thank my parents – Awedis and Taline – for nourishing the gift of life in me and for believing in me, and my brothers – Hampig and Shant – and Meghety for all the serious and fun times, particularly during the last six months.

Last but not least, I would like to express my infinite thanks and love to my partner, Natasha. You have been a constant amongst the chaos: I would have gone insane without your relentless love, care, and support. Thank you for always being there for me, even when the world seemed hopeless and dark; for believing in me, even during times when I did not believe in myself; and for being my best friend. I cherish every moment we've spent together over the last 14 years, and am looking forward to the adventures and challenges ahead. I owe you everything!

Antranik Awedis Sefilian

Beirut, 17^{th} January, 2022

لِبيروت،
من قلبي سـلَامٌ لِـبيروت.

To Beirut,
From my heart, peace to Beirut.

Contents

Chapter 1

Introduction

1.1 Historical Overview and Motivation

Since the dawn of time, humanity has gazed into the night sky with wonder. With the birth of astronomy with ancient stargazers millennia ago, it has been a consistent human endeavour to understand the nature of the universe, our location within it, and the origins of life: What are the origins of the sparkling objects in the sky? Is the Solar System unique? What are the origins of life on Earth? Are we alone in the universe?

Astronomy has had a long and dramatic history, one filled with discoveries which have and are continually affecting our view of Earth and our place in the universe. These discoveries, beginning with mapping out the phases of the Moon as it orbits the Earth by ancient stargazers, the identification of planets in our solar system by the ancient Greeks, to the shift from Ptolemy's geocentric model in 200 A.D. to the heliocentric model of Copernicus during the Renaissance, followed by the discovery of galaxies in the early 20^{th} century, have all shown that our place in the cosmos is not as unique as previously thought: neither the Earth nor our Sun is at the centre of the Universe, but rather that the Solar System is a mere member of some ~ 400 billion stars in our Galaxy, which itself represents one out of more than a trillion other galaxies in the observable Universe!

The golden age of astronomy began in the early 1600s, when Galileo first pointed a telescope to the skies and discovered that Jupiter is encircled by its own moons. This essentially opened the door to the discovery of additional planets within the Solar System and beyond. Through the centuries that followed, our understanding of the motion of planets developed tremendously with the invention of celestial mechanics by Kepler and Newton. In parallel, improvements in optics and telescopes opened the door to the discovery of additional planets within our Solar System, and indeed led to

the detection of Uranus by Herschel in 1781. At the same time, Laplace published his treatise on celestial mechanics, setting the foundations of modern celestial mechanics. The triumph of celestial mechanics came when irregularities in the orbit of Uranus were taken as a gravitational one-way sign for the existence of an additional planet in the Solar System beyond Uranus, and such a planet – Neptune – was discovered by Adams and Le Verrier in the 1840's. This marked the first instance when mathematics, rather than a telescope, was used to predict and discover the presence of a planet.

For most of the 20$^{\text{th}}$ century, however, celestial mechanics remained devoted to the study of the Solar System, with theories of planet formation and evolution designed and tailored to understand its origins and characteristics. However, the possibility that planets existed or not around other stars was not out of the question, with debates that can be traced back to ancient times. Indeed, Aristotle (384–322 BC), a Greek philosopher and polymath, and Copernicus (1473–1543), opposed the idea that planets existed outside the Solar System, while on the other hand, Epicurus (341–270 BC), another ancient Greek philosopher, postulated the presence of an infinite number of planetary systems – see Dick (1996) for a historical review. These ideas evolved throughout the centuries that followed, and the consensus in early 1900s was that exo-planetary systems, if any, must be extremely rare. For instance, Eddington (1928) noted in his *The Nature of the Physical World* that "I should judge that perhaps not one in a hundred millions of stars can have undergone this experience in the right stage and conditions to result in the formation of a system of planets"; see also Jeans (1919).

The evidence for the existence of planets outside our Solar System did not come to light until the early 1990s, when the first two exoplanets were discovered by Wolszczan & Frail (1992). These planets were found to be orbiting around not an ordinary star but, somewhat surprisingly, around a millisecond pulsar – the remnant of a massive star that has exploded as a supernova. Soon after in 1995, the first discovery of an exoplanet orbiting a main-sequence star, namely 51 Peg, was made by Mayor & Queloz (1995); a discovery which kicked the discipline of exoplanet astronomy into high gear. Indeed, during the three decades that followed these discoveries, several thousands of exoplanets have and are being discovered and characterised at an exponential rate using different observational techniques; from radial velocity and transit, to microlensing and direct imaging – each with its own strengths and biases of detection. Such observational campaigns have revealed that exoplanets are (i) more common than previously thought, with an overall planet occurrence rate of $\sim 50\%$, and (ii) vastly diverse (Winn & Fabrycky, 2015). Indeed, the plethora of exoplanets has revealed the existence of a large variety of exoplanets which are very different from what we

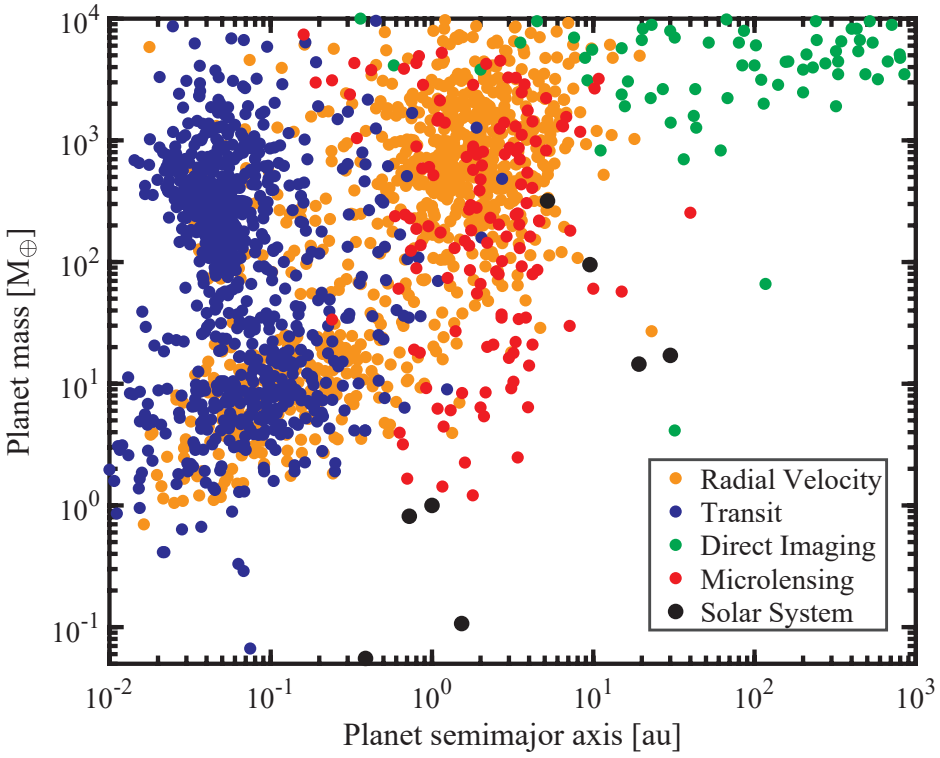

Figure 1.1. The distribution of masses and semimajor axes of currently confirmed exoplanets as discovered by various methods: radial velocity (orange), transit (blue), direct imaging (green), and microlensing (red). For reference, the Solar System planets are shown by black dots. The data were compiled from www.exoplanet.eu on 12th of August, 2021.

known from the Solar System's planetary album, with some being much more massive than Jupiter, while others orbiting their stars at extremely small (~ 0.01 au) or large separations (~ 100 au) on highly non-circular and/or non-coplanar orbits. All of these discoveries have presented serious challenges to our understanding of planet formation and evolution. Figure 1.1 shows the masses and orbital semimajor axes of currently confirmed exoplanets (as of August 12, 2021), along with the Solar System planets for comparison.

Concurrently with the first discoveries of exoplanets, astronomers also detected and confirmed the presence of the long hypothesised population of minor bodies in the outskirts of the solar system beyond Neptune (Kowal et al., 1979; Scotti et al., 1992; Jewitt & Luu, 1993): the Edgeworth–Kuiper belt (Edgeworth, 1943; Kuiper, 1951a), of which Pluto is a member. Since then, the number of known Kuiper belt objects has increased to over a thousand, dramatically changing our view of the Solar System. Indeed, the swarm of Kuiper belt objects represents the relics leftover from the

epoch of planet formation, and as such its orbital structure is thought of as preserving a record of gravitational interactions that took place throughout the lifetime of the Solar System; especially in its early ages. Thus, decoding the orbital architecture of the Kuiper belt has been the common goal of numerous dynamical studies, which have provided many clues into the formation, architecture, and evolution of the Solar System (see the review by Morbidelli et al., 2008, and references therein).

During this same period, i.e., from 1980s and onward, it also became clear that similar to the Solar System, exo-planetary systems are comprised not just of planets, but also of belts of debris which are populated by bodies ranging from micron-sized dust up to kilometre-sized planetesimals – somewhat analogous to the asteroid and Kuiper belts. The first of such *exo-Kuiper belts*, or *debris discs*, was discovered around the main-sequence star Vega (Aumann et al., 1984), and since then the number of detected debris discs has increased to over a thousand. Simultaneously, thanks to significant technical advancements in observational astronomy at all wavelengths, the spatial (sub-)structures of debris discs have also come into sharp resolution. Indeed, observational campaigns have revealed a large variety of debris disc morphologies such as radial gaps, vertical warps, and eccentric rings – hinting at sculpting by planets that are not detected as of yet (see e.g., the reviews by Hughes et al., 2018; Wyatt, 2018). Simultaneously with these discoveries, the field of planet formation has also been revolutionised through imaging of protoplanetary discs, the nurseries in which planets are ultimately born, allowing us to probe ongoing planet formation processes (Williams & Cieza, 2011; ALMA Partnership et al., 2015). A remarkable discovery in the recent years is that of two protoplanets orbiting within the cavity of a protoplanetary disc surrounding the PDS 70 star (Keppler et al., 2018; Müller et al., 2018; Haffert et al., 2019), with one of them seen to be surrounded by its own circumplanetary disc where moons can be formed (Isella et al., 2019); see Figure 1.5. With all of these developments, especially during the last three decades or so, I think it is fair to say that we are now living through very exciting times in the field of planetary sciences. Indeed, we are amassing an unprecedented amount of observational information on all stages of planetary system formation and evolution, which is continually challenging our theoretical understanding of the involved processes and mechanisms.

This dissertation is my small contribution to the study of planetary systems outside our own Solar System. My research focuses on understanding the dynamical evolution of exo-planetary systems and how they are shaped through gravitational interactions amongst the different components: star, planets, and debris discs. Just as the dynamical structures of the asteroid and Kuiper belts have revealed unique

insights into the formation and evolution of our own Solar System, detailed images of debris discs – and in particular those with high enough resolution – can be used to infer the formation and evolutionary history of exo-planetary systems, and to even reveal the presence of yet-unseen planets that are potentially shaping the debris discs. For any accurate interpretation of debris disc observations, however, we must have a good understanding of the various processes affecting the evolution of dust and planetesimals in debris discs, including planet–disc interactions. This is where my research comes in. Prior to describing the work conducted during my PhD, however, it is useful to take a step back and describe the current paradigms of star and planetary system formation and evolution, which would also help to put my research into context.

1.2 A Tale of Planetary System Formation and Evolution

Planetary systems represent the final stage of a long and complex process that starts with the formation of stars, to the ultimate formation of planets and circumstellar discs from the remnant gas and dust in the stellar environment, followed by shaping the final architecture of a planetary system due to interactions between planets and discs. Here I provide a brief overview of our current understanding of this process, framing the explanation – when possible – in the context of the planetary system that we known best: our Solar System.

1.2.1 Star Formation

The formation of stars is a relatively well understood process (see e.g. Shu et al., 1987; Carroll & Ostlie, 1996; de Pater & Lissauer, 2010). Stars are born within clouds of interstellar gas and dust, known as molecular clouds. These vast clouds, with typical diameters ranging from ~ 10 to 200 parsecs, are massive and dense enough to be self-gravitating; indeed, molecular clouds can contain more than $10^4 M_\odot$ of material. The structures of molecular clouds are generally complex, and are supported by turbulent, thermal, and magnetic pressures. Following Shu et al. (1987), the process of star formation can be categorised into four stages (see Figure 1.2). First, the interaction of turbulence with gravity and magnetic fields leads to the fragmentation of molecular clouds, or parts of it, into intricate filaments in which slowly rotating cores are formed (Figure 1.2(a)). When the core is of high enough density that its own self-gravity overcomes the internal pressure forces pushing it out, it collapses on itself inside out.

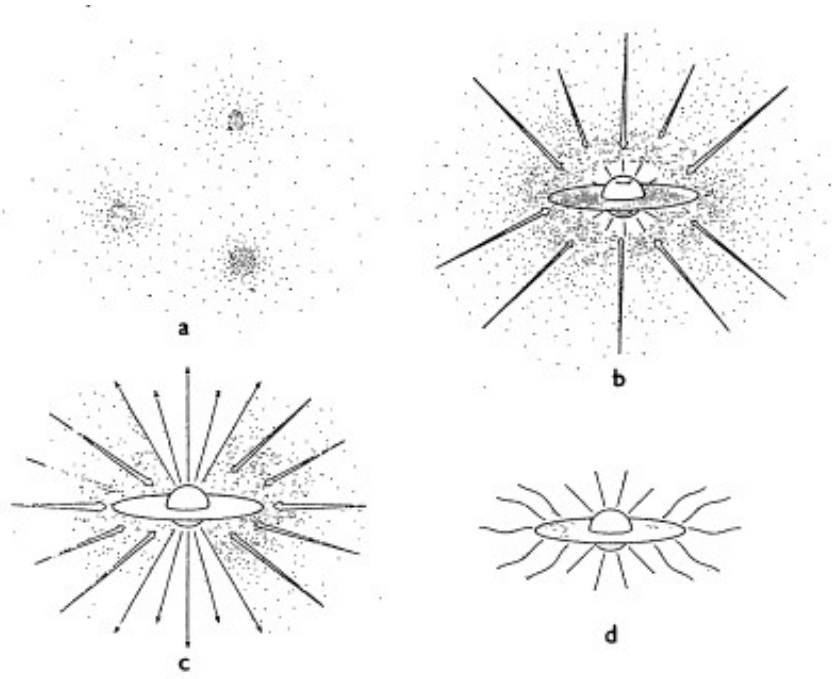

Figure 1.2. A diagram depicting the four stages of star formation (taken from Shu et al., 1987). *Panel a:* Slowly rotating cores form within the molecular cloud as a result of collapsing clumps of gas and dust. *Panel b:* The core collapses inside out forming a central protostar surrounded by an accretion disc, both embedded in an infalling envelope of gas and dust. *Panel c:* The protostar accretes material from the disc and launches a jet. *Panel d:* The envelope dissipates, leaving behind a newly formed star with a protoplanetary disc.

This leads to the formation of a protostar towards the core's centre with a typical mass of $\sim 0.1 - 1 M_{\odot}$, along with an accretion disc surrounding it and lying perpendicular to its axis of rotation; both of which would be embedded within an infalling envelope of dust and gas (Figure 1.2(b)). During this protostellar phase (Classes 0 and I), which is expected to last $\sim 0.1 - 1$ Myr, the protostar and the disc continue to accumulate material from the surrounding envelope through accretion, causing the protostar to grow in mass, heat up, and launch collimated jets (Figure 1.2(c)). Eventually, the envelope dissipates, leaving behind a pre-main sequence star (Class II, e.g. a T-Tauri or Herbig star) surrounded by a protoplanetary disc (Figure 1.2(d)).

1.2.2 Protoplanetary Discs

Protoplanetary discs represent the nurseries in which planets are ultimately born (for an excellent review from both observational and theoretical perspectives, see, e.g., Williams & Cieza, 2011). They typically have a radial extent on the order of ~ 100 au, and are composed of gases (mainly, hydrogen and helium) and dust grains which, collectively,

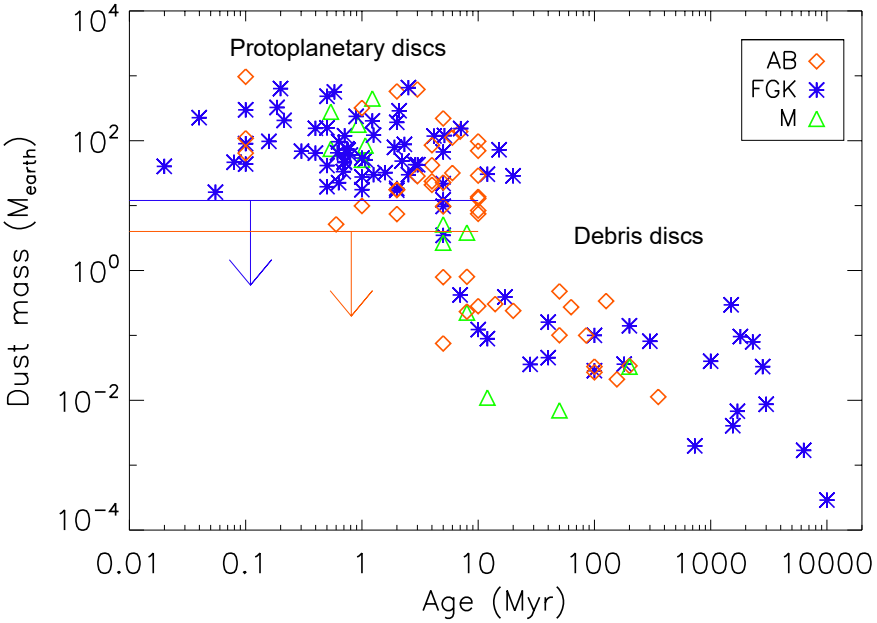

Figure 1.3. The evolution of dust mass in circumstellar discs derived from sub-millimetre observations as a function of stellar age (adapted from Panić et al., 2013). The blue and orange lines correspond to the sensitivity limits for T Tauri stars and Herbig Ae stars, respectively. Note the abrupt drop in dust masses for systems older than ~ 10 Myr.

weigh around $\sim 0.01-0.1 M_{\odot}$ – with the gaseous component dominating the entire mass budget ($\sim 99\%$ in mass). Thus, protoplanetary discs contain a sufficient reservoir of the key ingredients to form multiple gas giant and terrestrial planets. For instance, the lowest mass budget – known as Minimum Mass Solar Nebula (MMSN; Weidenschilling, 1977; Hayashi, 1981) – required to build the Solar System's planets is estimated to be around $0.01 - 0.07 M_{\odot}$. However, protoplanetary discs evolve over time due to various physical processes, such as viscous transport, accretion, photoevaporation, and dynamical interactions with (sub-)stellar companions if any (e.g. Williams & Cieza, 2011, and references therein). These processes cause protoplanetary discs to gradually lose their mass, until their eventual dispersal by $\sim 3 - 10$ Myr (Panić et al., 2013; Ercolano & Pascucci, 2017) – see Figure 1.3. By this time, accretion onto the central star terminates and the latter reaches its main-sequence phase (Class III).

In recent years, thanks to the advent of new generation of instruments, a large number of young protoplanetary discs have been observed, allowing for detailed studies of not only their properties but also their spatial (sub-)structures. Indeed, observations have revealed a variety of structures, such as concentric rings and spiral arms (e.g. ALMA Partnership et al., 2015; Andrews et al., 2016, 2018; Long et al., 2018; van der Marel et al., 2019). As an example, Figure 1.4 shows images of four young

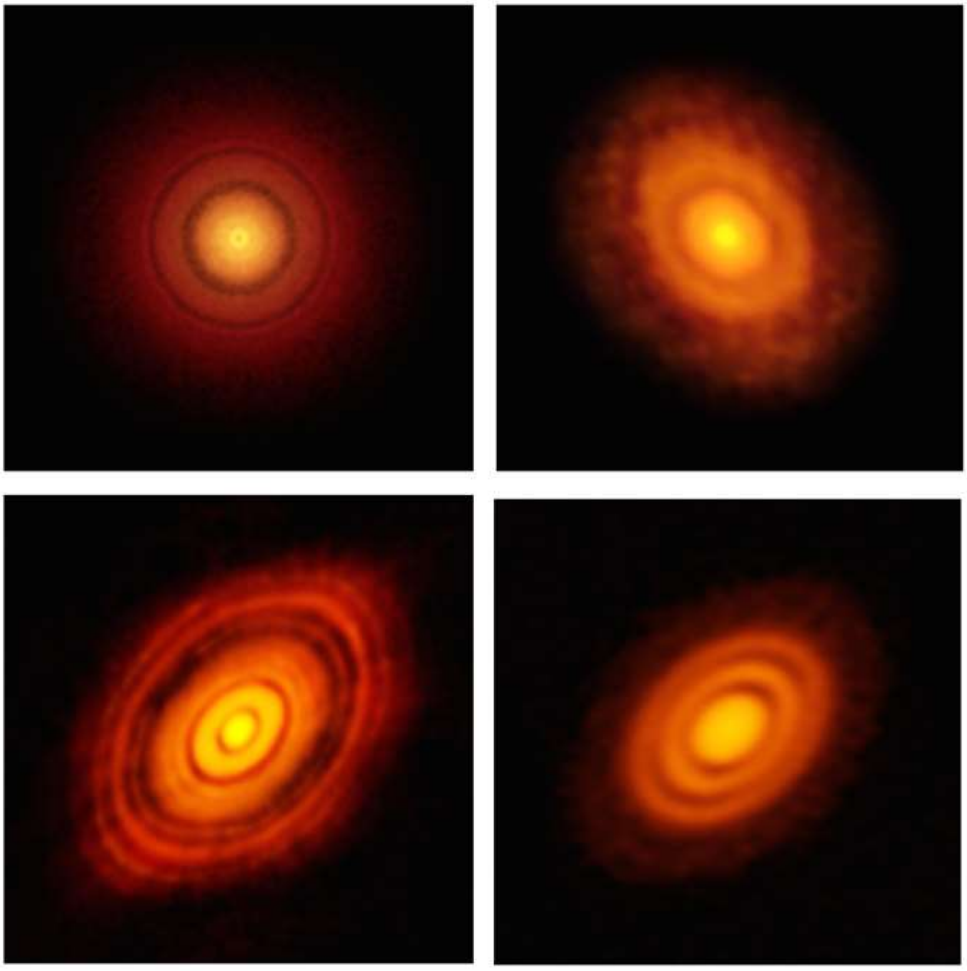

Figure 1.4. High-resolution ALMA images of four protoplanetary discs characterised with ringed (or gapped) structures in their ∼mm dust distributions. From top left to bottom right, these are: TW Hyd (Andrews et al., 2016), V883 Ori (Cieza et al., 2016), HL Tau (ALMA Partnership et al., 2015), and HD 163296 (Isella et al., 2016). One the most commonly accepted explanations for such structures is the presence of yet unseen planets. Credit: ESO.

protoplanetary discs exhibiting a series of concentric rings separated by gaps in their ∼ mm dust distribution as seen by the Atacama Large Millimetre/sub-millimetre Array (ALMA). Such observations have revolutionised the view of protoplanetary discs, which were previously thought to have a smooth density profile.

1.2.3 Planetesimal and Planet Formation

The dispersal timescale of protoplanetary discs presents one of the major challenges to planet formation processes, as it sets a limit on the time available for planets to form. Indeed, within this time period, sub-μm dust grains must grow by at least nine orders

of magnitude, i.e., until they form $\sim$ km-sized objects. This multiplex process can be generally divided into three main steps: planetesimal formation, terrestrial planet formation, and giant planet formation – which are reviewed below. For a more detailed overview of these processes, the reader is referred to the literature (e.g., Lissauer, 1993; Armitage, 2010; Morbidelli et al., 2012; Morbidelli & Raymond, 2016).

The formation of planetesimals, which represent the building blocks of planets, is not well understood and remains an active area of research. In the classical picture, μm-sized dust grains first settle to the midplane of the protoplanetary disc rather quickly – typically in a few thousand years – due to the stellar gravity and strong coupling with the gas. During this process, μm-sized grains collide with each other and grow to $\sim$ mm sizes via pairwise accretion and sticking due to electrostatic forces (e.g., Dominik & Tielens, 1997). The most economical hypothesis following this step is that the same process continues until the formation of $\sim$ km-sized planetesimals (e.g., Weidenschilling, 1980; Weidenschilling & Cuzzi, 1993). This classical picture, however, comes with severe problems (see e.g., Morbidelli & Raymond, 2016). On one hand, when grains become $\sim$ cm- or mm-sized, mutual collisions do not lead to sticking and net growth, but rather to bouncing of grains off of each other, inhibiting further growth (e.g., Zsom & Dullemond, 2008; Güttler et al., 2009). This is the so-called "bouncing barrier". On the other hand, once particles attain $\sim$metre sizes, they would decouple from the gas and migrate towards the star due the drag force they experience because of the deviation between the gas and Keplerian orbital velocities. This gives rise to large relative velocities among particles of various sizes, and hence destructive collisions. At the same time, the migration happens quickly, so much so that particles are lost from the system way before they can grow any further (Weidenschilling, 1977). For instance, a particle of 1 meter at 1 au would drift towards the Sun in a few hundred years. This is the well-known "meter-size barrier".

Over the past decade, several mechanisms have been proposed to overcome these barriers (see, e.g., the review by Blum, 2018). One promising mechanism, known as the streaming instability, is that planetesimals can be formed directly through the gravitational collapse of a concentrated clump of mm-sized dust pebbles, bypassing the barriers described above (Youdin & Goodman, 2005; Johansen et al., 2009; Bai & Stone, 2010; Yang et al., 2017). Another mechanism which has been proposed is that dust grains grow as fractals composed of sub-micrometre monomer particles which remain sticky and well coupled to the gaseous component of the disc (Kataoka et al., 2013). Other possibilities involve modified collisional processes (Boley et al., 2014), clumping of small particles due to disc turbulence (Cuzzi et al., 2001; Johansen et al.,

2007), and even accounting for the gravitational effects of the protoplanetary disc as a means of reducing collisional velocities particularly in more hostile systems hosting binary stars (e.g., Silsbee & Rafikov, 2015a,b, 2021).

I now move on from planetesimals to discuss the formation of terrestrial planets. Once ∼km-sized planetesimals are formed in the protoplanetary disc, their further growth is reasonably well understood. Indeed, planetesimals can keep growing by accreting smaller bodies (i.e., pebbles) present in the disc through a process known as "runaway growth" (e.g., Greenberg et al., 1978; Ida & Makino, 1993). This results in several objects which are large and massive enough (i.e., ~ 1000 km in size and $\sim 10^{-2} - 10^{-1} M_\oplus$ in mass) that they can be considered as planetary embryos, or protoplanets. Such embryos interact with each other gravitationally, while also perturbing the material near their orbits, hampering the growth of additional planetesimals. From this stage on, embryos grow through mutual collisions under gravitational focusing, resulting in terrestrial planets similar to those in the Solar System. This final stage could extend well beyond the protoplanetary disc lifetime, with the final assembly of terrestrial planets taking around ~ 100 Myr.

As to giant planet formation, there are two competing models – core accretion and gravitational instability – each with its own strengths and weaknesses. In the core accretion scenario (see e.g. Pollack et al., 1996; Rafikov, 2004; Levison et al., 2010), a solid core of $\sim M_\oplus$ has already formed according to the scenario described above. In addition to accreting planetesimals, however, the core now starts to accrete gas and form an atmosphere. Once the mass of the newly acquired atmosphere becomes comparable to the solid core ($\sim 10 M_\oplus$), the gas accretion accelerates, acquiring more atmosphere mass, until no more gas remains along the planetary orbit. This scenario can readily result in gas giant planets, similar to Saturn and Jupiter, and even ice giant planets, similar to Uranus and Neptune. This said, core accretion has difficulty in explaining the origin of $\gtrsim 1 M_J$ exoplanets, especially at large distances, as the timescale for enough gas to be accreted onto the planetary embryo can well exceed the protoplanetary disc lifetime (Matsuo et al., 2007; Rafikov, 2011).

In the gravitational instability scenario (see e.g., Kuiper, 1951b; Boley, 2009; Boss, 1997, 2011), on the other hand, giant planets are formed as a result of local instabilities in the protoplanetary disc. Such gravitational instabilities would fragment the disc into dense clumps of material, which then quickly accrete the surrounding gas forming giant planets within the space of a few thousand years. The development of the gravitational instability, however, requires the disc to be massive or cold enough. As such, this

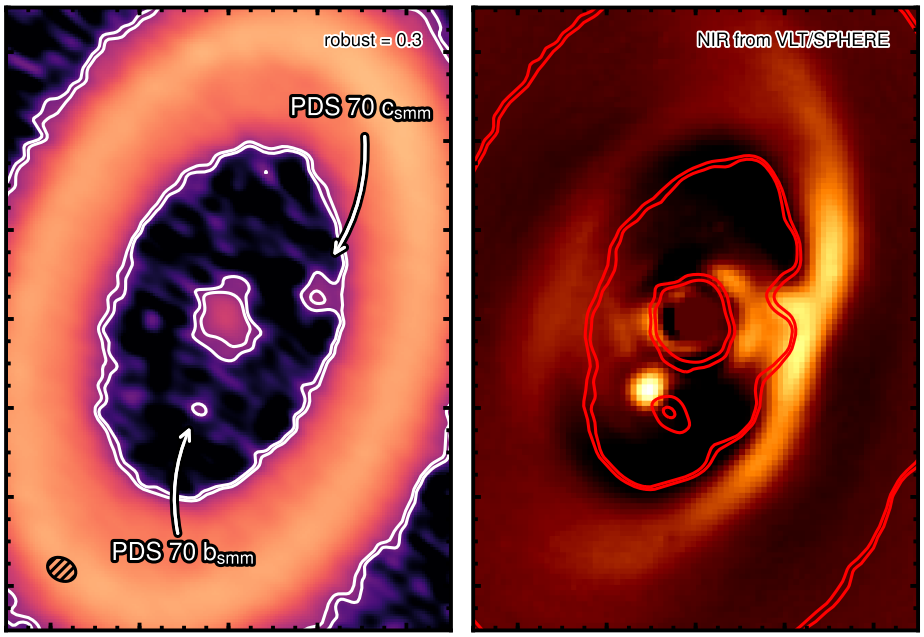

Figure 1.5. Observations of the PDS 70 system (taken from Isella et al., 2019). *Left:* ALMA continuum observations of the PDS 70 protoplanetary disc, along with the locations of the two planets, PDS 70b and PDS 70c. The latter is thought to be surrounded by a circumplanetary disc. *Right:* Near-infrared observation of the system using VLT/SPHERE (Müller et al., 2018).

scenario tends to form giants (or even brown dwarfs) at large separations from the star (i.e., $\gtrsim 50$ au), as it is very unlikely to operate close to the star (Rafikov, 2005).

Once planets are formed in one way or another, they generally do not remain on the same orbits indefinitely. Gravitational interactions both amongst themselves as well as the surrounding protoplanetary disc can lead to various dynamical outcomes – for instance, migration of planets, modulation of orbital eccentricities and inclinations, and even scattering and ejection of planets (e.g., Armitage, 2010, and references therein). Planets can also leave observational fingerprints on their surrounding protoplanetary disc e.g. through carving inner cavities, launching spiral arms, and even forming gapped or ringed structures – features which have come into increasingly sharp resolution with the advent of new generation instruments (e.g., Andrews, 2020, and references therein); see also Figure 1.4. While other processes that do not require the presence of planets can also explain such structures, the planet-disc interaction hypothesis remains as the most widely accepted explanation. A remarkable confirmation of this hypothesis has been provided recently by observations of the PDS 70 system, which hosts a ~ 5 Myr T-Tauri star surrounded by a protoplanetary disc extending from $\sim 60 - 140$ au (see Isella et al., 2019, and references therein). Indeed, two gas giant planets, PDS 70b

and PDS 70c, with estimates masses of $\sim 4 - 12 M_J$ were detected to be orbiting the star well within the protoplanetary disc's cavity, i.e., at ~ 25 and 35 au separations, respectively (Keppler et al., 2018; Müller et al., 2018; Haffert et al., 2019) – see Figure 1.5. Further analysis with resolved Hα emissions also indicated that PDS 70 c is surrounded by a circumplanetary disc (Isella et al., 2019). This represents the first time that young planets have been detected in their birth environment, and suggests that planet formation might occur early in the history of a young stellar system.

1.2.4 Debris Discs

Following the dispersal of a protoplanetary disc, the host star (which has now reached its main sequence phase) may have its own planetary system composed of (i) any planets that formed successfully during the protoplanetary phase, (ii) as well as – or, at the very least – of smaller solid materials left over from the epoch of planet formation. Such residual bodies may have failed to accrete into full-sized planets either because their agglomeration into larger bodies was hampered by the gravitational perturbations of newly formed planets (e.g., Safronov, 1991), or because surface densities were too low and accretional timescales too high for planet formation to progress efficiently (e.g., Weissman, 1995). These relics of planet formation are the constituents of planetesimal belts, or debris discs (see reviews by Wyatt, 2008; Krivov, 2010; Hughes et al., 2018; Wyatt, 2018; Wyatt, 2020, from which the following is summarised). Figure 1.6 depicts a schematic diagram of a typical planetary system composed of (both terrestrial and giant) planets and a debris disc.

A good example of debris discs is found in our cosmic backyard (see e.g. reviews by Chapman et al., 1978; Wyatt, 2020; Gladman & Volk, 2021, and references therein). Indeed, our Solar System has its own debris disc, with the majority of its mass being concentrated in two distinct belts: the Asteroid belt, spanning from $\sim$ 2 to 4 au between the orbits of Mars and Jupiter, which is composed of objects ranging from $\sim$ cm-sized pebbles up to $\sim$ 100 km-sized rocky bodies; and the Kuiper belt – of which Pluto is a member – located beyond the orbit of Neptune from $\sim$ 30 to 50 au, which is mainly populated by bodies that are composed of frozen ices. Beyond this region, there also resides a sparse population of icy minor bodies known as trans-Neptunian objects which, collectively, populate the so-called scattered disc, which is further surrounded by the Oort cloud of comets which extends out to thousands of au. Last but not least, another component of the Solar System's debris disc is the Zodiacal cloud which is primarily found within ~ 5 au and composed of $\lesssim 1$ mm dust produced by asteroid collisions and comets (Nesvorný et al., 2010, see also Figure 1.7).

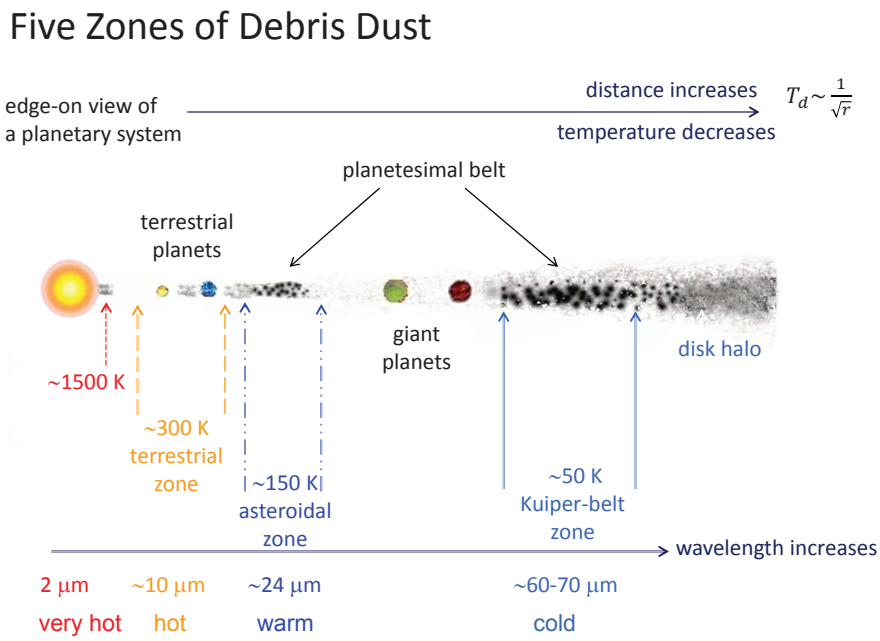

Figure 1.6. A cartoon schematic of a typical planetary system composed of both planets and a debris disc (taken from Su & Rieke, 2014). The debris disc has been categorised into five different zones as a function of observational wavelength.

Debris discs are the natural by-products of star and planet formation processes and hence are expected to be ubiquitous around main-sequence stars. Indeed, according to recent observational surveys, debris discs have been detected around at least $\sim 20\%$ of main-sequence stars in the Solar neighbourhood (Montesinos et al., 2016; Sibthorpe et al., 2018). They typically have radii ranging from a few to hundreds of au, and contain $\lesssim 1 M_\oplus$ in emitting small (i.e., $\lesssim$ mm-sized) grains. As in the Solar System, extrasolar debris discs are comprised not only of single-sized bodies, but rather a range of sizes ranging from (sub-)μm all the way up to ~km sizes – which cannot be detected with current technologies. As such, estimates of debris disc masses are uncertain, as they require extrapolating the masses of the observed dust to the masses of the unobservable planetesimals (Krivov & Wyatt, 2021). Nevertheless, it is expected that the total disc mass is larger than the observed dust mass by several orders of magnitude, with estimates on the order of $\sim 10 - 100 M_\oplus$ (see Wyatt & Dent, 2002; Krivov et al., 2018; Krivov & Wyatt, 2021, and Section 2.6). As such, extrasolar debris discs that are detected thus far are much more massive than the asteroid and Kuiper belts, which weigh $\sim 5 \times 10^{-4} M_\oplus$ (DeMeo & Carry, 2013) and $\sim 0.1 M_\oplus$ (Gladman

Figure 1.7. The zodiacal light as seen from the summit of Mauna Kea, Hawaii (Source: Courtesy M. Ishiguro, ISAS.).

et al., 2001), respectively. Indeed, if the latter were situated around another star, they would have not been detected using current technologies (Vitense et al., 2012).

In general, debris discs are qualitatively different from protoplanetary discs in various and important ways. First, while protoplanetary dust is primordial and considered as the ingredient for the formation of larger bodies, the dusty component of debris discs is secondary, in the sense that it is the result of the fragmentation and breakup of large ∼km-sized planetesimals (e.g., Backman & Paresce, 1993; Wyatt, 2008). Second, and unlike protoplanetary discs, debris discs are generally poor in gas, with the majority containing only tenuous amounts of it (e.g., Hughes et al., 2018; Wyatt, 2020). Another distinction between the two is the total amount of dust that they harbour: protoplanetary discs are seen to have large amounts of ∼mm-sized dust, with a total mass of $\sim 30 - 300 M_{\oplus}$, while debris discs generally contain only $\lesssim 1 M_{\oplus}$ of dust (e.g., Panić et al., 2013). This is illustrated by Figure 1.3, which shows that the circumstellar dust masses derived from millimetre observations drop abruptly for systems that are older than ~ 10 Myr, i.e., by the time that the protoplanetary phase is expected to have ended and transitioned to the debris phase. It is, however, worth

noting that the distinction between these two phases has become less clear in recent years (see Wyatt et al., 2015). This is because there are some protoplanetary discs with ages much longer than 10 Myr (Scicluna et al., 2014), and a few debris discs which are younger than 10 Myr (e.g., Carpenter et al., 2009b; Riviere-Marichalar et al., 2013; Lovell et al., 2021). Additionally, there are a few debris discs – particularly, around $\sim 10-40$ Myr A stars – which contain significant amounts of gas (e.g., Kóspál et al., 2013; Moór et al., 2017), somewhat comparable to those found in protoplanetary discs. As such, it is possible that there is an intermediate stage, referred to as hybrid or transitional discs, that bridges the connection between protoplanetary and debris phases (see also Michel et al., 2021).

Since the first detection of a debris disc around Vega by Aumann et al. (1984), significant advances in observational astronomy – from optical to radio wavelengths – have enabled resolved imaging of debris discs. High-resolution imaging of debris discs has revealed a rich variety of (sub-)structures: gaps or double-ringed structures, warps, spirals, and eccentric rings – see e.g., Figures 2.2, 2.5, 2.6, and Section 2.3. Studying such structures provides unique information on the various physical processes affecting the evolution of debris discs, the nature of the planetesimals (e.g., composition and sizes), and the connection of debris discs with planet formation (e.g., Krivov, 2010; Hughes et al., 2018; Wyatt, 2018). For instance, much of our understanding of the Solar System's formation and evolutionary history has come from decoding the structure of our asteroid and Kuiper belts (e.g., Heppenheimer, 1980; Malhotra, 1993; Ida et al., 2000; Moro-Martín & Malhotra, 2002; Batygin et al., 2019).

1.2.5 Planetary System Evolution

After the formation of a planetary system, a diverse array of dynamical evolution can ensue, reshaping its architecture. This is because planets and debris discs are closely linked, and can have significant interactions which could affect both the orbital evolution of planets and the morphology of debris discs[1].

Evidence of planet–disc interactions can be found in the Solar System. For instance, according to the Nice model – whose name comes after the French city where it has been developed – the giant planets underwent major orbital evolution early in their lifetimes, i.e., after the protoplanetary disc dissipated, giving rise to many of the present-day architectural features of the Solar System (see Gomes et al., 2005; Morbidelli et al., 2005; Tsiganis et al., 2005; Levison et al., 2008, from which the following is summarised).

[1] In a broader context, qualitative changes to the architecture of a planetary system can also be caused by stellar flybys in stellar clusters or dynamical interactions with binaries (if any).

Indeed, according to this model, the four giant planets were initially on circular orbits in the region from ~ 5.5 to ~ 15 au, i.e., much more closely spaced and compact than today, with Uranus orbiting exterior to Neptune. The giant planets were also surrounded by a massive debris disc ($\sim 30 - 50 M_{\oplus}$) of rocky and icy planetesimals – the primordial Kuiper belt – which extended from Uranus's initial orbit out to $30 - 35$ au. As the giant planets interacted with the debris disc, Saturn, Neptune, and Uranus gradually migrated outwards whilst Jupiter migrated inwards. After several hundreds of millions of years, this migration caused Jupiter and Saturn to enter into an orbital resonance (to be specific, $1:2$ mean-motion resonance), which pumped their orbital eccentricities. This destabilised the entire planetary system, causing a re-arrangement of the giant planets. Indeed, Saturn moved away from the Sun to its present location, and Uranus and Neptune exchanged their places which, as a result of gravitational encounters with Saturn, were put onto much wider and elliptical orbits. As a result, the ice giants started to cross the population of Kuiper belt objects along their orbits, which they interacted with and scattered onto highly eccentric orbits all over the Solar System. This process also led to the circularisation of the orbits of Uranus and Neptune and – to a lesser extent– of Jupiter and Saturn, along with the depletion of around $\sim 99\%$ of the primordial Kuiper belt, leaving behind the modern Kuiper belt together with an orbital architecture of giant planets similar to that today. Additionally, as Neptune's orbit moved outwards, it dynamically excited the orbits of Kuiper belt objects, trapping many of them (such as Pluto) into orbital resonances, giving rise to the cold and hot populations of the Kuiper belt. During the course of planetary migration, some of the planetesimals were also scattered into the inner parts of the Solar System, producing a sudden influx of impact on the terrestrial planets and the Moon, a period known as the "Late Heavy Bombardment". This series of events can also explain the formation of the Oort cloud, the existence of Jupiter and Neptune trojans, along with the dynamical structure observed in the asteroid and Kuiper belts. For instance, the Kirkwood gaps in the asteroid belt – which represent narrow regions within which the number of asteroids dips significantly – are understood to be sculpted by dynamical interactions between asteroids and Jupiter. Additionally, the location of the edges of the asteroid and Kuiper belts are set by resonant interactions with Jupiter and Neptune, respectively.

Outside of the Solar System, the diversity of exoplanetary systems – both in terms of the known population of exoplanets and morphologies of debris discs – remains to be explained. Nevertheless, it is natural to think that dynamical events, in particular in the context of planet–disc interactions, may be common to all planetary systems. As such,

studying planet–disc interactions could provide unique insights into the formation and dynamical history of planetary systems. A triumph of planet–disc interaction studies in the context of exo-planetary systems is the discovery of the planet β Pic-b (Lagrange et al., 2010) based on theoretical predictions stemming from the spatial structure observed in the debris disc surrounding the star (Mouillet et al., 1997; Augereau et al., 2001). The persistent challenge for astronomers and astrophysicists working in the field of (exo-)planetary sciences, however, is that it is generally easier to detect dusty debris discs than to detect planets. This is largely because the total cross sectional area of debris discs is generally larger than that of a planet. Thus, detailed modelling of planet–disc interactions and comparing the outcomes with observed systems is essential for our overall understanding of the formation, evolution, and architecture of planetary systems; and to even reveal the presence of planets that might otherwise not be detected with current technologies. This forms a significant fraction of the work presented in this dissertation.

1.3 Aims and Outline of the Thesis

The broad theme that this dissertation aims to contribute to is our quest to understand the formation, architecture, and evolution of planetary systems. My research focuses on developing models of the interactions between planets and debris discs, with applications to the interpretation of imaged debris discs. More specific questions that the dissertation aims to answer are:

1. What is the dynamical role of disc (self-)gravity in planet– debris disc interactions?
2. What are the observational signatures of long-term interactions between planets and *massive* debris discs?
3. Can we infer the presence and evolution of planets in systems that might otherwise remain undetected?
4. What are the insights that we can draw about the masses of debris discs based on planet-disc interactions?

I address these questions using a synergistic approach involving celestial mechanics theory, numerical simulations, and applications to observed planetary systems.

The outline of this dissertation is as follows. First, Chapters 2 and 3 provide a review of the observational and theoretical material relevant to debris discs and modelling

of planet–disc interactions. More specifically, Chapter 2 first reviews the different methods for observing debris discs (Section 2.1), and discusses the various physical processes governing their dynamics and evolution (Section 2.2). It then describes the diversity of structures imaged (Section 2.3), and how they might be interpreted as a result of planet–disc interactions (Section 2.4). In the subsequent sections, a summary is presented about the incidence of debris discs and exoplanets around the same stars (Section 2.5), followed by a review of disc masses (Section 2.6), including how they are usually estimated and their implications in the context of planet–disc interactions. Then, Chapter 3 provides an overview of the basics of celestial mechanics, starting from the Keplerian problem (Section 3.1) and then moving on to perturbation theories relevant for this dissertation (Section 3.2).

Chapter 4 focuses on theoretical and numerical aspects concerning the computation of the long-term, secular effects of self-gravitating discs. This chapter begins by first presenting the challenges involved in computing the gravitational potential of discs using standard techniques of celestial mechanics (Section 4.1), and then provides an overview of the two existing methods found in the literature that are used to alleviate these problems: softened and unsoftened gravity (Section 4.2). It then presents the results of a comparative analysis between these two methods, analysing the performance of approaches using softened forms of gravity in reproducing the expected unsoftened dynamics for various disc models (Sections 4.3 – 4.5). Then, the main findings of this work are discussed (Section 4.6), including how they can be understood within the context of a generalised Laplace–Lagrange theory for discs (which is formulated in Appendix A), and their implications for numerical treatments of self-gravitating discs. Finally, the chapter concludes with a summary of the key findings (Section 4.7).

The subsequent two chapters – i.e., Chapters 5 and 6 – examine the role of disc gravity in secular planet–debris disc interactions. Both chapters focus on what is arguably the simplest setup: a host star orbited by a single eccentric planet and a massive, external debris disc. Chapter 5 first presents a simplified analytic model that accounts for both the gravitational coupling between the disc and the planet, as well as the disc self-gravity – with the limitation that it ignores the non-axisymmetric component of the disc (self-)gravity (Section 5.2). This is achieved using the unsoftened treatment of disc gravity studied in Chapter 4. The chapter then continues by describing the essential features of planetesimal dynamics in the combined planet–disc potential (Section 5.3), and proposes a mechanism based on secular resonances as the origin of gapped structures observed in several debris discs. It then explores the characteristics of such secular resonances over a broad region of parameter space (Section 5.4), and

provides an exemplary application to a specific system (namely, HD 107146) with the aim of identifying the planet–disc parameters that could explain the observations (Section 5.5). In the subsequent sections, the main results obtained for the evolution of the disc morphology using some of these parameters is presented (Section 5.6), followed by a discussion of the results and their implications (Section 5.7), both from theoretical and observational point of views. Finally, after assessing some of the model limitations (Section 5.8), the chapter concludes with a summary of the key results (Section 5.9).

Chapter 6 expands the investigation in the previous chapter by incorporating the full secular gravitational potential of the disc, i.e., both axi- and non-axisymmetric components. This is achieved by the aid of a numerical tool referred to as the N-ring model, which is developed here based on the results of Chapter 4 for softened forms of disc gravity. The chapter begins by first presenting the details of the N-ring model (Section 6.2), and then reports on the results of several performance tests, including its ability to reproduce the analytical results of Chapter 5. Owing to the more complex nature of the problem, the chapter then proceeds to describe the results of N-ring simulations when all but the non-axisymmetric component of the disc self-gravity is accounted for (Section 6.3). These results are discussed and analysed in the subsequent section (Section 6.4), where quantitative explanations for the various phenomena observed, such as the circularisation of the planetary orbit, is developed using dynamical theory. Following this, the results obtained by accounting for all gravitational perturbations between the planet and the disc are presented (Section 6.5), including a detailed discussion of the two main outcomes found in terms of the system evolution. Finally, the chapter closes with an overall discussion and conclusion (Section 6.6), where the limitations of the study and directions for future work are also presented.

As an epilogue, Chapter 7 summarises the work presented in this dissertation and discusses potential future directions of research.

Most of these Chapters also contain supplementary material, which are provided in the form of Appendices at the back end of the dissertation. These may be skipped at first reading. Additionally, animated versions of some of the figures in Chapters 5 and 6 have been made available online, as explicitly mentioned in the corresponding figure captions.

Chapter 2

Debris Discs: An Overview

Debris discs are ubiquitous around main-sequence stars, with current detection rates of $\sim 20\%$ in the solar neighbourhood (Montesinos et al., 2016; Sibthorpe et al., 2018). They are optically thin, almost devoid of gas, and are composed of objects ranging from micron-sized dust grains up to kilometre-sized planetesimals. Since the dust grains are short-lived compared to the stellar age (e.g., Dominik & Decin, 2003), their sustained presence requires a massive reservoir of large planetesimals continually supplying fresh dust via mutual collisions (Backman & Paresce, 1993). Observed discs typically contain $0.01 - 1 M_{\oplus}$ in millimetre/centrimetre-sized grains (Wyatt et al., 2003; Holland et al., 2017), which, when extrapolated, yields masses of $\sim 1 - 100 M_{\oplus}$ for the parent planetesimal population (e.g., Wyatt & Dent, 2002; Greaves et al., 2005; Krivov & Wyatt, 2021). The spatial distribution of these planetesimals is probed indirectly with observations at millimetre wavelengths, e.g., by ALMA. At such wavelengths, observations trace the distribution of millimetre-sized dust that are largely insensitive to radiation forces, thus serving as proxy for the distribution of parent planetesimals.

Recent high-resolution observations of debris discs by ALMA and direct imaging have revealed a rich variety of radial and azimuthal structures, e.g., gaps or double-ringed structures, warps, spirals, and eccentric rings (e.g., Hughes et al., 2018; Wyatt, 2018; Wyatt, 2020). Analogous to the studies of the asteroid and Kuiper belts (e.g. Malhotra, 1993; Ida et al., 2000; Moro-Martín & Malhotra, 2002; Nesvorný et al., 2010), investigating the structure of debris discs can provide unique insights into the architecture and evolution of exoplanetary systems. For instance, the presence of a giant planet around β Pictoris, dubbed as β-Pic b, was predicted based on the warp in the debris disc (Mouillet et al., 1997), and such a planet was later discovered by direct imaging (Lagrange et al., 2010). As such, modelling of disc morphology is often focused on investigating the dynamical imprints of (invoked) massive perturbers, e.g. planets

(e.g., Wyatt et al., 1999; Wyatt, 2005b; Lee & Chiang, 2016) or stellar companions (e.g., Nesvold et al., 2017), through their interactions with the debris particles.

In this chapter, I first describe the different methods for observing debris discs (Section 2.1), together with the various physical processes governing the dynamics and evolution of the constituent dust grains and planetesimals (Section 2.2). I then move on to review the diversity of imaged spatial structures (Section 2.3), the theoretical studies which show that such structures might be explained by often yet-unseen planets (Section 2.4), and then the occurrence of debris discs and exoplanets around the same stars from an observational point of view (Section 2.5). Finally, I present a review of disc masses (Section 2.6), including how they are estimated based on observations of dust and their implications in the context of modelling planet–disc interactions.

2.1 Observational Methods

Debris discs are detected through observations of the stellar radiation which is either scattered, or absorbed and re-emitted by the constituent smallest dust grains (typically, $\lesssim 1$ mm in size). Here, I first review the different methods for observing debris discs (Sections 2.1.1 and 2.1.2), and then discuss the importance of observations at multiple wavelengths (Section 2.1.3).

2.1.1 Photometry and Spectral Energy Distributions

The earliest and most efficient method of detecting a debris disc has been through infrared photometry by measuring the Spectral Energy Distributions (SEDs) of stars – i.e., the stellar flux as a function of wavelength. Essentially, when and if circumstellar dust is present, it would absorb the incident stellar light and then re-emit at longer wavelengths that depend on the dust temperature T, or radial location r. For instance, assuming that dust grains can be approximated as black body spheres, the peak wavelength of the emission can be approximated by Wien's displacement law, $\lambda_{\rm peak} = 2800 \mu m/T$, where the temperature (in Kelvins) depends both on the dust location and the star's luminosity L_c via

$$T(r) = 278.3 \left(\frac{L_c}{L_\odot}\right)^{1/4} \left(\frac{r}{\rm au}\right)^{-1/2}. \tag{2.1}$$

As a result, the SED of the system would show excess emission flux in either the infrared or sub-millimetre (corresponding to dust at several to tens of au) when compared to

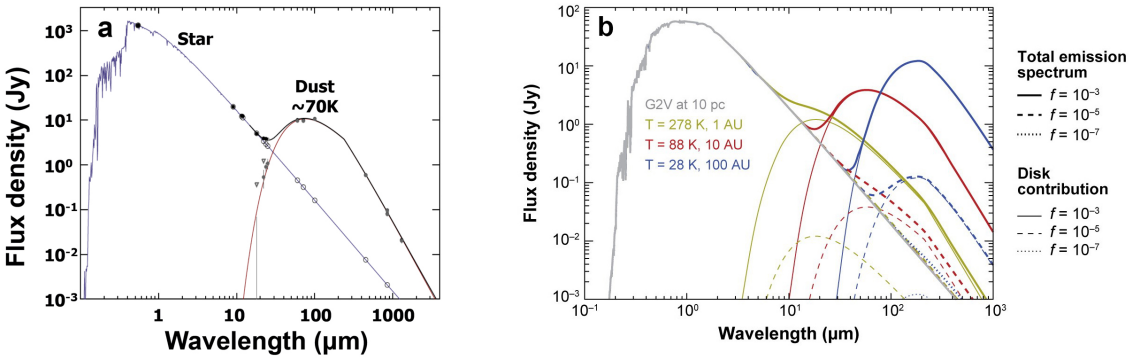

Figure 2.1. The Spectral Energy Distribution (SED) of a real star and a hypothetical star. *Panel a*: The SED of the star Fomalhaut A (taken from Wyatt, 2020). The emission at $\sim 0.1 - 10\mu$m is dominated by the stellar photosphere, while the excess emission at $\sim 100\mu$m is indicative of a dusty debris disc surrounding the star (see also Figure 2.2). *Panel b*: Synthetic SED of a G2V star located at 10 pc with a debris disc (taken from Wyatt, 2008). Calculations assume different dust temperatures assuming blackbody dust located at 1, 10, and 100 au from the star (see Equation (2.1)), as shown by different colours. The solid, dashed, and dotted lines represent discs with different levels of fractional luminosities: $10^{-3}, 10^{-5}$, and 10^{-7}, respectively. The total emission spectrum is represented by the thick lines, while the contribution of the disc alone is shown by the thin lines.

that expected from the photospheric stellar emission alone – this is usually referred to as "infrared excess" in the literature (e.g., see reviews by Krivov, 2010; Wyatt, 2020). The excess flux also depends on the amount of dust and as such is further characterized by the fractional dust luminosity $f_{\rm IR}$, i.e., the ratio of the luminosity from the dust to that of the star, with typical values ranging from 10^{-6} to 10^{-3} (Wyatt, 2020). Figure 2.1 shows an example of such an infrared excess as observed in the Fomalhaut system, along with some synthetic SEDs for various values of $T(r)$ and $f_{\rm IR}$.

Historically, the first ever detection of a debris disc around a main-sequence star, namely that around Vega, was made through observations of such infrared excesses by the Infrared Astronomical Satellite (IRAS) (Aumann et al., 1984). Since then, over a thousand of debris discs have been detected by surveys done by e.g., IRAS (Oudmaijer et al., 1992; Mannings & Barlow, 1998), the Infrared Space Observatory (ISO) (Kessler et al., 1996; Habing et al., 2001; de Muizon, 2005), the Spitzer Space Telescope (Werner et al., 2004; Meyer, 2006; Su et al., 2006; Carpenter et al., 2009a), and the Herschel Space Observatory (Pilbratt et al., 2010; Eiroa et al., 2013; Matthews et al., 2014a; Hughes et al., 2018). Detailed modelling of SEDs has allowed us to infer and estimate various fundamental parameters of debris discs, such as their radial

location, masses of the emitting dust, as well as the size distribution of the constituent dust and their composition (e.g., Lisse et al., 2007; Lebreton et al., 2012; Olofsson et al., 2012; Ballering et al., 2014; Kennedy & Wyatt, 2014; Pawellek et al., 2014).

Finally, although it is not in the remit of this dissertation to go into the details of observational methods, it is worthwhile to note that fitting SEDs with a black body spectrum is only a first order approximation. In reality, dust grains can have different emission efficiency or albedos depending on their composition and size, and so the true properties of the debris disc could differ from those inferred from the SEDs assuming blackbody grains (e.g., Schneider et al., 2006; Booth et al., 2013; Pawellek et al., 2014). At the same time, SED modelling does not always result in a unique solution due to degeneracies between the disc geometry and optical properties of dust grains (e.g., Wyatt, 2008; Pawellek et al., 2014). One avenue for breaking such degeneracies is by resolving discs spatially (e.g., Wyatt, 2008; Pawellek, 2016), which is discussed next.

2.1.2 Imaging

For debris discs that are close and bright enough to be spatially resolved, they can be detected by direct imaging. This technique is possible not only for the thermal emission of dust particles, but also for the starlight scattered off them (provided that the stellar light is blocked by a coronagraph). Indeed, thanks to the advent of both space- and ground-based instruments – including Spitzer, HST, Herschel, GPI, and more recently ALMA – this technique has been successful across the electromagnetic spectrum, from optical through mid- and far- IR to (sub-)millimetre wavelengths (e.g., Hughes et al., 2018). A spectacular example of such multi-wavelength imaging is shown in Figure 2.2, which showcases the narrow eccentric debris disc around Fomalhaut A.

Historically, such imaging at optical wavelengths provided the first evidence that the circumstellar dust around β Pictoris – first identified through infrared excess by IRAS – is distributed in a disc and not a spherical shell as was thought previously (Smith & Terrile, 1984). As of the time of writing, there are more than a hundred debris discs which have been spatially resolved with a high signal-to-noise ratio, of which a few dozen at multiple wavelengths (e.g., Hughes et al., 2018).

While limited spatial information can be derived from SED modelling, imaging represents a better method for characterising debris disc features. For instance, it can reveal the radial extent of debris discs together with the distribution of the constituent dust particles. It can also reveal the level of dynamical excitation in the disc, for instance, through characterising the scale height of nearly edge-on discs. With the improvement of imaging techniques, it has also been possible to resolve the substructures

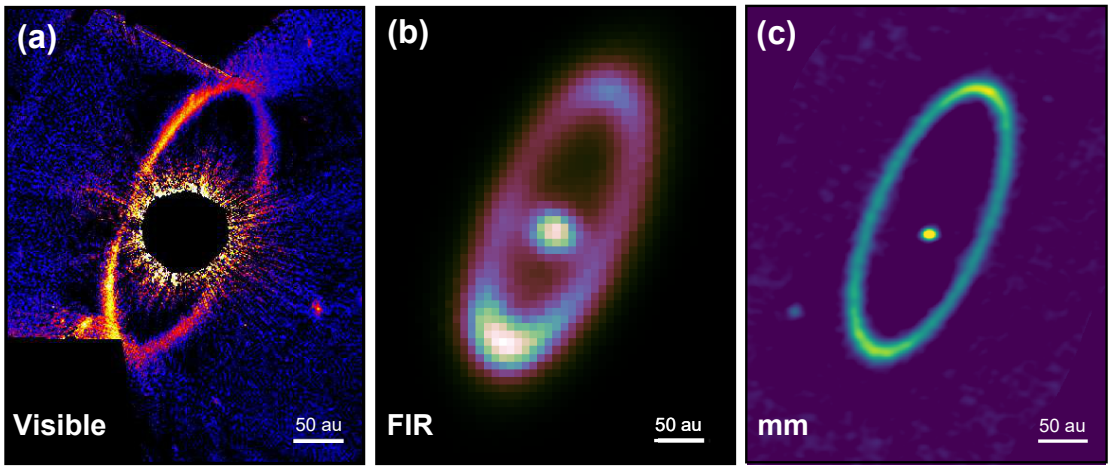

Figure 2.2. Images of the eccentric debris disc around Fomalhaut A at different wavelengths (taken from Hughes et al., 2018). *Panel (a):* in optical scattered light as seen by the Hubble Space Telescope (Kalas et al., 2013); *Panel (b):* at 70 μm as seen by the Herschel space telescope (Acke et al., 2012); *Panel (c):* the 1.3 mm continuum emission as seen by the Atacama Large Millimeter Array (MacGregor et al., 2017).

of the dust revealing a large diversity of morphologies: e.g., gaps or double-ringed structures, warps, spirals, and eccentric rings (Hughes et al., 2018; Wyatt, 2018; Wyatt, 2020). Such structures, which I will discuss in detail in Section 2.3, provide vital clues to the underlying processes as well as planetary architecture shaping the disc.

2.1.3 The Importance of Multi-wavelength Observations

It is well known that observations at different wavelengths trace dust grains of different sizes. Indeed, as a general rule of thumb, observations at a specific wavelength λ are most sensitive to dust grains of sizes s that are comparable to the observing wavelength, i.e. $\lambda \sim s$ (see e.g. Hughes et al., 2018). This can be understood as follows. First, the size distribution of grains in the disc is dominated by small particles which in turn dominate the emitting surface area or the total cross-sectional area of the disc (see Section 2.2.2). Second, the emission efficiency of dust particles is generally greater at wavelengths shorter than their physical sizes and is known to drop significantly at longer wavelengths (e.g., Draine & Lee, 1984). Thus, the total disc brightness at a given wavelength is mostly dominated by the thermal emission or scattered light of particles with sizes approximately equal to the observing wavelength.

Since particles of different sizes are affected by different type of physics (see Section 2.2), debris discs could exhibit different structures depending on the wavelength at which they are observed (Wyatt, 2006). This is particularly important as it implies

that multi-wavelength observations of the same disc, by probing different grain sizes, allows us to piece together a better understanding of the underlying physics that is shaping the disc. For instance, observations at optical wavelengths (e.g. Figure 2.2(a)) can provide insights into the spatial distribution of small (i.e., $\sim \mu$m-sized) dust grains which are strongly affected by radiation forces and stellar winds (Burns et al., 1979). On the other hand, observations at millimetre wavelengths trace larger dust grains ($\sim 0.1 - 10$ mm) for which the effects of stellar radiation and winds are negligible (see e.g. Figure 2.2(c)). Given this, such large grains are thought to be located in close proximity to the dust-producing parent planetesimals of kilometre sizes. In other words, observations at millimetre wavelengths *indirectly* probe the surface density distribution of the large planetesimals within the debris disc (e.g., Wyatt, 2006; Wilner et al., 2011). Note that bodies of centimetre size and above essentially cannot be observed at any wavelength due to their small cross-section for any plausible size distribution (e.g. Figure 2.4), rendering millimetre observations the only reliable way for probing the spatial distribution of km-sized planetesimals. This also highlights another important aspect of millimetre observations such as those with ALMA, in that they can provide insights into the gravitational forces perturbing the planetesimal orbits and could, for instance, pinpoint yet unseen planets in the system.

2.2 Physical Processes in Debris Discs

Observations of debris discs provide a unique tool to probe the formation, evolution, and architecture of exoplanetary systems. For any accurate interpretation of debris disc observations, however, a model that is based on an understanding of the physical and dynamical processes shaping the distribution of dust and planetesimals in the debris disc is necessary. Although it has been recently shown that some debris discs do contain some amount of gas (Hughes et al., 2018; Wyatt, 2020), except for a couple of debris discs, the gas is so tenuous that its dynamical effect is negligible. As a result, the evolution of particles in debris discs is mainly affected by radiation forces, collisional activity, and gravitational forces. These and other physical processes are the subject of this section.

2.2.1 Stellar Gravity and Radiation Forces

Consider a particle of mass m in orbit around a central star of mass M_c such that $m \ll M_c$. The gravitational force $\mathbf{F}_g$ due to the star acts on all particles in proportion

to their mass and distance from the star:

$$\mathbf{F}_{\mathrm{g}} = -\frac{GM_c m}{r^3}\mathbf{r}, \tag{2.2}$$

where G is the gravitational constant, and $\mathbf{r}$ is the particle's position vector such that $r = |\mathbf{r}|$ is its distance from the star. In the absence of additional forces and perturbations, particles would orbit the star on Keplerian orbits. The shape of the orbit and the position of the particle along it can be uniquely defined by a set of parameters known as the orbital elements (e.g., Murray & Dermott, 1999). These are as follows: the semimajor axis a and the orbital eccentricity e, which together define the size and shape of the orbit; the orbital inclination I and the longitude of ascending node Ω, which together dictate the orbital plane defined relative to an arbitrary reference plane; the longitude of pericentre ϖ which defines the orientation of the orbit within the plane of the orbit (relative to an arbitrary reference direction); and the true anomaly f, which defines the location of the body along its orbit. For the sake of completeness, a brief review of the Kepler problem – including orbital elements and perturbations beyond the two-body problem – is set out in Chapter 3.

The stellar gravity is the dominant force that acts on the largest bodies within the debris disc, which remain on Keplerian orbits with fixed orbital elements in the absence of other perturbing forces. The orbits of small dust grains, however, are also perturbed by non-gravitational forces resulting from their interaction with the stellar radiation field. This is because dust grains absorb and scatter the incident photons from the star, removing energy from the radiation field, which they then re-radiate.

The resulting net force experienced by a dust grain in the star's radiation field can be written as (Burns et al., 1979):

$$\mathbf{F}_{\mathrm{r.f.}} = \frac{GM_c\beta m}{r^3}\left[\mathbf{r} - \frac{\mathbf{r}\cdot\mathbf{v}}{c}\frac{\mathbf{r}}{r} - \frac{r\mathbf{v}}{c}\right], \tag{2.3}$$

where $\mathbf{v}$ is the velocity vector of the particle, and c is the speed of light. In Equation (2.3), the parameter β – the ratio of the force of stellar radiation pressure to that of stellar gravity – depends on the particle's size, density, and optical properties as well as the star's luminosity and mass; see below for details. Note that the radiation force (Equation 2.3) can be split into two parts, which are usually treated separately: (i) a velocity-independent radial part, which is known as radiation pressure; and (ii) a velocity-dependent part, which is known as Poynting–Robertson drag (or P–R drag). Both of these are discussed next.

2.2.1.1 Radiation Pressure

The radiation pressure force is determined by the velocity-independent terms in Equation (2.3), i.e., $\mathbf{F}_{\rm rad} = \frac{GM_c\beta m}{r^3}\mathbf{r}$. Similar to stellar gravity, it is inversely proportional to the square of a particle's distance from the star, however, it points radially away from the star. Thus, radiation pressure dampens the stellar gravity so that particles experience an effective gravitational force of $\mathbf{F}_{\rm eff} = \mathbf{F}_{\rm g}(1-\beta)$. That is, particles experience the gravity of an "effective star" with a mass reduced by a factor of $1-\beta$.

It is important to note that, since both radiation pressure and stellar gravity forces fall off as r^{-2}, the parameter β is a distance-independent quantity. It is given by (Burns et al., 1979; Gustafson, 1994):

$$\beta(D) \equiv \frac{|\mathbf{F}_{\rm rad}|}{|\mathbf{F}_{\rm g}|} = \left(\frac{\sigma}{m}\right)\left(\frac{L_c\langle Q_{\rm pr}\rangle_\lambda}{4\pi cGM_c}\right) \approx 0.42\left(\frac{L_c}{L_\odot}\right)\left(\frac{M_\odot}{M_c}\right)\left(\frac{2.7{\rm gcm}^{-3}}{\rho}\right)\left(\frac{1\mu{\rm m}}{D}\right), \quad (2.4)$$

where L_c is the star's luminosity, σ is the particle's cross-sectional area, $\langle Q_{\rm pr}\rangle_\lambda$ its Mie scattering coefficient averaged over all wavelengths[1], and the numerical estimate assumes that particles are perfect absorbers (i.e., a black body) and are spherical of diameter D and bulk density ρ.

Equation (2.4) shows that $\beta \propto 1/D$, meaning that the smaller a particle, the more effective radiation pressure is[2]. Thus, particles sharing the same orbital semimajor axis but differing in sizes will have different orbital periods since, when $\beta \neq 0$, $n = [GM_c(1-\beta)/a^3]^{1/2}$. This also means that when dust grains are produced, for instance, by the fragmentation of a large parent body with $\beta \approx 0$, their orbits differ from that of the parent body depending on their sizes. For a parent body which fragments at a true anomaly f along its orbit defined by semimajor axis a and eccentricity e, the daughter fragments will move in the same orbital plane as the parent body – i.e., $I^{'} = I$ and $\Omega^{'} = \Omega$ – but on orbits with semimajor axis $a^{'}$, eccentricity $e^{'}$, and longitude of pericentre $\varpi^{'}$ given by (Burns et al., 1979; Wyatt et al., 1999):

$$a^{'} = a(1-\beta)(1-e^2)[1-e^2-2\beta(1+e\cos f)]^{-1}, \quad (2.5)$$

$$e^{'} = (1-\beta)^{-1}\sqrt{e^2+2\beta e\cos f+\beta^2}, \quad (2.6)$$

$$\varpi^{'} - \varpi = f - f^{'} = \arctan[\beta\sin f/(\beta\cos f+e)]. \quad (2.7)$$

[1]The value of $\langle Q_{\rm pr}\rangle_\lambda$ ranges from 0 for perfect transmitters to 2 for perfect backscatterers. For perfect absorbers (i.e., black bodies), $\langle Q_{\rm pr}\rangle_\lambda = 1$.

[2]This is a good approximation for particles of $\gtrsim \mu$m sizes, below which β levels off and approaches a finite asymptotic value (e.g., Gustafson, 1994, and Figure 3 therein.)

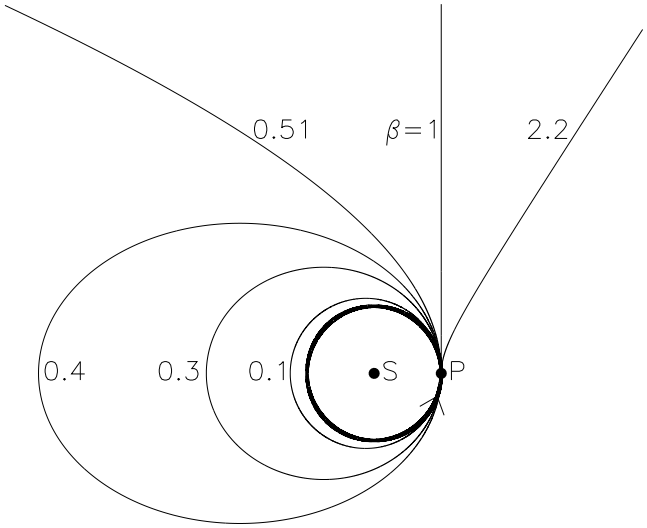

Figure 2.3. The orbits of dust particles (denoted by thin lines) produced as a result of the fragmentation of a large planetesimal P on a circular orbit (denoted by the thick line) around the star S. The dust particles have different physical sizes and so radiation β (Equation (2.4)), the values of which are shown in the figure. There are three types of orbits: (i) particles with $0 < \beta < 0.5$ are on bound eccentric orbits; (ii) those with $0.5 < \beta < 1$ are on unbound hyperbolic orbits; (iii) while those with $\beta > 1$ are on unbound, anomalous hyperbolic orbits – see the text (Section 2.2.1.1) for details. Figure taken from Wyatt et al. (1999).

Note that the pericentric distances of daughter fragments, $r_p^{'} = a^{'}(1 - e^{'})$, do not differ from that of the parent body, $r_p = a(1 - e)$, since $r_p^{'}/r_p = 1 + e(1 - \cos f) + \mathcal{O}(e^2)$ irrespective of β; see Equations (2.5) and (2.6).

Equations (2.5) – (2.7) indicate that there are three outcomes for the orbits of dust particles depending on their physical size or β. For illustrative purposes, here I follow Wyatt et al. (1999) and describe the possible outcomes for particles released by a parent body on a circular orbit ($e = 0$); see also Figure 2.3. First, if $0 < \beta < 0.5$, dust particles are released into bound elliptic orbits, $0 < e^{'} < 1$, with semimajor axes larger than that of the parent body ($a^{'} > a$). Note that particles for which $\beta \lesssim 0.1$, the effect of radiation pressure is negligible. Second, if $0.5 < \beta < 1$, then dust particle orbits are unbound and hyperbolic with $e^{'} > 1$ (or parabolic if $\beta = 0.5$) . Note that in general, i.e., for parent bodies on eccentric orbits, the boundary between bound and unbound orbits is defined by $\beta = 0.5(1 - e^2)/(1 + e\cos f)$, see Equation (2.6). Third, and finally, particles with $\beta > 1$ experience a repelling effective gravitational force and so move on anomalous hyperbolic orbits which open outward.

Note that particles with $\beta > 0.5$ will be blown out of the system on timescales comparable to the orbital period of the parent body. It is thus expected that debris discs will not contain dust grains smaller than a characteristic size $D_{\rm bl}$ corresponding

to $\beta \approx 0.5$. This is known as the *blowout size* which is given by (Wyatt, 2008):

$$D_{\rm bl} \approx 0.84 \mu m \left(\frac{L_c}{L_\odot}\right)\left(\frac{M_c}{M_\odot}\right)^{-1}\left(\frac{\rho}{2.7 {\rm gcm}^{-3}}\right)^{-1}, \tag{2.8}$$

see Equation (2.4). Finally, I remark that the minimum size of grains inferred from observations is often larger than $D_{\rm bl}$, see Pawellek & Krivov (2015) for further details.

2.2.1.2 Poynting–Robertson (P-R) Drag

The Poynting–Robertson (P-R) drag force is determined by the velocity-dependent terms of the radiation force $\mathbf{F}_{\rm r.f.}$, i.e., the second and third terms in Equation (2.3). Given the dissipative nature of this force, particles subject to it lose orbital energy and angular momentum. This causes particles to spiral in toward the star while at the same time circularising their orbits. The induced migration and circularisation rates depend on the particle size (or β) which, to lowest order in eccentricities, read as (Burns et al., 1979):

$$\frac{da_{\rm PR}}{dt} = -\frac{2\alpha(\beta)}{a} + \mathcal{O}(e^2), \quad \frac{de_{\rm PR}}{dt} = -\frac{5\alpha(\beta)}{2a^2}e + \mathcal{O}(e^2), \tag{2.9}$$

where $\alpha = 6.24 \times 10^{-4}(M_c/M_\odot)\beta$ au^2.yr^{-1} (Wyatt et al., 1999). Note that P-R drag does not affect the particle's orbital plane nor its longitude of pericentre, i.e., $dI_{\rm PR}/dt = d\Omega_{\rm PR}/dt = d\varpi_{\rm PR}/dt = 0$.

Equation (2.9) implies that P-R drag acts on relatively short time scales. For a particle initially on a circular orbit at a distance r from the star, the time $t_{\rm PR}$ taken for it to spiral into the star is on the order of thousand of years (Wyatt, 2005a),

$$t_{\rm PR} = 400 \text{ yr } \beta^{-1}\left(\frac{M_c}{M_\odot}\right)^{-1}\left(\frac{r}{\rm au}\right)^2. \tag{2.10}$$

Thus, P-R drag is another mechanism which could remove dust particles from the system either through accretion by the central star or through sublimation close to it (assuming they are not e.g. trapped by a planet or destroyed by collisional processes as they spiral in; see e.g. Wyatt (2008), Wyatt (2020), and references therein).

Evidence of P-R drag is found in the solar system e.g., in the zodiacal cloud around Earth's orbit (Figure 1.7), which is composed of $\lesssim 1$ mm dust created in the asteroid belt and dragged in towards the Sun (Dermott et al., 1994, 2001). In the case of exoplanetary systems, however, it has been argued that P-R drag is normally

insignificant when modelling known debris discs (Wyatt, 2005a). This is because for debris discs that are bright enough to be detected, the mean time of collisions between small dust grains is shorter than their P-R drag lifetime; so much so that they are processed through collisions (until they reach the blowout size, Equation 2.8) before P-R drag becomes significant – see Figure 1 in Wyatt (2005a). It is worth noting that since the balance of P-R drag and collisions is not perfect, some dust grains do get dragged into the inner parts of a system and produce excess emission at mid- to near-infrared wavelengths near the star (if not ejected or scattered by intervening planets, e.g., Wyatt, 2005a; Kennedy & Piette, 2015; Rigley & Wyatt, 2020). With the advent of new instruments over the last decade, such "exo-zodiacal dust" has now been observed at small distances from several nearby stars (e.g. Absil et al., 2006; Kennedy & Wyatt, 2013; Ertel et al., 2014, 2020)

2.2.2 Collisional Evolution

The discussion in Section 2.2.1 implies that as a result of radiation forces, small dust grains are removed from the system in relatively short timescales compared to the ages of debris discs (Dominik & Decin, 2003). Thus, dust grains in debris discs cannot be primordial or relics from the protoplanetary disc phase, but rather must be continually replenished. The accepted theoretical picture is that a "collisional cascade" initiated by a reservoir of large planetesimals – which are leftover from the planet formation process – continually supplies fresh dust (Backman & Paresce, 1993; Wyatt & Dent, 2002; Wyatt, 2008). Within this picture, km-sized planetesimals collide and produce many smaller objects that then collide and break up into smaller objects, and so on – until the blowout size (Equation 2.8) is reached.

For a collisional cascade to be ignited, the orbits of parent planetesimals must be dynamically excited – or "stirred" – so that relative velocities are high enough for collisions to be destructive. The stirring mechanism is not well constrained. For instance, it is possible that debris discs are *pre-stirred* by processes acting during the protoplanetary disc phase (Matthews et al., 2014a; Wyatt, 2008). Another possibility is that debris discs could be *self-stirred* by the gravitational perturbations of Pluto-sized ($\sim$ 1000 km) planetesimals within the disc (Kenyon & Bromley, 2008; Kennedy & Wyatt, 2010). Alternatively, the disc could be stirred through long-term secular interactions with planets (or stellar companions) in the system – referred to as *planet stirring* (Mustill & Wyatt, 2009).

The collisional evolution after the disc is stirred one way or another, however, has been studied in detail. It is well known that the size distribution of particles resulting

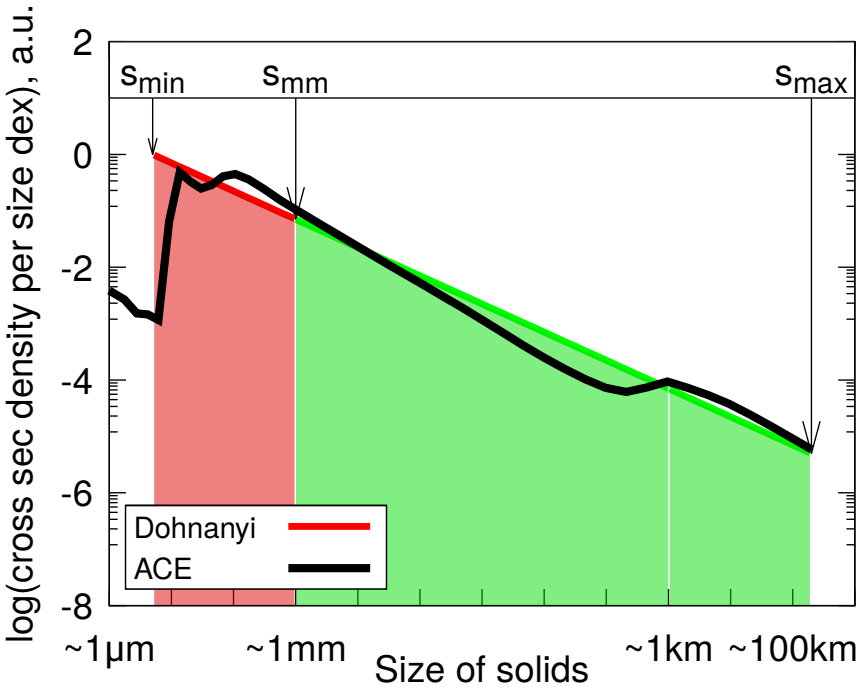

Figure 2.4. A schematic illustrating the size distribution of particles in a debris disc, as plotted in the space of cross section per unit size decade and particle radius (taken from Krivov & Wyatt, 2021). The red-green coloured line corresponds to the Dohnanyi size distribution with a slope of $\Theta = 3.5$ (Equation 2.11). For comparison, a typical size distribution obtained from collisional simulations (using the ACE code, Krivov et al., 2006) is also shown (black line). The red-filled area represents the range of dust sizes that can be observed.

from a collisional cascade can be well described by a power-law profile (e.g., Wyatt, 2008, and references therein),

$$n(D) \propto D^{-\Theta} \tag{2.11}$$

for $D_{\rm min} \leq D \leq D_{\rm max}$ and $n(D) = 0$ otherwise, where the index Θ reads as $\Theta = 3.5$ for a steady-state collisional cascade with particles of equal strengths and velocities (Dohnanyi, 1969). At the small-size end, the distribution is typically cut off by the blowout size (i.e., $D_{\rm min} \approx D_{\rm bl}$, Equation 2.8) where particles get removed from the system by radiation forces; while, at the large end, by the size of the largest body (typically, $D_{\rm max} \sim 1 - 100$ km) that has experienced a destructive collision (Dominik & Decin, 2003; Wyatt et al., 2007). Numerical simulations of collisional cascades support this general behaviour, although with slight departures from a pure power-law; see Figure 2.4 which is taken from a recent study by Krivov & Wyatt (2021).

Equation (2.11) implies that small dust particles dominate the total cross section of the disc, whilst the largest planetesimals dominate the total disc mass – as long

as $3 \leq \Theta \leq 4$. Indeed, this is supported by observations of several discs from which values of Θ within the $\approx 3-4$ range have been inferred (see Norfolk et al., 2021, and references therein). It is also interesting to note that more complicated treatments of collisional cascades than that presented by Dohnanyi (1969) predict Θ to be between 3 and 4 (e.g., O'Brien & Greenberg, 2005; Pan & Sari, 2005; Krivov et al., 2006; Thébault & Augereau, 2007; Gáspár et al., 2012; Pan & Schlichting, 2012; Schüppler et al., 2015), depending on the adopted assumptions e.g. about the material strength and velocity dispersion of bodies, and the physics involved (see also Belyaev & Rafikov, 2011, for non-power law behaviour).

Another implication of collisional evolution is that as mass is passed from the largest planetesimals to smaller dust grains (which are then lost), the total disc mass gradually depletes in time; usually on $\sim$ Gyr timescales (Dominik & Decin, 2003; Wyatt et al., 2007). Thus, the initial mass of a debris disc $M_d(0)$ is usually much larger than its present value $M_d(t)$. For instance, assuming a steady-state collisional cascade of a pre-stirred disc, $M_d(t)$ can be written as (Dominik & Decin, 2003; Wyatt et al., 2007)

$$M_d(t) = \frac{M_d(0)}{1 + t/\tau_0}, \tag{2.12}$$

where τ_0 is the collisional lifetime of the largest planetesimals in the disc at the initial epoch. Equation (2.12) implies that $M_d \propto t^{-1}$ at late times (see also Gáspár et al. (2013) for a summary of the various decay slopes as predicted by more detailed modelling), which is in general agreement with observations; see e.g. Figure 1.3. However, it is important to note that estimating the total mass of a debris disc is difficult (Krivov & Wyatt, 2021), because it requires extrapolating the mass of the observable dust ($\lesssim 1$ mm) to the mass of the unobservable large planetesimals, let alone our limited knowledge of optical and physical properties of the disc particles. Thus, a debris disc's total mass is a very uncertain parameter. A detailed discussion of this issue is provided in Section 2.6.

2.2.3 Gravitational Interactions

Dust grains and planetesimals comprising debris discs are also influenced by the gravitational perturbations induced by any massive objects in the system, apart from the central star. Generally, such perturbations could result from 'external' massive sources, i.e., planets and stellar companions, and/or 'internal' sources, i.e., the collective gravity of the debris disc itself. Regardless of the perturbation origin, however, the orbits of debris particles, together with the perturbers themselves (whether those are the

planets residing in the system or the disc particles themselves), will be perturbed away from their otherwise fixed Keplerian orbits in the absence of gravitational perturbations and thus evolve in time. Indeed, particle orbits will experience variations in their orbital parameters – such as orbital eccentricities, inclinations, and longitudes of pericentre and ascending node – in ways that depend on the type of the perturbation. This will in turn affect the spatial distribution of the debris particles.

Generally, gravitational perturbations can be categorised, both mathematically and conceptually, into three distinct types (Murray & Dermott, 1999): (i) secular, long-range interactions which act on timescales much longer than the typical orbital periods; (ii) resonant, relatively short-range interactions which act on timescales comparable to the involved orbital periods; and (iii) short-period interactions, which become important during close encounters and scattering events. A detailed discussion of such perturbations – both from a mathematical and physical point of view – is provided in Chapter 3. The effects of each of these types of gravitational interactions, both separately and in combination, have been the subject of multiple studies over the last decades (e.g., see review by Krivov, 2010; Hughes et al., 2018; Wyatt, 2018), finding that they can imprint various signatures in the spatial appearance of debris discs, such as spiral arms, warps, radial gaps, etc. A detailed review of the imaged structures in debris discs and studies of planet–disc interactions aiming to explain them is provided in Sections 2.3 and 2.4, respectively.

2.2.4 Other Processes and Forces

In addition to the physical processes described in Sections 2.2.1 – 2.2.3, a number of other processes and forces may influence the spatial distribution and evolution of debris particles: for instance, sublimation, stellar wind forces, the Lorentz force, the Yarkovsky effect, and gas drag. These processes, while of no direct relevance for the debris discs studied and modelled in this dissertation, are discussed briefly below for the sake of completion.

• **Sublimation:** Icy dust grains and planetesimals can sublimate, i.e. transit from a solid state of matter to a gaseous state, in the vicinity of a star as a result of the higher temperatures in that zone. The radius of the sublimation zone depends on the physical properties of the particles (e.g., material, porosity, etc.) as well as the stellar luminosity: as a general rule of thumb, though, it ranges from ~ 1 to ~ 10 stellar radii. Thus, when particles move on orbits that bring them close to $\sim 0.01 - 0.1$ au from the central star, they sublimate and release a population of atomic and molecular gas. This can happen as a result of dust grains migrating inwards towards the star

under radiation forces or exo-comets being scattered into the inner parts of a planetary system – as is the case in the Solar System (e.g., Kimura et al., 1997; Krivov et al., 1998). Such sublimation processes assisted by radiation forces and/or scattering events have been invoked in extra-solar settings as a viable scenario for e.g. explaining the observed gaseous component of some debris discs (e.g., Beust et al., 1990; Wyatt, 2018, and references therein). Within such models, the debris disc gas is considered as having a secondary origin, rather than primary (i.e., remnant from the protoplanetary phase). More recently, it has also been shown that the sublimation of icy grains and the subsequent release of gas could potentially explain the observed near infrared excesses due to close-in, hot dust around various main-sequence stars (Pearce et al., 2020): within this context, the released gas acts as a trap, holding dust grains that enter the sublimation zone subsequently on stable orbits.

• **Stellar wind forces:** In the same way that stellar electromagnetic radiation gives rise to radiation pressure forces, $\mathbf{F}_{\text{r.f.}}$ (see Section 2.2.1), stellar particulate radiation gives rise to stellar wind forces, $\mathbf{F}_{\text{s.w.}}$ (Burns et al., 1979). This force, similar to radiation pressure forces, can be decomposed into two components: *stellar wind pressure* and *stellar wind drag*. Indeed, as a corpuscular analogue to $\mathbf{F}_{\text{r.f.}}$, the stellar wind force $\mathbf{F}_{\text{s.w.}}$ has the same form as in Equation (2.3), but with c and β replaced by $v_{\text{s.w.}}$ and $\beta_{\text{s.w.}}$, respectively. Here, $v_{\text{s.w.}}$ is the stellar wind speed and $\beta_{\text{s.w.}} = \left(\frac{\sigma}{m}\right)\frac{\dot{M}_c v_{\text{s.w.}} C_D/2}{4\pi G M_c}$, where $\dot{M}_c$ is the stellar mass loss rate, σ is the particle's cross-sectional area, and C_D – typically taken to be equal to 2 – is a measure of the momentum transfer from wind particles to a dust grain.

The effects of the stellar wind pressure is known to be negligible, since stellar winds carry much less momentum and energy than stellar radiation. However, the effects of stellar wind drag can be significant and non-negligible since $v_{\text{s.w.}} < c$. For instance, in the Solar System, with typical values of $v_{\text{s.w.}} \sim 300 - 800\text{km.s}^{-1}$ and $\dot{M}_c = 2 \times 10^{-14} M_\odot \text{yr}^{-1}$ (Allen, 1973), the strength of the stellar wind drag is $(\beta_{\text{s.w.}}/\beta)(c/v_{\text{s.w.}}) \approx 30\%$ of that of the P–R drag (Gustafson, 1994). For stars with stronger stellar winds or late type stars, it is also possible for stellar wind drag to overcome P-R drag (Reidemeister et al., 2011). Thus, in practice, when studying the evolution of dust grains in debris discs, P–R and stellar wind drag forces are usually combined together and an effective $\beta_{\text{eff.}}$ parameter is introduced.

• **Lorentz force:** Circumstellar dust grains can also be charged as a result of interacting with the incident stellar light due to e.g. the emission of photo-electrons or capture of ions and electrons (e.g., Horanyi, 1996). As a result, dust particles are also subject to the Lorentz force induced by the magnetic field of the central star (or

that of a planet). This introduces additional perturbations to the orbit of a given dust particle – an effect well-studied within the context of the Solar System (e.g., Morfill & Gruen, 1979; Consolmagno, 1980). Generally, the perturbations induced by the Lorentz force could either trap dust grains at small distances from the star – giving rise to hot exozodiacal dust (Rieke et al., 2016) – or contribute to the ejection of dust particles out of the planetary system (Hamilton et al., 1996). These effects, however, become prominent for relatively small dust grains which are typically of $\lesssim 0.1 - 1\mu$m in size, i.e., those which contribute only marginally to both the mass budget and cross sectional area of a debris disc. This is so mainly because $\lesssim 1\mu$m dust grains are relatively scant in number since they are quickly removed from the system by radiation forces (e.g., recall that $D_{\rm bl} \sim 1\mu$m for a solar-type star; Equation (2.8)). As such, the Lorentz force is usually ignored in the literature when studying debris discs[3].

• **Yarkovsky effect:** The dynamics of large ($\sim$ km) objects can also be influenced by the Yarkovsky (Radzievskii, 1952; Peterson, 1976) and the YORP effects (Radzievskii, 1952; Paddack, 1969; O'Keefe, 1976). The Yarkovsky effect arises due to the anisotropic thermal emission of photons from the surface of a rotating body: the photons impart a recoil force on the body as they leave its surface, perturbing its orbit. The YORP effect, on the other hand, results due to the torque due to the thermal emission from irregularly shaped bodies: this can increase or decrease the body's spin rate and can affect the spin axis orientation. These effects have been well-studied within the context of asteroids in the Solar System (e.g., Bottke et al., 2006, and references therein), and have also found applications within the context of debris discs, especially around white dwarfs (e.g., Veras et al., 2014).

• **Gas drag:** Particles constituting a debris disc may also be subject to gas drag, which arises due to their relative velocity with respect to the gaseous component. This is because the gas, unlike solids, is partially supported by its own pressure, and thus orbits the star with a sub-Keplerian velocity. The effects of gas drag is inversely proportional to the radius of planetesimals (Adachi et al., 1976), and so more significant for smaller particles. It generally acts to dissipate orbital energy and angular momentum, damping the orbital eccentricity and inclination of a particle while causing its orbit to decay towards the star (e.g., Marzari & Scholl, 2000). Given that most of the known debris discs are gas-poor, the effects of gas drag is usually ignored when studying them. Nevertheless, it could be important for some of the discs which have recently been found to contain a significant amount of gas (e.g. Hughes et al., 2018;

[3]An exception to this is for debris discs around young M-dwarfs (such as that around AU Mic, e.g. Chiang & Fung, 2017), known to generate the strongest magnetic fields among main-sequence stars.

Olofsson et al., 2022). For instance, Lin & Chiang (2019) showed that the drag force exerted by second-generation gas in debris discs – e.g., released by volatile-rich solids due to collisions – could force the orbits of small dust grains to apsidally align, giving rise to "moth"-like debris disc morphologies in scattered light images (e.g. Figure 2.5(h)). It is also possible that the gas drag acting during the protoplanetary phase, in combination with the perturbations due to a young planet, could shape the distribution of solids and thus leave an imprint on the structure – or at least, the initial conditions – of the descendant debris disc once the gas is dissipated (Zheng et al., 2017; Kennedy, 2020).

2.3 The Diversity of Debris Disc Structures

As already mentioned above, recent advances in imaging capabilities across multiple wavelengths have revealed a rich variety of debris disc structures at unprecedented detail including narrow and broad rings, gaps, warps, and spiral arms, to name a few[4]. In this section, I provide a brief review of the diversity of spatially resolved structures in debris discs which can be categorised into three types: radial, azimuthal, and vertical. At the outset, however, it is important to note that some disc structures could fall into more than one category, or be the result of a combination of structures coupled with viewing angle effects (e.g. Lee & Chiang, 2016). Thus, the categorisation is not unique per se. For reference, a gallery of images illustrating examples of the different disc structures is presented in Figures 2.5 and 2.6.

2.3.1 Radial Structure

• **Eccentric Rings:** Debris discs seem to have a wide range of radial extents, with some being very narrow (i.e., on the order of $\sim$ 10 au) and others being very broad (i.e., on the order of $\sim$ 100 au). A unifying theme to both, however, is that they seem to be eccentric, in the sense that their centre is offset from the central star by some distance. Indeed, measurements of debris rings' eccentricities typically yield values ranging from $\sim$ 0.05 (which is the lower limit of what can be measured by existing instrumentation) up to $\sim$ 0.2 (see e.g., Hughes et al., 2018; Kennedy, 2020, and references therein). This feature seems quite common particularly in narrow rings, which often are found to be eccentric (compared to broad rings). Canonical examples of narrow eccentric rings

[4]Catalogues of spatially resolved images of debris discs can be found online at www.circumstellardisks.org and www.astro.uni-jena.de/index.php/theory/catalog-of-resolved-debris-disks.

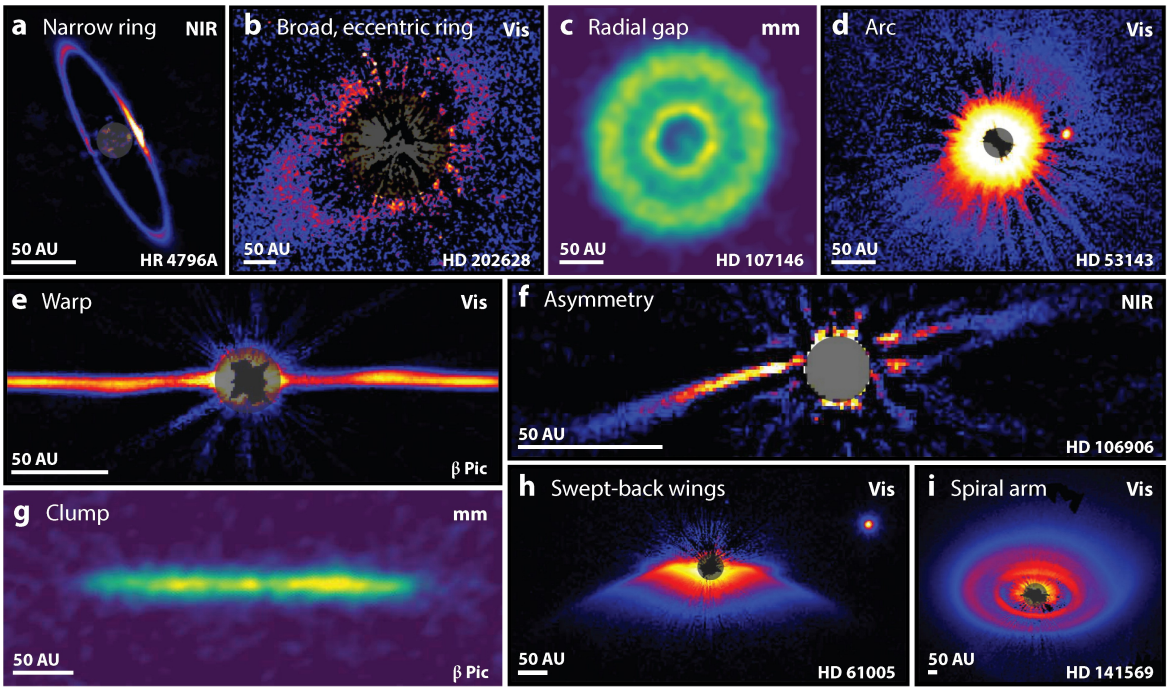

Figure 2.5. Compilation of scattered light and/or thermal emission images of debris discs around eight nearby stars (taken from Hughes et al., 2018). The collection of images showcases the diversity of observed debris disc structures: radial, azimuthal, and vertical – see the text for more details (Section 2.3).

are provided by Fomalhaut (Kalas et al., 2013, see also Figure 2.2), HR 4796A (Milli et al., 2017b; Kennedy et al., 2018, see also Figure 2.5(a)), and HD 181327 (Schneider et al., 2014); while HD 202628 on the other hand represents a good example of a broad eccentric ring (Schneider et al., 2016, see also Figure 2.5(b)).

Owing to the stellar offset, such eccentric rings are also seen to exhibit asymmetries in their surface brightness distribution, which peak either at the pericentre or apocentre side depending on the observational wavelength (Wyatt et al., 1999; Pan & Wu, 2016; Lynch & Lovell, 2021). The former case, known as pericentre glow, is observed at short wavelengths as a result of the dust at pericentre being hotter and glowing more brightly than the dust at apocentre (Wyatt et al., 1999); see e.g. panel b of Figure 2.2. The latter case, on the other hand, which is referred to as apocentre glow, is observed at long wavelengths, as they essentially trace the surface density distribution of the dust which is expected to be larger at the apocentre than at the pericentre (Pan & Wu, 2016); see e.g. Figure 2.2(c).

Finally, I point out that eccentric rings could also manifest themselves as having various morphologies, e.g. swept-back wings and asymmetries (see Sections 2.3.2 and 2.3.3), depending on the observing angle (Lee & Chiang, 2016). For instance, an

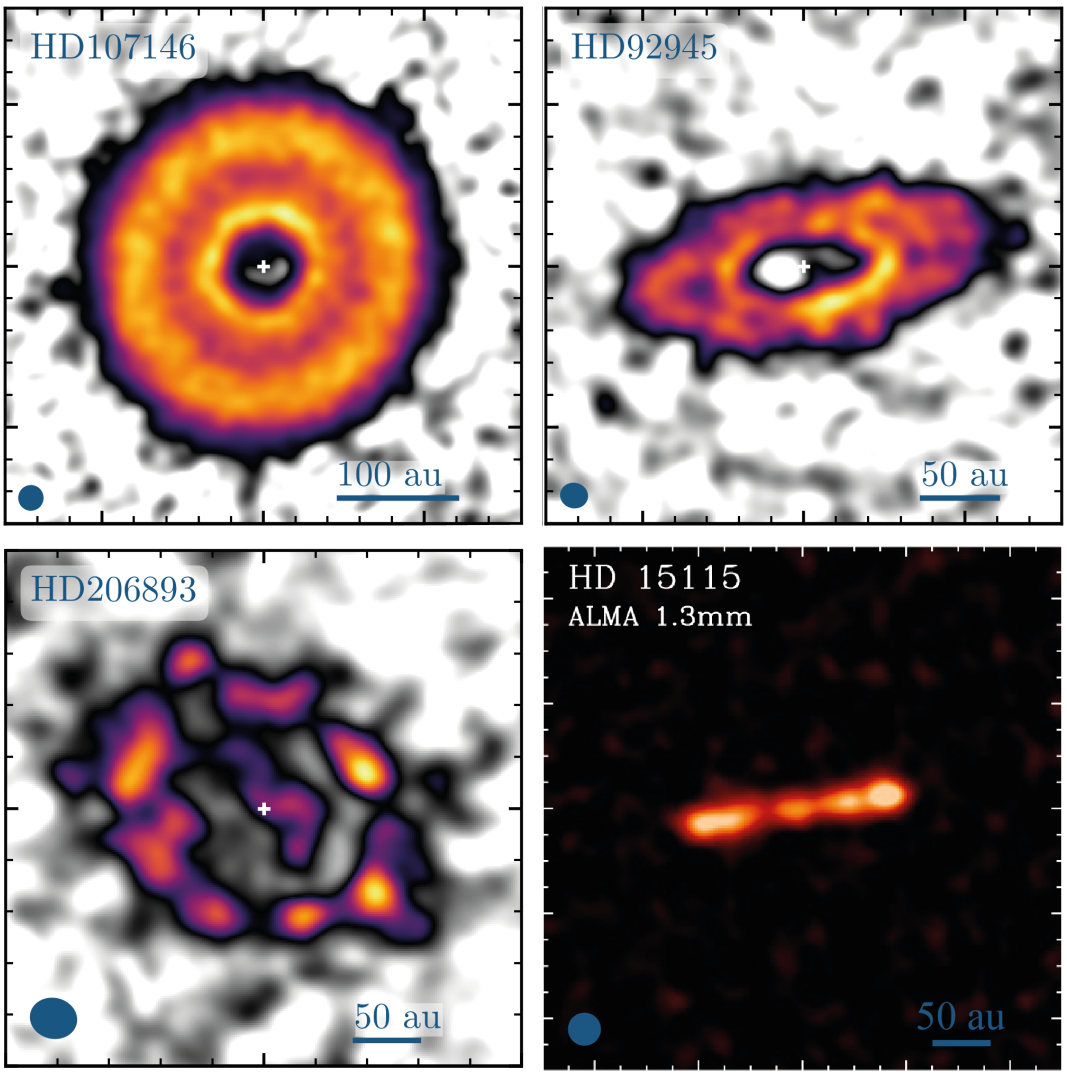

Figure 2.6. Compilation of ALMA images of the distribution of mm-sized dust in the debris discs of four nearby stars exhibiting gapped, or double-belt, structures (adapted from Marino (2021) and MacGregor et al. (2019)). The individual discs are: HD 107146 (Marino et al., 2018a), HD 92945 (Marino et al., 2019), HD 206893 (Marino et al., 2020), and HD 15115 (MacGregor et al., 2019). The resolution beam sizes and radial scales are shown by blue ellipses and lines, respectively, at the bottom of each panel.

eccentric ring seen nearly edge-on may appear to be shorter on one side compared to the other (see Figure 2.5(f) and Section 2.3.2).

• **Radial gaps:** Another common radial structure observed in debris discs is the presence of a double-belt structure, or a gap. In other words, the debris disc appears to be separated into two distinct belts, separated by a certain region in which there is

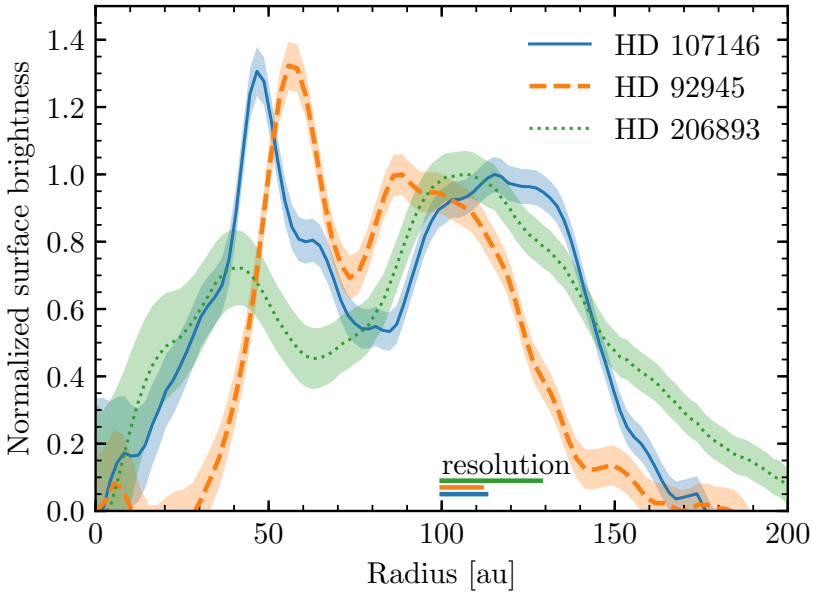

Figure 2.7. The radial profiles of the deprojected surface brightness of the gapped debris discs around HD 107146 (blue), HD 92945 (orange), and HD 206893 (green), as derived from ALMA observations (taken from Marino et al., 2020). The shaded regions represent the associated 1σ uncertainties. Note the depletion centred around ≈ 70 au in all three discs.

a void, or at least a significant underdensity, of planetesimals; see e.g. Figure 2.5(c). This is somewhat similar to the asteroid and Kuiper belts in our own Solar System, which are separated by a $\sim$ 27 au wide radial gap. Such radial gaps seem to be quite common in relatively wide (i.e., on the order of $\sim$ 100 au) debris discs, with at least 4 out of 6 wide belts that have been observed with ALMA at high enough resolution[5] showing evidence of gaps. Namely, these are the debris discs surrounding HD 107146 (Ricci et al., 2015; Marino et al., 2018a), HD 92945 (Marino et al., 2019), HD 15115 (MacGregor et al., 2019), HD 206893 (Marino et al., 2020; Nederlander et al., 2021), and potentially AU Mic (Daley et al., 2019). For reference, the images of these four gapped debris discs as obtained by ALMA are displayed in Figure 2.6. Note that perhaps somewhat coincidentally, the gaps in these four debris discs are centred at around the same radial distance, $\sim$ 70 au (Marino et al., 2020). This can be seen more clearly in Figure 2.7, which shows the deprojected surface brightness profiles of these discs (with the exception of HD 15115 which is nearly-edge on).

[5]Here, what is meant by "high enough resolution" is that the disc is observed with at least 4 beams across its width, and a signal-to-noise ratio of $S/N \gtrsim 10$ in terms of the deprojected radial profile; see e.g. Marino et al. (2020).

Radial gaps have also been detected in scattered light observations of at least 6 debris discs, including HD 92945 and HD 15115 (Golimowski et al., 2011; Schneider et al., 2014) – which already exhibit gaps at millimetre observations – but also those around HD 141569A (Perrot et al., 2016), HD 131835 (Feldt et al., 2017), HD 120326 (Bonnefoy et al., 2017), HD 141943 (Boccaletti et al., 2019), and potentially HD 38858 (Beichman et al., 2006). While these gaps may be indicative of depletion of planetesimals in that region, it is also possible that they are the result of gas-dust interactions which could be affecting the disc structure (e.g. Lyra & Kuchner, 2013), particularly for HD 141569 and HD 131835, which are known to be gas-rich (Zuckerman et al., 1995; Kral et al., 2019). Thus, the gaps observed by ALMA in the four debris discs described above are more reliable (Figure 2.6), in the sense that in those cases observations have constrained the level of gas to be minimal, and that the depletion in the millimetre dust emission could be interpreted as depletion in the distribution of large ($\sim$ km-sized) planetesimals.

Finally, I remind the reader that the work presented in the second part of this dissertation mainly focuses on explaining the origin of radial gaps. This said, more detailed information on the systems imaged by ALMA – particularly, HD 107146, HD 92945, and HD 206893– is provided in the relevant chapters.

• **Spirals:** Another type of radial structure observed in debris discs are spiral features. Indeed, spirals have been observed in various forms, ranging from tightly-wound structures to large-scale one- or two-armed structures. A prototypical example is HD 141569A which, in optical imaging, shows that the radial distribution of dust is peaked at $\sim$ 150 and $\sim$ 250 au as a result of a tightly wound spiral (Clampin et al., 2003); see Figure 2.5(i). Additionally, the more diffuse emission from distances ranging over $\sim$ 300 to 1200 au forms an open spiral structure with one, or possibly two arms (Mouillet et al., 2001; Konishi et al., 2016). Other discs characterised with spiral features are those around TWA 7 (Olofsson et al., 2018), which shows evidence of an one-armed spiral arm, and perhaps HD 92945 (Marino et al., 2019, see also the top right panel in Figure 2.6). Finally, note that depending on the viewing geometry, the appearance of spiral structures could be due to the projected view of a disc which is warped; see Section 2.3.3. In Chapters 5 and 6, I will demonstrate that spiral arms – both as a wound feature and an open one-armed one – could result from the secular interactions between planets and massive debris discs.

2.3.2 Azimuthal Structure

• **Clumps:** A common type of azimuthal structure observed in debris discs is a variation of the disc brightness as a function of the azimuthal angle, with much of the emission being concentrated in one or more clumps; see Figure 2.5(g). There are several debris discs characterised with this feature: for instance, β Pic, which exhibits a clump on one side of its debris disc as seen in millimetre observations (e.g. Dent et al., 2014); and AU Mic, which showcases several clumps distributed over a range of locations as seen in scattered light imaging (e.g. Krist et al., 2005). Other discs potentially showcasing clumpy structure are those around Vega (Wilner et al., 2002) and ϵ Eridani (Greaves et al., 2005), although follow-up observations have not yet recovered them robustly (e.g. Hughes et al., 2012; Chavez-Dagostino et al., 2016; Booth et al., 2017).

• **Arcs:** Another azimuthal structure observed in debris discs are arc-shaped features. Arcs manifest as a morphology similar to a horseshoe, whereby the emission seems to be organised in a crescent-shaped structure ranging over a limited span in terms of azimuth on either side of the disc. Good examples of this structure are seen in the debris discs around HD 53143 (Schneider et al., 2014, see panel d of Figure 2.5) and in the extremely extended halo around the HR 4796A debris disc (Schneider et al., 2018). Note that such arc-shaped structures may look like swept-back wings (or vice versa), depending on the viewing geometry.

• **Asymmetries:** Debris discs are also found to display azimuthal asymmetries, in the sense that they appear to be extended to significantly larger distances towards one side of the star compared to the other. In its extreme limit, this is also known as the "needle" structure. A prototypical example of such a structure is seen in HD 106906's debris disc (Kalas et al., 2015, see panel f of Figure 2.5), for which the asymmetries are particularly pronounced. Note, however, that it is often difficult to discern whether the two ansae of the disc are at different radii or the seemingly shorter side extends to the same distance as the longer side but has a lower level of emission that it evades detection. Additionally, such structures could also be the result of eccentric rings (see panels a and b of Figure 2.5 and Section 2.3.1), particularly when found in nearly edge-on configurations.

2.3.3 Vertical Structure

• **Warps:** Several debris discs have been observed to have vertical, warped structures. A warp represents variations of the disc's plane of symmetry as a function of radial

distance from the central star. In other words, a warped disc is characterised by having an inner part which is inclined with respect to its outer parts. Though such warps are difficult to detect in debris discs which are not nearly edge-on with respect to the line of sight, several warped debris discs have been imaged to date, such as those around β Pic (Heap et al., 2000, see also Figure 2.5(e)), AU Mic (Wang et al., 2015), and HD 110058 (Kasper et al., 2015). Note that it is possible that many more discs which are nearly face-on or at moderate inclinations are also warped. For instance, a warp has been identified in the face-on TW Hydra *protoplanetary* disc by analysing the emission spectrum, finding that a warp blocks the stellar light reaching the outer disc parts (Roberge et al., 2005). However, this method is futile when it comes to debris discs, which are optically thin (e.g. Hughes et al., 2018).

• **Swept-back wings:** Another vertical structure observed in debris discs is the so-called swept-back wings feature, also referred to as the "moth". Moths are a manifestation of bowed disc midplanes, whereby the disc midplane is bent sharply at the disc's ansae (similar to an elbow) with some deflection angle either upward or downward from the star. The best examples of this structure are provided by the debris discs around HD 61005 (see panel h of Figure 2.5) and HD 32297, as evidenced by both scattered light and millimetre observations (Schneider et al., 2014; MacGregor et al., 2018).

2.4 Debris Discs as Signposts of Planets

The question on the origins of the observed diversity of debris disc structures is still a fairly open one. Nevertheless, and primarily motivated by the studies of the Solar System and its debris discs (e.g., Heppenheimer, 1980; Malhotra, 1993; Ida et al., 2000; Moro-Martín & Malhotra, 2002, see also Section 1.2.5), the common consensus is that the majority of the observed debris disc structures are of dynamical origin, in the sense that they are sculpted by the various physical processes affecting the evolution of the dust and planetesimals in debris discs (Section 2.2); particularly, planet–disc interactions. As such, much effort has been put into understanding how planets sculpt debris disc morphologies through their gravitational interactions, both in the presence and absence of other physical processes such as collisions, radiation forces, etc (e.g., see reviews by Krivov, 2010; Hughes et al., 2018; Wyatt, 2018; Wyatt, 2020).

Of course, it is possible to argue that associating debris structures to gravitational perturbations due to planets (or companions) is overstretched, in the sense that perhaps observed debris structures are instead simply inherited from the distribution

of planetesimals during the protoplanetary phase – which typically exhibit structures very similar to those observed in debris discs (e.g. Figure 1.4 and Andrews (2020)) – and thus no planet is required. While this hypothesis might be true for some of the observed features such as the location of a debris ring and the size of its inner hole (e.g. Michel et al., 2021; Najita et al., 2021), it might not easily explain the observed non-axisymmetric features since they need to be maintained over long timescales, i.e., long after the evolution of the circumstellar disc from its protoplanetary to debris phase. Consequently, debris disc structures are often considered as signposts and indicators of planets which usually remain undetected as yet.

As mentioned in Section 2.2.3, there are essentially three ways in which planets can dynamically interact with debris discs and, in turn, sculpt them: secular, resonant, and short-period perturbations such as scattering. Here, I review some of the most notable and recent works on planet–debris disc interactions, highlighting the structures that are often associated with planets.

2.4.1 Secular Structures

Secular interactions are interactions that operate over both long timescales and distances, see Section 3.2.4.1 for details. The secular effects of planets on debris discs has been studied in detail in the literature, finding that it can explain a large variety of the observed structures in debris discs. The foundational work in this aspect has been carried out by Wyatt et al. (1999), who investigated how a single or multiple planets which are on eccentric and/or inclined orbits affect the evolution of debris particles.

For planets which are on nearly circular and coplanar orbits, it is well known that the secular perturbations cause planetesimal orbits to become eccentric and precess at rates that depend on their distance from the planet (Murray & Dermott, 1999, see also Chapter 3). This causes planetesimals that are on initially circular orbits to start crossing each other, "stirring" the disc and thus igniting a collisional cascade (Mustill & Wyatt, 2009). During this process, a spiral pattern develops in the disc which starts off at the disc edge closer to the planet and propagates radially throughout the disc with time as it wraps around the star (Wyatt, 2005b). After enough time has elapsed, i.e., so that the spiral arm is very tightly wound, the disc then becomes eccentric, exhibiting a large scale asymmetry as a result of the shifted centre of symmetry away from the central star (Wyatt et al., 1999). This in turn gives rise to brightness asymmetries which manifest as pericentre or apocentre glow, depending on the observational wavelength – similar to those observed in many discs (Section 2.3.1).

Building on the work of Wyatt et al. (1999), Lee & Chiang (2016) carried out a detailed exploration of the morphologies that ensue from the secular interactions between a single, eccentric planet and a coplanar, narrow ring of parent planetesimals. Despite the simplicity of the model, Lee & Chiang (2016) found that as a result of the secular interactions, coupled with radiation forces that act on the smaller dust produced due to collisions, a surprisingly large variety of observed morphologies can be reproduced. Indeed, depending on the viewing geometry, they find that e.g. eccentric rings (Figure 2.5(a)), swept-back wings or moths (Figure 2.5(h)), and asymmetries or needles (Figure 2.5(f)) are easily reproduced.

For planets that are on moderately inclined orbits, on the other hand, it has been shown that planetesimal orbits evolve somewhat similarly to the case of eccentric planets, but now the interaction leads to the appearance of a warp within the disc when viewed edge-on (e.g. Mouillet et al., 1997; Augereau et al., 2001; Dawson et al., 2011; Dong et al., 2020); similar to that in Figure 2.5(e). The warp is a transient feature if induced by a single inclined planet, i.e., its location changes over long (secular) timescales (e.g. Dawson et al., 2011), and may be more localised if induced by two (or more) mutually inclined planets (e.g. Wyatt et al., 1999; Dong et al., 2020). Here, it is important to note that the presence of the planet around β Pic was first predicted based on the observed warp in its debris disc and its modelling via secular planet–disc interactions (Mouillet et al., 1997; Augereau et al., 2001), and later discovered by direct imaging (Lagrange et al., 2010); see also Section 2.5.2.

The role of planetary secular perturbations in sculpting debris discs has also been studied in more extreme setups involving planets on highly eccentric and/or inclined orbits (e.g. Beust et al., 2014; Faramaz et al., 2014; Pearce & Wyatt, 2014, 2015). In such cases, it has been shown that the secular interaction can result not only in elliptical discs, but also hollow bell-shaped structures enveloping the planet's orbit as well puffed-up discs that are orthogonal to the planetary orbit.

Finally, it is important to note that since secular interactions are long-range interactions, planets – or stellar companions, for that matter – could sculpt debris discs whether they are interior or exterior to a debris disc. For instance, Nesvold et al. (2017) have shown that an external inclined companion could induce Kozai–Lidov oscillations of debris particles, resulting in disc asymmetries and eccentric structures (see also Shannon et al., 2014; Kaib et al., 2018).

To summarise, secular planet–disc interactions could explain a wide range of debris features: spirals, offsets and brightness asymmetries, warps, moths, and needles. Consequently, resolved images of such structures can be connected to the underlying

planetary system, revealing valuable information e.g. on the mass and orbital elements – i.e., semimajor axis, eccentricity, and inclination – of the planet(s) sculpting the disc.

2.4.2 Resonant Structures

Resonant interactions are interactions that operate over timescales that are generally shorter than secular interactions and are constrained to narrow radial ranges corresponding to planetesimal orbital periods comparable to that of the perturbing planet, see Section 3.2.4.2 for details. While seemingly narrower in scope compared to secular interactions, resonant planet–disc interactions have been shown to be apt in reproducing several of the observed debris structures: clumps, disc edges and gaps.

For instance, as a planet migrates inward or outward, it may resonantly interact with planetesimals which were otherwise not affected by its perturbations. As a result, in the course of the evolution, the planet may capture planetesimals inside or outside its orbit into resonant orbits which may then remain trapped, following the resonance as the planet migrates. It has been shown that this process results in the azimuthal redistribution of a swarm of planetesimals which get trapped into certain longitudes depending on the nature of the resonance, giving rise to clumpy structures (e.g., Wyatt, 2003; Reche et al., 2008, and references therein) – similar to those observed in various debris discs (Figure 2.5(g)). Given that capture into resonance is a probabilistic event depending on the planet's orbital parameters and migration rate (e.g. Murray & Dermott, 1999), clumpy structures can thus constrain not only the masses and orbital parameters of the underlying planets, but also their past migration history (similar to the studies of Neptune and the Kuiper belt, e.g. Moro-Martín & Malhotra (2002)).

Resonant planet–debris interactions have also been invoked to explain the edges and the cleared inner regions of planetesimal belts (e.g. Quillen & Faber, 2006; Chiang et al., 2009; Mustill & Wyatt, 2012); similar to those in panels a and b of Figure 2.5. This follows from the destabilising nature of mean-motion resonances that occur in the proximity of a planet (Section 3.2.4.2): since such resonances have finite width over which they operate, there is a region surrounding a planetary orbit in which the resonances overlap, driving planetesimals orbiting within that region to chaotic motion and their eventual ejection from the system over relatively short timescales (Wisdom, 1980). As such, instabilities of this type may be used to relate the imaged radial position of debris rings'/discs' inner edges to potential planets residing nearby, i.e., either interior or exterior to the debris ring, allowing one to constrain the planet's mass and semimajor axis. Additionally, since the planet also imposes some eccentricity

on the surviving planetesimals, the sharpness of the edges of debris rings may encode further information about the planetary orbit, including its eccentricity.

For single or multiple planets which are embedded within the debris discs, on the other hand, such destabilising resonant interactions may lead to the formation of a wide gap within the debris disc, i.e., surrounding the planet(s) (e.g. Shannon et al., 2016; Schüppler et al., 2016; Lazzoni et al., 2018; Morrison & Kratter, 2018) – similar to those observed in many systems (e.g. Figure 2.6). Depending on the number of planets and their orbital parameters, the resultant gap could be devoid of debris particles either completely or partially. Note, however, that while this is one of the most widely considered mechanisms to explain the origin of observed radial gap structures, there are many other alternative mechanisms proposed in the literature (Pearce & Wyatt, 2015; Yelverton & Kennedy, 2018; Zheng et al., 2017). I will describe these in more detail in Chapter 5, where I present a novel pathway to sculpting such gaps based on secular interactions between a planet and an exterior but massive debris disc.

Finally, it has also been shown that resonant interactions may lead to the formation of very narrow gaps in debris rings, with a dearth of planetesimals at specific mean-motion resonances (Tabeshian & Wiegert, 2016); somewhat akin to the Kirkwood gaps in the asteroid belt. Nevertheless, such substructures have not been observed yet.

2.4.3 Structures due to Scattering

Apart from secular and resonant interactions, planetesimals may also be subject to e.g. scattering events as a result of close encounters with planets (Section 3.2.4.3). While scattered particles typically have short dynamical lifetimes (e.g. Levison & Duncan, 1997), they may also form long-lived scattered discs in the outer parts of planetary systems which could, in turn, continually replenish the population of exocomets (Muñoz-Gutiérrez et al., 2015; Wyatt et al., 2017) – similar to the trans-Neptunian objects in the Solar System. Depending on the architecture of the planetary system, exocomets could then be scattered into the inner parts, where they may disintegrate and release detectable amounts of dust and even gas. This mechanism has been invoked to explain the presence of dust – or exozodi – in the inner regions of exoplanetary systems, which are coming into increasingly sharper view (e.g., see review by Kral et al., 2017), and even to deliver cometary material including volatiles on to inner planets via impacts (Marino et al., 2018b).

To summarise, modelling of planet–debris disc interactions shows that the majority, if not all, of observed debris structures (Section 2.3) can be readily explained. Thus,

detailed modelling of debris disc morphologies provides an indirect way of probing and revealing the presence of planets, giving insights both in terms of their orbital elements, as well as their evolutionary history. Before moving on though, note that this is particularly important as often dynamically inferred exoplanets belong to a parameter space which remains inaccessible to existing instrumentation and so, if the interpretation is correct, they help in expanding our understanding of planetary populations. For instance, if the radial gaps observed within several debris discs are indeed sculpted by in-situ planets that have cleared their orbits of debris, as is often predicted, then their potential detection in the future will represent the first time ever that ice giants at such large separations (i.e., where the gap is centred, ~ 70 au) are discovered. Conversely, if the planet(s) are already identified, modelling of debris morphologies provide an avenue to better constrain their parameters and evolutionary history.

2.5 Incidence of Exoplanets and Debris Discs around the Same Stars

In the previous section I discussed how debris disc structures could reveal the presence and orbital parameters of often suspected – but as-yet undetected – planets in exoplanetary systems. However, some natural questions that arise are the following: do stars with debris discs also have planets? And, conversely, do stars with planets also have debris discs? Perhaps, considering the general similarity of debris discs to our own solar system's asteroid and Kuiper belts (e.g., Currie et al., 2015; Booth et al., 2017; Matrà et al., 2019), the incidence of planets in debris disc-hosting systems may be construed as plausible. Nevertheless, hard evidence is required to support this hypothesis.

The aim of this section is to briefly summarise what we know about the incidence of debris discs and exoplanets around the same stars, beginning with the exoplanetary system population considered as a whole and then proceeding to exemplary individual systems in which planets and debris coexist.

2.5.1 Correlation between Debris Discs and Planets on a Population Level

As already mentioned, debris discs are ubiquitous around main-sequence stars, with recent surveys detecting infrared dust excess around $\sim 20-30\%$ of nearby stars with spectral classes of A–K (e.g., Eiroa et al., 2013; Thureau et al., 2014; Montesinos et al.,

2016; Sibthorpe et al., 2018). On the other hand, it is also known that exoplanets are highly abundant around solar-type stars, with detection rates ranging from $\sim 1\%$ for directly imaged super-Jupiters at $\sim 10 - 100$ au separations (e.g., Bowler & Nielsen, 2018, and references therein) to more than $\sim 50\%$ for Doppler and Kepler planets of any mass at distances of ~ 1 au or less (e.g., Mayor et al., 2011; Winn, 2018, and references therein). These figures, however, are surely a lower limit of the prevalence of debris disc and/or exoplanet hosting stars due to limited sensitivity of current instrumentation. For instance, current sensitivity limits are not sufficient, e.g., (i) for a debris disc comparable to that of our solar system's Kuiper belt to be detected around a neighbouring star (Wyatt, 2008; Vitense et al., 2012), and (ii) for planets with $\gtrsim 10 M_J$ to be detected at separations larger than ~ 10 au.

With the expansion of the exoplanetary and debris disc albums over the last decades, several observational efforts have been made in order to identify correlations between the two. Historically, the first to consider this problem was Greaves et al. (2004), who searched for thermal dust emission in the sub-millimetre around eight stars known back then to harbour giant planets orbiting within a few au. They found that there is no evidence of a correlation between the presence of massive close-in planets and detectable debris discs (with dust masses down to $\sim 10^{-8} M_{\odot}$). A qualitatively similar conclusion was arrived at by Moro-Martín et al. (2007), this time based on the analysis of higher sensitivity mid-IR Spitzer observations of nine F, G, and K stars with one or more known radial-velocity planets.

On the other hand, by analysing a sample of the nearest 60 G-type stars, Wyatt et al. (2012) found a tentative evidence for higher incidence of detectable debris discs in close-in but low-mass planetary systems. Indeed, Wyatt et al. (2012) showed that debris discs are more common around stars which harbour planets of masses less than that of Saturn. This dependence of the correlation between planets and debris discs on planet mass was also confirmed by Marshall et al. (2014) based on PACS observations of 37 nearby exoplanetary systems.

However, it was later shown by Moro-Martín et al. (2015) that this correlation is not significant for stars harbouring either low-mass or high-mass planets at close separations. Their study, though, was based on small number statistics (with $N = 22$), making it difficult to ascertain whether the lack of correlation was a genuine physical effect. In light of this, the problem was recently revisited by Yelverton et al. (2020), who considered a much larger sample than Moro-Martín et al. (2015) (with $N = 138$) and confirmed, at high significance, that there is indeed no correlation between close-in planets and debris discs. In other words, the current data – with all of its limitations

due to detection constraints – indicates that debris discs are not more prevalent in systems with close-in planets than in systems that are selected without regard as to whether they have planets at several au or not.

Finally, while the works mentioned thus far have focused on systems hosting close-in planets that are typically discovered by radial velocity[6], Meshkat et al. (2017) have more recently considered the case of giant planets at $\sim 10 - 100$ au separations, i.e., those probed by direct imaging. Meshkat et al. (2017) found that debris disc-hosting systems are more likely to have large-separation giant planets than in those where no debris disc is detected, at the 88% confidence level. Indeed, they report a frequency of $\sim 6\%$ for the occurrence of distant giant planets around stars with debris discs, compared to $\sim 1\%$ for stars without discs. However, it is important to note that this tentative correlation is caveated with small number statistics – let alone potential selection biases (e.g., Bowler, 2016) – and larger samples are needed for a firm confirmation.

In conclusion, it is clear that the existence (or lack thereof) of a correlation between debris discs and planets cannot be generally ruled out at present, and further studies with improved detection limits of both planets and debris discs are required (see, e.g., Yelverton et al., 2020, for a detailed discussion).

2.5.2 Exemplary Systems with both Debris Discs and Planets

As discussed above, most of the presently known systems with debris discs do not have known companion(s) of planetary mass (or sub-stellar mass, for that matter). To date, however, there are few tens of planetary systems in which both planets and debris discs are detected[7]. Such systems are of particular importance as e.g. they provide an ideal test bed for our understanding of planet-disc interactions, and for advancing our theoretical frameworks. Here, I describe three systems which I have chosen with the intention of covering as wide a variety of planetary system architectures as possible.

• *Planets interior to the disc:* One of the most notable exoplanetary systems is β Pic, one of the first systems ever to have its debris disc detected and resolved (e.g., Smith & Terrile, 1984; Matrà et al., 2019, and references therein). Located at 19.3 pc (Crifo et al., 1997) and with an age of ~ 23 Myr (Nielsen et al., 2016), the A-type star β Pic is known to host an extremely broad debris disc extending from ~ 20 au out to beyond 1000 au. The disc is seen edge-on, and multi-wavelength observations have

[6]While it is true that most of the known exoplanets have been discovered by Kepler – i.e., using the transit method – they are usually too far away for a debris disc to be detected in those systems.

[7]A catalogue of systems in which both planets and debris discs are detected can be found online at http://svo2.cab.inta-csic.es/vocats/debris2/.

revealed various structures and asymmetries, with the warp at ~ 80 au being the most notable (Burrows et al., 1995; Kalas & Jewitt, 1995; Heap et al., 2000; Apai et al., 2015; Millar-Blanchaer et al., 2015). Remarkably, the warp was used to predict the presence of a giant planet dubbed as β Pic-b interior to the disc (at ~ 9 au, Mouillet et al., 1997; Augereau et al., 2001), and such a planet was later discovered by direct imaging (Lagrange et al., 2010). After the discovery of the planet, it was also demonstrated that β Pic-b could potentially also explain the x-shaped morphology observed in scattered light, in addition to the clumpy structure imaged in both CO and mid-IR (Nesvold & Kuchner, 2015b). Nevertheless, there remains some uncertainty in the interpretation of the structures in the β Pic disc. For instance, the orbital plane of β Pic-b is misaligned by $\sim 4°$ with the inner disc tilt, when it is expected for the mutual inclination to be closer to zero (e.g., Dawson et al., 2011; Millar-Blanchaer et al., 2015). Additionally, an additional giant planet of mass $\sim 9M_J$, dubbed as β Pic-c, was recently discovered to be orbiting the star at $\sim 2-3$ au (Lagrange et al., 2019). At the time of writing, it is not clear whether the scenario of warp production by β Pic-b alone would hold in light of this discovery. Finally, and relatedly, the location of the inner disc edge is thought to be too far from the giant planets for them to be responsible for truncating the disc at the observed location (e.g., Millar-Blanchaer et al., 2015). Clearly, then, there is a great deal yet to learn about this system, both from future observations of the system and tailored modelling of planet–disc interactions.

• *Planets in between two belts of debris:* Another notable example is the HR 8799 system, one of the most well-known directly imaged systems. HR 8799, a $\sim 50-70$ Myr A-type star located at 41.3 pc (Gaia Collaboration, 2018), is known to host four directly imaged planets between ~ 15 and 70 au, each with masses between ~ 5 and $10M_J$ (Marois et al., 2008, 2010). The system also hosts a narrow ring of warm debris interior to the planets, as well as a cold planetesimal belt exterior to the planets which extends from ~ 150 to 400 au[8] (e.g., Sadakane & Nishida, 1986; Su et al., 2009; Hughes et al., 2011; Patience et al., 2011; Matthews et al., 2014b; Booth et al., 2016). Given these observations, the HR 8799 is often considered as a scaled-up sibling to our own solar system, at least its outer portion. It is also notable that, somewhat akin to the solar system, the inner and outer edges of the warm debris belt in HR 8799 seem to coincide with the 4:1 and 2:1 mean motion resonances with the innermost planet (Marois et al., 2010). Additionally, and relatedly, there are indications, based on planet formation arguments (e.g., Dodson-Robinson et al., 2009), that the planets in

[8]I note that a halo of small grains has also been detected outside the planetesimal belt, extending out to ~ 1000 au (e.g., Su et al., 2009; Matthews et al., 2014b).

HR 8799 could have undergone significant migration to their current location (Marois et al., 2010). Hence, much effort has been put into understanding the HR 8799 system, as it could also be useful for transforming our knowledge of solar system formation and evolution. Currently, however, one of the outstanding open issues related to the external HR 8799 disc is its inner edge location, which cannot be explained by dynamical interactions with the known planets, and hence has led to the postulation of an additional planet between the outermost known planet and the cold disc (Booth et al., 2016; Read et al., 2018).

• *A planet exterior to a circumbinary disc:* Another system which stands out is HD 106906 AB, a ~ 15 Myr spectroscopic binary consisting of two F-type main-sequence stars with similar masses and projected separation of ~ 0.5 au (Lagrange et al., 2016; Pecaut & Mamajek, 2016). Unlike most other systems, HD 106906 AB contains a planet which is external rather than internal to its debris disc. Indeed, the $\sim 11 M_J$ planet, dubbed as HD 106906-b, orbits the central binary with a semimajor axis of ~ 850 au (Bailey et al., 2014), farther away from the debris disc. The debris disc – which is seen almost edge-on to the line of sight – extends between ~ 50 and a few hundreds of au and is characterised with an asymmetric dust distribution that is more extreme toward the planet's location on the sky (e.g., Kalas et al., 2015; Lagrange et al., 2016; Wu et al., 2016). Interestingly, previous work has shown that the perturbations by HD 106906-b could potentially explain the observed features of the disc (Jílková & Portegies Zwart, 2015; Nesvold et al., 2017; Rodet et al., 2017). This said, however, it is worthwhile to note that the orbital parameters of HD 106906-b have since been observationally refined (Nguyen et al., 2021), finding that its orbit is of lower eccentricity (~ 0.5) and larger mutual inclination with respect to the disc ($\sim 40°$) than previously thought (~ 0.7 and $\sim 10°$, respectively). The effects of this remain to be examined, but nevertheless it is interesting from a dynamical and theoretical point of view that such systems are already observed.

2.6 Masses of Debris Discs

As already mentioned briefly in Sections 1.2.4 and 2.2.2, obtaining a reliable estimate for the total masses of debris discs – which is arguably one of their most fundamental characteristics – is not a straightforward exercise. This is mainly due to the fact that observations can only probe the mass of the emitting dust grains (i.e., $\lesssim$mm-sized), and not that of the larger dust-producing planetesimals which dominate the disc mass (Section 2.1). In this section, I first review existing methods for estimating the total

masses of debris discs (Section 2.6.1), providing a rough estimate for their allowed range, and then discuss their implications within the context of modelling debris discs and planet–disc interactions (Section 2.6.2). This discussion will facilitate not only in justifying my choices for the values of disc masses that I later consider when examining interactions between planets and debris discs (Chapters 5 and 6), but also in highlighting the importance of my research findings.

2.6.1 Methods of Estimation and Outcomes

One of the major problems in the field of debris discs is that their masses are uncertain and can vary by orders of magnitude depending on the estimation technique used and the assumptions that go into them. For a recent discussion of this subject see Krivov & Wyatt (2021), from which the following is summarised.

There are two methods in the literature for measuring the masses of debris discs. The first and standard method is by extrapolating the measured masses of the dust to the unobservable large planetesimals. This procedure requires knowledge of the size distribution of bodies in the disc, from the smallest dust grains of size $D_{\rm min}$ to the largest planetesimals of size $D_{\rm max}$, as well as parameters that govern the collisional evolution (e.g. planetesimal strengths). Within this context then, the total disc mass M_d can be calculated via (Krivov & Wyatt, 2021):

$$M_d = M_{\rm dust}\left(\frac{D_{\rm max}}{D_{\rm mm}}\right)^{4-\Theta}, \tag{2.13}$$

where $M_{\rm dust}$ is the total mass of the dust grains with diameters from $D_{\rm min}$ up to $D_{\rm mm}$, the latter corresponding to the sizes of particles that are probed by (sub-)mm observations, i.e., $D_{\rm mm} \sim 1$ mm.

The most basic assumption in the literature is to adopt the classical size distribution of Dohnanyi (1969) valid for a constant strength of bodies, i.e., $n(D) \propto D^{-\Theta}$ with $\Theta = 3.5$ (Equation 2.11), which is a good approximation to more detailed models (e.g. see Figure 2.4; as well as Krivov & Wyatt (2021) and references therein). Equation (2.13) then implies that $M_d \propto D_{\rm max}^{1/2}$: that is, the total mass of a debris disc containing up to, say, $\sim$ 100 km-sized bodies will be larger than the measured dust masses by at least 4 orders of magnitude. As a rough estimate, taking the nominal values of $M_{\rm dust} \sim 10^{-3} - 10^{-1} M_\oplus$ from Figure 1.3, one then obtains $M_d \sim 10 - 1000 M_\oplus$. Exemplary values of disc masses resulting from this extrapolation method, as applied to a sample of 20 debris discs observed by ALMA, is shown in Figure 2.8. Note that

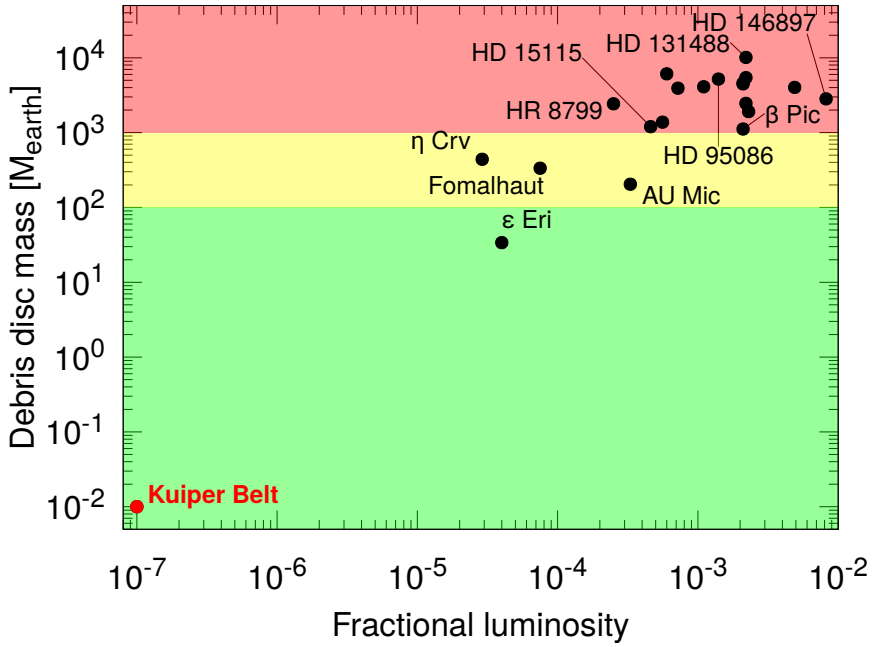

Figure 2.8. The estimated values of the total masses of a sample of 20 debris discs observed by ALMA (taken from Krivov & Wyatt, 2021). The disc masses are calculated by extrapolating the masses of observed dust grains up to the unobservable large planetesimals, assuming ideal, Dohnanyi collisional cascade: that is, using Equation (2.13) with $\Theta = 3.5$, $D_{\rm min} = 1$ mm and an assumed value of $D_{\rm max} = 200$ km. The different coloured areas correspond to the allowed (green), forbidden (red), and intermediary (yellow) ranges; see the text (Section 2.6.1) for details.

this 'back of envelope' calculation yields values similar to those obtained with more detailed collisional cascade models (Krivov & Wyatt, 2021).

It is, however, worth noting that this extrapolation procedure yields relatively large values for the total masses of some bright debris discs, i.e. on the order of $\sim 10^3 - 10^4 M_\oplus$ (see e.g. Figure 2.8, as well as Table 1 in Krivov & Wyatt, 2021). Such values are unrealistic and problematic (highlighted for reference in red in Figure 2.8), as they clearly exceed the masses of solids available in the protoplanetary disc phase (highlighted in yellow in Figure 2.8) – a problem referred to as the "debris disc mass problem" in the literature (Krivov et al., 2018; Krivov & Wyatt, 2021). As a solution to this conundrum, it has been suggested that uncertainties and various assumptions entering in this extrapolation procedure would need to be revisited. For instance, Krivov & Wyatt (2021) conclude that perhaps the easiest solution is that the largest planetesimals in debris discs are on the order of $\sim$ km – rather than ~ 100 km. Note, however, that such planetesimals would be smaller than the largest objects observed in

the Kuiper belt (e.g., Pluto's size is ~ 1000 km, see also Fraser et al., 2014) or those produced in theoretical planetesimal formation models (e.g. Simon et al., 2016).

The second method for estimating the total mass of a given debris disc is by measuring its mass-loss rate $|\dot{M}_d|$ at its current age t, and interpreting that within the context of collisional cascade models. This procedure allows one to recover the initial total disc mass, $M_d(0)$, that can collisionally sustain the observed amounts of dust in a given debris disc (recall that dust grains are short-lived compared to the stellar age; Section 2.2). This can be understood by considering the mass decay law given by Equation (2.12), which, upon differentiation, reads as:

$$M_d(0) = |\dot{M_d}(t)|\tau_0(1 + t/\tau_0)^2, \tag{2.14}$$

where we recall that τ_0 is the collisional lifetime of the largest planetesimals at the initial epoch. Then, using the fact that the collisional cascade is fed by planetesimals of a given size D_c for which the collisional lifetime is equal to the age of the system (e.g. Wyatt & Dent, 2002; Krivov & Wyatt, 2021), i.e., $t = \tau_0$, one arrives at:

$$M_d(0) = 4|\dot{M}_d|t. \tag{2.15}$$

Note that it is possible for the disc to contain a population of large planetesimals with $D > D_c$ that do not contribute to the collisional cascade within the age of the system: thus, Equation (2.15) represents the *lower* limit on the initial total disc mass.

The rate $|\dot{M}_d|$ entering in Equation (2.15) can be inferred from observations of dust in two distinct ways. The first relies on linking physical properties of the disc (which are described below) to the observed mass of small grains and their collision rate, assuming an idealized collisional cascade (Dohnanyi, 1969). Matrà et al. (2017) showed that this allows one to analytically express the mass-loss rate $|\dot{M}_d|$ as a function of the disc's fractional luminosity, radial location, and radial width – which can be well characterised through SED fitting and high-resolution imaging – as well as the stellar mass and luminosity; see their equation (21). Alternatively, $|\dot{M}_d|$ can be inferred from the debris disc's vertical thickness at sub-mm wavelengths (Daley et al., 2019; Matrà et al., 2019), if it happens to be nearly edge-on to the line of sight. This is because the disc's vertical thickness at such wavelengths is an indicator of the inclination distribution of the large dust-producing planetesimals which, through equipartition of energy (assuming equilibrium), can be related to the orbital eccentricities and thus the relative velocities of colliding planetesimals. Based on this, it is then straightforward to compute $|\dot{M}_d|$ through standard formulae for collisional rates (e.g. Wyatt, 2008).

Application of this method typically yields values of $\sim 10 M_{\oplus}$ for the *minimum* total mass of a debris disc (e.g. Wyatt & Dent, 2002; Daley et al., 2019; Matrà et al., 2019; Krivov & Wyatt, 2021).

On the other end of the spectrum, however, a question that naturally arises is: how massive is a debris disc allowed to be? To answer this question, one could consider the total mass of dust grains that were present in the protoplanetary disc phase, from which the planetesimals constituting a debris disc – and potentially planets – formed. For simplicity, assuming a Sun-like star with a typical protoplanetary disc mass of $0.1 M_{\odot}$ and dust-to-gas ratio of 100 : 1 (e.g. Williams & Cieza, 2011), the maximum possible mass of a debris disc would be $0.001 M_{\odot}$, or $300 M_{\oplus}$ – assuming planetesimals were neither ejected from the system nor agglomerated into planets. Obviously, this estimate can also differ by a factor of two or so, depending on the stellar mass (as more massive stars are known to host more massive protoplanetary discs, see Figure 5 in Williams & Cieza, 2011) and the assumed dust-to-gas ratio (which, for instance, in the Solar System could be 60 : 1, see Lodders, 2003). Given this, values on the order of $\sim 10^2 - 10^3 M_{\oplus}$ might be a good estimate for the *maximum* masses of debris discs – this range is highlighted in yellow in Figure 2.8.

To summarise, estimating the total masses of debris discs based on observations of their constituent dust is a difficult and uncertain task. Depending on the adopted method and the assumptions that go into it, debris discs could be as massive as $\sim 10 M_{\oplus}$, if not $\sim 10^4 M_{\oplus}$ (see also Wyatt & Dent, 2002; Krivov et al., 2018; Krivov & Wyatt, 2021). The commonly accepted – or perhaps, plausible – maximum masses of disc masses, however, is $\lesssim 100 M_{\oplus}$; which is also consistent with the theoretical predictions of planetesimal formation models. For instance, it has been recently shown by Carrera et al. (2017) that starting with an $0.1 M_{\odot}$ protoplanetary disc around a Sun-like star, a total of $\sim 80 - 150 M_{\oplus}$ of planetesimals can robustly form through streaming instability, of which $\sim 60 - 130 M_{\oplus}$ populate the region beyond 100 au from the star.

2.6.2 An Unexplored Alternative: Debris Disc Structures

A potential alternative method for accessing the total masses of debris discs is through studying processes that depend only on the total disc mass, but not the finer details which are often poorly constrained, such as the size distribution of planetesimals and their maximum sizes. To this end, modelling of gravitational interactions between planets and massive debris discs, and examining the observational consequences in terms of the spatial structure of debris discs represents the best example. Such an

approach would render the finer details about the physical properties of dust grains and planetesimals irrelevant, allowing one to indirectly infer the *dynamical* mass of the debris disc; of course, provided that the predicted debris disc structures reproduce the observations. Nevertheless, this approach has not been well appreciated in the debris disc literature, which is what I discuss next.

Indeed, motivated by the fact that debris discs could be relatively massive ($\sim 10-100 M_{\oplus}$), the need to account for their gravitational effects when modelling planet–disc interactions has been previously cited in the literature – though rarely considered in detail. Indeed, debris discs are often treated as a collection of massless particles subject only to the gravity of external perturbers (e.g. planets and stellar companions); see Section 2.4 and references therein. A literature review on the subject reveals two approaches in handling the effects of debris disc gravity: (i) they are either not included but acknowledged as a limitation (Wyatt, 2005b; Pearce et al., 2021); or (ii) when accounted for, it is not done fully self-consistently, in the sense that only the back-reaction of the disc on the planet is modelled, but not the gravitational effects of the disc on itself (Beust et al., 2014; Pearce & Wyatt, 2015).

To the best of my knowledge, the only exceptions to this are the studies presented by Nesvold et al. (2016), Zheng et al. (2017), and Jalali & Tremaine (2012). In Nesvold et al. (2016), the authors numerically studied the evolution of massive debris discs being perturbed by an inclined exterior perturber, accounting for the full gravitational potential due to the disc. They found that the inclusion of the disc self-gravity can suppress the Kozai-Lidov oscillations induced by the external body (somewhat akin to massive protoplanetary discs in binary star systems; e.g., Batygin et al. (2011)). On the other hand, with the help of numerical simulations, the authors in Zheng et al. (2017) considered the evolution of a massive gaseous disc (i.e., a transitional disc) in the presence of a single planet embedded within. They showed that, as the gaseous disc is depleted, its waning gravitational potential could carve a gap around the planetary orbit via sweeping secular resonances. They found that the resulting gap in the gas-poor 'debris' disc is wider than that anticipated otherwise based on considerations of overlapping first-order MMRs in massless discs. Last but not least, Jalali & Tremaine (2012) analytically studied the evolution of debris discs influenced only by their self-gravity (and, surely, the gravity of the host star), in the absence of external perturbers. They showed that self-gravitating debris discs can support large-scale 'slow' ($m = 1, 2$) modes, which, if and when excited (e.g., by stellar flybys), could reproduce many of the observed debris discs features – challenging the common practice of ascribing observed features to the presence of yet-unseen planets. Before

moving on, it is worth noting that somewhat intriguingly, the oft-neglected role of debris disc (self-)gravity pertains only to extra-solar settings: indeed, within the context of the solar system, the gravitational role of the asteroid and Kuiper belts both on their own dynamical evolution and the planets in the solar system has been studied extensively (e.g. Heppenheimer, 1980; Ward, 1981; Ward & Hahn, 1998a,b; Nagasawa & Ida, 2000; Nagasawa et al., 2001; Hahn, 2003, 2007, 2008; Sefilian & Touma, 2019).

One of the hurdles in accounting for the gravitational effects of debris discs (or generally, astrophysical discs, for that matter) has been the computational and theoretical challenges involved. From a computational point of view, and in spite of remarkable advances in the domain within the last decades, traditional direct N-body simulations – such as `MERCURY` (Chambers, 1999) and `REBOUND` (Rein & Liu, 2012) – remain very expensive and inefficient in terms of CPU time to allow for an extensive exploration of the role of disc self-gravity in planet–disc interactions. This is mainly because conventional N-body codes both require $\mathcal{O}(N^2)$ computations per time step and converge slowly especially when orbits have a close approach or intersect (leading to a discontinuity in the force calculation). From a theoretical point of view, on the other hand, most existing methods for computing the gravitational potential of discs remain somewhat specialised and limited in scope for practical purposes. This is mainly because such analytical calculations hold inherent assumptions about the disc structure and its time variability: for instance, an axisymmetric power-law surface density distribution that does not change in time due to, say, external eccentric perturbers (e.g. Heppenheimer, 1980; Ward, 1981; Batygin et al., 2011), or non-axisymmetric surface density distribution but with pre-defined ansatz assumptions on the eccentricity and apsidal angle distributions of the constituent disc particles (e.g. Silsbee & Rafikov, 2015a; Davydenkova & Rafikov, 2018; Sefilian & Touma, 2019; Teyssandier & Lai, 2019, Davydenkova & Rafikov, in preparation.). Such limitations can be easily alleviated by treating the disc as composed of $N \gg 1$ massive rings (or wires) with prescribed spacing – justified by the constancy of the semimajor axis in the secular regime, see e.g. Section 3.2.4.1 – interacting gravitationally with each other. However, this requires the use of softened forms of gravity – whereby the Newtonian point-mass is spatially smoothed by the introduction of some form of softening parameter (e.g. Tremaine, 2001; Touma, 2002; Touma et al., 2009; Hahn, 2003; Batygin, 2012; Teyssandier & Ogilvie, 2016) – and a physical justification for a specific form of softening often remain unclear, making its introduction somewhat arbitrary or *ad hoc*.

These difficulties, coupled with the general consensus that the gravitational effects of discs can be ignored as long as their masses are less than that of external perturbers

(e.g. Wyatt, 2005b; Beust et al., 2014; Pearce & Wyatt, 2014) – which, in my opinion, has not been rigorously addressed – seem to have left the problem untouched, without a detailed investigation. This is, in essence, where my research comes in. Indeed, in Chapters 4 – 6, I will first investigate different methods of computing the gravitational potential of discs, and then employ my findings to study the gravitational role of debris discs in planet–disc interactions, and demonstrate how observed structures – specifically, radial gaps such as those in Figure 2.6 – could be used to e.g. indirectly measure the total masses of debris discs. Before presenting my work, though, a brief overview of the basics of celestial mechanics is presented in the upcoming chapter.

Chapter 3

Basics of Celestial Mechanics

This dissertation is mainly concerned with the orbital evolution of planetary systems harbouring both planets and many smaller debris particles. This is a generally complex problem involving coupled gravitational interactions between N bodies, each with its mass m_i and position vector $\mathbf{r}_i$ ($i = 1..N$), defined relative to some inertial reference frame with origin O. Indeed, the equation of motion describing the evolution of each body can be written as:

$$\frac{d^2\mathbf{r}_i}{dt^2} = -G \sum_{j=1, j\neq i}^{j=N} m_j \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3}, \tag{3.1}$$

where G is the gravitational constant.

Generally speaking, such N-body systems are not integrable and do not admit exact analytic solutions; hence, are commonly investigated using N-body simulations. Nevertheless, in cases where one of the bodies (e.g. the star) is much more massive than the other bodies (e.g. planets and debris particles) – as is the case in many astrophysical problems – the problem can be probed analytically using perturbation theories. The starting point of such perturbation theories is the well-known integrable "Keplerian problem", which concerns the motion of two bodies under their mutual gravitational attraction. The solution of the two-body problem allows to model perturbations arising from additional bodies as small deviation from the Keplerian problem, yielding approximate analytic solutions to N-body systems.

In this Chapter I provide a brief overview of the basics of celestial mechanics, starting from the Keplerian problem and then moving on to perturbation theories relevant for the work in this dissertation. To this end, I consider particles to be affected by Newtonian gravity only, neglecting effects that could arise due to general relativity,

tides, and non-sphericity of particles. For a complete theoretical treatment of the subject, the reader is referred to the treatise by Plummer (1918) and the textbook by Murray & Dermott (1999), from which the following is summarised. The reader well familiar with the subject may safely skip to the next Chapter without any loss of continuity.

3.1 The Two-Body Keplerian Problem

The two-body Keplerian problem serves as the starting point in developing perturbations theories. Here I outline the main results of the Kepler problem starting with the planar case, and then moving on to the spatial case.

3.1.1 The Planar Case

Consider a system composed of two point-mass particles. Without loss of generality, let us assume that $m_1 > m_2$, so that m_1 and m_2 are referred to as the central and secondary bodies, respectively. In this case, and after some straightforward algebra, the equations of motion given by Equation (3.1) can be reduced to:

$$\mu \frac{d^2\mathbf{r}}{dt^2} = -\frac{G\mu(m_1 + m_2)}{r^3}\mathbf{r}, \tag{3.2}$$

where the vector $\mathbf{r}$ represents the relative position of the two bodies, i.e., $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$, in the frame centred at an arbitrary point O, and $\mu = m_1 m_2/(m_1 + m_2)$ is the reduced mass of the system. It is then straightforward to show that the vectorial quantity $\mathbf{h}$, which represents the total orbital angular momentum of the system, $\mathbf{h} \equiv \mu\mathbf{r} \times \dot{\mathbf{r}}$, is a constant of the motion. This is a restatement of the fact that the gravitational force is radial. Additionally, since $\mathbf{h} \cdot \mathbf{r} = 0$, it follows that the motion of the two bodies relative to one another is restricted to a plane defined by the vectors $\mathbf{r}$ and $\mathbf{v} = \dot{\mathbf{r}}$, whose normal is parallel to $\mathbf{h}$. Mathematically, this essentially reduces the two-body problem to the problem of a single body of mass μ and position vector $\mathbf{r}$ subjected to an external force given by the right hand side of Equation (3.2).

The solution of Equation (3.2) is well known to describe a conic section: i.e., an ellipse, parabola, or hyperbola. Throughout this dissertation, we are mainly interested in the elliptical solution, as it represents a bound periodic trajectory. In this case, and

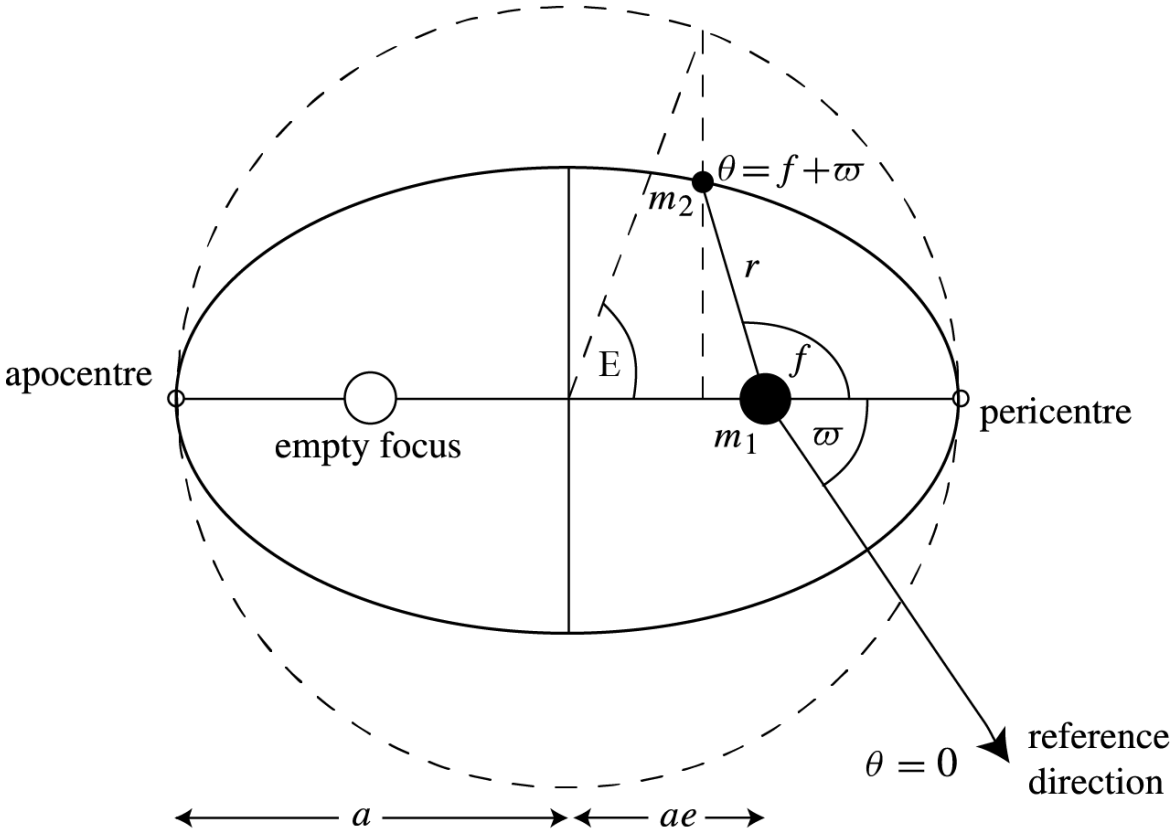

Figure 3.1. An ellipse representing the Keplerian orbit of a particle (m_2, small circle) around the central body (m_1, large circle). The primary is located at the primary focus of the ellipse, while the secondary focus is empty. This diagram also illustrates the geometrical meanings of the various orbital elements defined in the text – see Section 3.1.1 for details.

in terms of polar coordinates (r, θ), the solution of Equation (3.2) reads as:

$$r = \frac{a(1-e^2)}{1+e\cos f}, \tag{3.3}$$

where $f \equiv \theta - \varpi$. Equation (3.3) describes the elliptical Keplerian orbit of the two bodies relative to each other, and is characterised with its semimajor axis a, eccentricity e, longitude of pericentre ϖ, and true anomaly f – the geometrical meanings of which are provided next and illustrated in Figure 3.1.

The shape and size of a conic section describing the secondary's orbit are determined by two parameters: the orbital *eccentricity* e and *semi-major axis* a. The eccentricity is a measure of the departure of the conic section's shape from a circle: (i) circular orbits have $e = 0$; (ii) elliptical orbits have $0 < e < 1$, such that the more elongated the ellipse is, the higher the value of e; (iii) parabolic orbits have $e = 1$; while (iv) hyperbolic

orbits have $e > 1$. In this thesis, I will not deal with unbounded (parabolic and hyperbolic) orbits, and as such I focus, hereafter, on the elliptical case. The semi-major axis represents half the length of the major axis of the conic section, which is equivalent to the average of the pericentre and apocentre distances, r_p and r_a, respectively. Here, the pericentre and apocentre distances are defined as the distance between the primary focus of the elliptical orbit – occupied by the central body about which the secondary orbits (while the second focus remains empty) – and the point of closest and furthest approach to it, respectively. Mathematically, these can be quantified as $r_p = a(1-e)$ and $r_a = a(1+e)$, respectively. The orientation of the orbit, on the other hand, is defined by the angle ϖ, known as the *longitude of pericentre*. This angle, defined such that $0 \leq \varpi \leq 360°$, is a measure of the angle between the line connecting the primary focus with the pericentre position and an arbitrarily chosen reference direction. Finally, the position of the particle along its orbit is determined by the *true anomaly*, f. This measures the instantaneous angle between the direction of the pericentre and the particle's position along its orbit around the central body. It is always in the range $0 \leq f \leq 360°$, so that $f = 0°$ at the pericentre and $f = 180°$ at the apocentre.

The knowledge of this set of orbital elements – i.e., a, e, ϖ, and f – completely defines the instantaneous position r (Equation 3.3) and speed v of the orbiting body along its elliptical trajectory, where the expression of the latter is given by:

$$v^2 = G(m_1 + m_2)\left(\frac{2}{r} - \frac{1}{a}\right). \tag{3.4}$$

Indeed, as the particle orbits the star, its position and velocity along the elliptical trajectory change in time in such a way that a, e, and ϖ remain constant (surely, assuming no external perturbations), whilst f advances in time. Note, however, that the true anomaly does not increase with time at a uniform rate: it evolves faster near the pericentre than near the apocentre, where particles spend most of their time along their orbits. This can be also seen by noting that the instantaneous speeds at the pericentre and apocentre are given by:

$$v_p = na\sqrt{\frac{1+e}{1-e}} \qquad \text{and} \qquad v_a = na\sqrt{\frac{1-e}{1+e}}, \tag{3.5}$$

respectively, where n is the mean motion given by Equation (3.7).

For this reason, it is customary in celestial mechanics to replace the time variable (or f) by a dimensionless variable M, known as the *mean anomaly*. This variable –

and unlike the true anomaly – advances linearly in time such that

$$M = n(t - t_0). \tag{3.6}$$

Here, n is the mean motion or orbital frequency of the particle orbit given by

$$n = \sqrt{\frac{G(m_1 + m_2)}{a^3}}, \tag{3.7}$$

so that $T = 2\pi/n$ is the orbital period, and t_0 is the time of pericentre passage. Note that the mean motion does not depend on the shape of the trajectory, i.e., eccentricity, but only on its size.

While the mean anomaly has no simple geometrical interpretation, it is related to another geometrical angle, known as the *eccentric anomaly*, which is customarily used in celestial mechanics. The eccentric anomaly E is the angle subtended at the centre of the ellipse by the projection of the body's instantaneous position on the circle with radius a and tangent to the ellipse at its extremities. Mathematically, the mean and eccentric anomalies are related through a transcendental equation known as the 'Kepler equation':

$$M = E - e \sin E. \tag{3.8}$$

Eccentric and true anomalies, on the other hand, are related by the following formulae:

$$\tan \frac{f}{2} = \sqrt{\frac{1+e}{1-e}} \tan \frac{E}{2}, \tag{3.9}$$

$$\cos f = \frac{\cos E - e}{1 - e \cos E}, \tag{3.10}$$

$$\sin f = \frac{\sqrt{1-e^2} \sin E}{1 - e \cos E}, \tag{3.11}$$

which can be obtained using elementary geometrical relationships (see Figure 3.1). Note that for circular orbits, the three anomalies become identical, i.e., $M = E = f$.

With these definitions at hand, one can also express the instantaneous position of the particle as follows:

$$r = a(1 - e \cos E). \tag{3.12}$$

The definitions of M, E, and f and their relationships (i.e., Equations 3.8–3.11) will prove useful particularly when a certain dynamical quantity needs to be orbit-averaged, i.e., averaged over the orbital period of a body, as done primarily in Chapter 4.

Before we move on to discuss orbits in three dimensional space, it is worthwhile to recall that thus far we have considered the orbit of one particle relative to the other. This is because mathematically, the choice of the central body is arbitrary. However, it is equally valid to define the orbits of both particles relative to the system's centre of mass located at

$$\mathbf{r}_{\mathrm{CM}} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2}. \tag{3.13}$$

In this case, in the frame whose origin is at $\mathbf{r}_{\mathrm{CM}}$, also known as the barycentric frame, it is trivial to show that both the central and secondary bodies orbit on elliptical trajectories around their common centre of mass with eccentricities $e_1 = e_2 = e$, longitudes of pericentre $\varpi_2 = \varpi_1 + 180°$, and semimajor axes given by

$$a_1 = \frac{m_2}{m_1 + m_2}a \quad \text{and} \quad a_2 = \frac{m_1}{m_1 + m_2}a. \tag{3.14}$$

Note that since $a_1/a_2 \propto m_2/m_1$, when the mass ratio tends to zero – as would happen naturally say e.g., for a planetesimal and a star – the star's orbit shrinks to the barycentre position, and the orbit of the small body relative to the star becomes equivalent to the small body's orbit in the inertial reference frame.

3.1.2 The Spatial Case

Thus far, I described how the orbital elements a, e, ϖ, and f uniquely characterise the orbit of a particle in two dimensional space. In general, however, the orbital plane of an object may also be inclined with respect to the reference plane, intersecting it in two points known as the *nodes* of the orbit. Therefore, two additional angles are required to uniquely characterise the orbit of a particle in three dimensional space.

The first of these is the angle I, known as the *orbital inclination*, which is a measure of the angle between the reference plane and the orbital plane. This angle is defined such that one always has $0 \leq I \leq 180°$, with orbits of $I < 90°$ referred to as prograde, whilst those with $I \geq 90°$ referred to as retrograde. Next is the angle Ω, known as the *longitude of ascending node*, which is a measure of the angle between a given reference direction and one of the nodes of the orbit, known as the ascending node. The latter is where the particle crosses the reference plane from negative to positive in the orthogonal direction (i.e. z). The longitude of ascending node is always in the range $0 \leq \Omega \leq 360°$. Finally, when considering orbits in three dimensional space, it is also customary to write the longitude of pericentre as the sum of two 'dog-leg' style

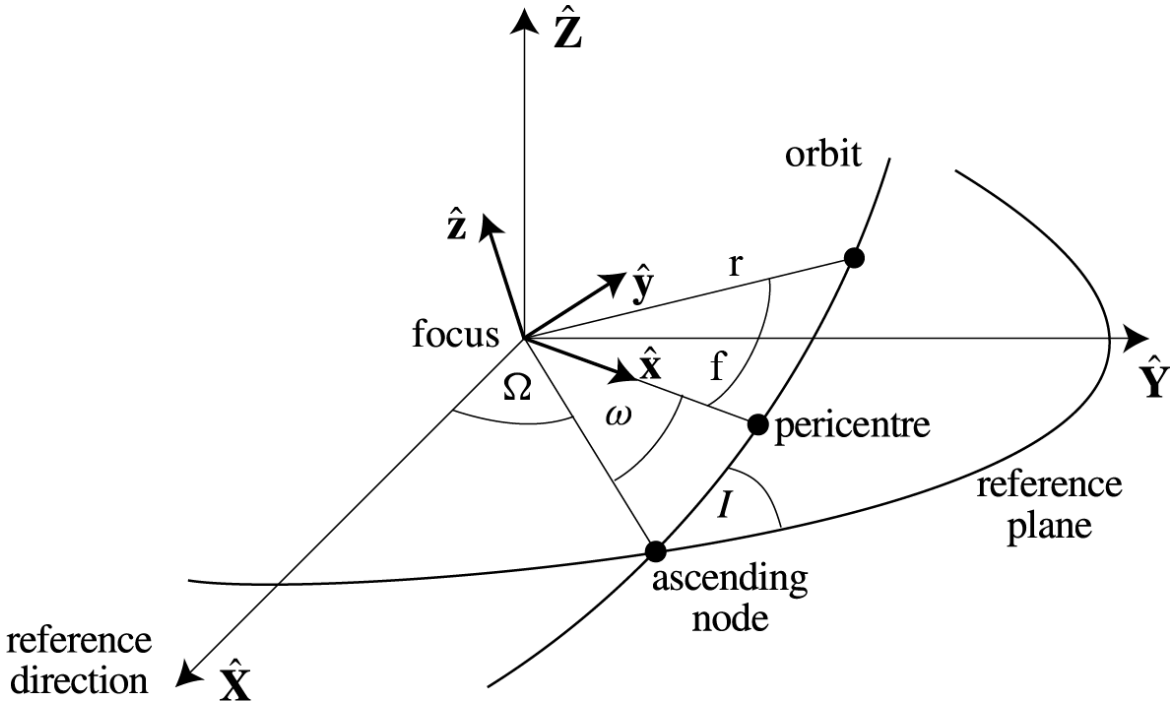

Figure 3.2. Visualisation of the Keplerian orbit of a particle about the central body in three dimensional space. The reference frame (X, Y, Z) is centred at the orbit's focus, occupied by the central body. The definitions of the various angles highlighted in this diagram – namely, I, Ω, and ω – are given in the text; see Section 3.1.2.

angles,

$$\varpi = \omega + \Omega, \tag{3.15}$$

which lie in different planes. Here, the angle ω is known as the *argument of pericentre*, which is a measure of the angle between the line connecting the focus of the orbit to its ascending node and the pericentre direction in the orbital plane. This angle is always in the range $0 \leq \omega \leq 360°$. These three angles – namely, I, Ω, and ω – are illustrated in Figure 3.2.

The set of the six orbital elements a, e, ϖ, I, Ω, and f completely define the position and velocity of the orbiting body. Indeed, it is trivial to show that there is one-to-one correspondence between the components of the position and velocity vectors and the orbital elements. For instance, in an arbitrary reference frame (X, Y, Z) centred on the central body (or the focus), with X being the reference direction and (X, Y) the reference plane (as represented in Figure 3.2), the position vector $\mathbf{r}$ of the

orbiting body reads as:

$$\mathbf{r} = r \begin{pmatrix} \cos\Omega\cos(\omega+f) - \sin\Omega\sin(\omega+f)\cos I \\ \sin\Omega\cos(\omega+f) + \cos\Omega\sin(\omega+f)\cos I \\ \sin(\omega+f)\sin I \end{pmatrix}. \tag{3.16}$$

A similar expression can be derived for the velocity vector $\mathbf{v}$; see Murray & Dermott (1999). Finally, it is worthwhile to note that when the orbital plane coincides with the reference plane, i.e., $I = 0$, the ascending nodes are no longer defined, and thus the angles ω and Ω too. However, their sum, i.e., $\varpi = \omega + \Omega$, remains well defined and all expressions, e.g., Equation (3.16), become naturally dependent on ϖ only, and not ω and Ω separately.

3.2 Dynamical Theory: beyond Keplerian Orbits

Thus far, we only considered the motion of two bodies and their Keplerian motion around each other. Throughout this thesis, however, we are mainly concerned with systems containing multiple bodies – namely, a planet and a debris disc composed of $N \gg 1$ planetesimals in orbit around a central star. Unfortunately, however, the general N-body problem – unlike the Kepler problem – is not integrable and has no simple analytical solution. As such, one of the main efforts of celestial mechanics in the last centuries has been the development of perturbation theories. In essence, such theories treat additional perturbations as small deviations from the integrable Kepler problem, allowing one to derive approximate solutions describing the orbital evolution of the N bodies involved. These deviations can be calculated by making use of the so-called *disturbing function*, which is the subject of this section. Within this context, the set of the six orbital elements defined in Chapter 3.1 still define the instantaneous position and velocity vectors of the interacting bodies, but they are now thought of as *osculating elements*; in the sense that, they define the orbit of a perturbed body if the perturbing force were to disappear at that instant of time. Below, I review the basic concepts of dynamical theory beyond Keplerian orbits, focusing on the definition, analysis, and use of the disturbing function.

3.2.1 The Disturbing Function

For heuristic purposes, let us consider the motion of two point-mass bodies m_1 and m_2 orbiting around a central star of mass M_c. Let the position vectors of the two bodies

relative to the central star be $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively. It is then trivial to demonstrate that the equations of motion can be written as (see Equation 3.1):

$$\frac{d^2\mathbf{r}_1}{dt^2} + G(M_c + m_1)\frac{\mathbf{r}_1}{r_1^3} = Gm_2\left(\frac{\mathbf{r}_2 - \mathbf{r}_1}{|\mathbf{r}_2 - \mathbf{r}_1|^3} - \frac{\mathbf{r}_2}{r_2^3}\right), \tag{3.17}$$

$$\frac{d^2\mathbf{r}_2}{dt^2} + G(M_c + m_2)\frac{\mathbf{r}_2}{r_2^3} = Gm_1\left(\frac{\mathbf{r}_1 - \mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2|^3} - \frac{\mathbf{r}_1}{r_1^3}\right). \tag{3.18}$$

When either of the two bodies is absent, the right hand sides of Equations (3.17) and (3.18) vanish, reducing the equations of motion to that of the Kepler problem (Equation 3.2). Thus, the right hand sides in the these equations represent the accelerations of m_1 and m_2 arising due to their mutual gravitational interaction. These relative accelerations can be obtained from the gradients of a perturbing potential known as the *disturbing function*. This can be seen by noting that the equations of motion can be written as:

$$\frac{d^2\mathbf{r}_1}{dt^2} = \nabla_1(U_1 + R_1) \quad \text{and} \quad \frac{d^2\mathbf{r}_2}{dt^2} = \nabla_2(U_2 + R_2), \tag{3.19}$$

where

$$U_1 = \frac{G(M_c + m_1)}{r_1} \quad \text{and} \quad U_2 = \frac{G(M_c + m_2)}{r_2} \tag{3.20}$$

are the standard two-body (or Kepler) part of the total potential, and

$$R_1 = \frac{Gm_2}{|\mathbf{r}_2 - \mathbf{r}_1|} - Gm_2\frac{\mathbf{r}_1 \cdot \mathbf{r}_2}{r_2^3} \quad \text{and} \quad R_2 = \frac{Gm_1}{|\mathbf{r}_1 - \mathbf{r}_2|} - Gm_1\frac{\mathbf{r}_1 \cdot \mathbf{r}_2}{r_1^3} \tag{3.21}$$

are the disturbing functions representing the perturbations arising due to the other secondary object. In Equations (3.21), the first terms are known as the *direct terms*, while the second terms are known as the *indirect terms*. The latter arise due to the choice of the coordinate system: if the origin was chosen to be the system's centre of mass, then the indirect terms would not have appeared. This is the most straightforward definition of the disturbing function.

Finally, it is worthwhile to note that the definition of the disturbing function is general, and not constrained to the three-body problem considered here. In other words, it can be extended to account for any number of bodies, or even for perturbations resulting from other sources and not just from the gravity of point-masses e.g. the oblateness of the central mass. In this dissertation, we shall use the disturbing function to describe the perturbations arising due to planets, in addition to the gravity of radially extended massive debris discs.

3.2.2 Expanding the Disturbing Function

By definition, the disturbing function holds all the information related to the strength and effects of any perturbations beyond the Kepler problem. In order to analyse the resulting perturbations analytically, however, it is more useful to express the disturbing function in terms of the standard orbital elements (i.e., a, e, I, ϖ, Ω, and f) rather than the position vectors of the interacting bodies. As such, much effort has been put towards this end, resulting in the development of a multitude of analytical and computer-assisted methods of expanding the disturbing function (e.g. Plummer, 1918; Brouwer & Clemence, 1961; Murray & Dermott, 1999; Laskar & Boué, 2010; Mardling, 2013). Skipping over the details, the end result of such expansion techniques is that the disturbing function takes on the following very general form:

$$R_{1,2} = Gm_{2,1} \sum_{j_1,...,j_6} S(a_1, a_2, e_1, e_2, I_1, I_2) \cos\varphi. \tag{3.22}$$

Here, the quantity φ is a permitted linear combination of the mean anomalies, longitudes of pericentre, and longitudes of ascending node given by:

$$\varphi = j_1 M_1 + j_2 M_2 + j_3 \varpi_1 + j_4 \varpi_2 + j_5 \Omega_1 + j_6 \Omega_2, \tag{3.23}$$

and j_n (with $n = 1, ..., 6$) are integer numbers which obey the d'Alembert relation,

$$\sum_{n=1}^{6} j_n = 0, \tag{3.24}$$

stemming from the rotational invariance of the disturbing function. Thus in essence, the disturbing function is written as the sum over an infinite number of terms involving the standard orbital elements. For completeness, I now briefly discuss the two classical methods of expanding the disturbing function, each coming with its own strengths and weaknesses.

3.2.2.1 Laplace-Lagrange Expansion

One of the classical ways of expanding the disturbing function is the so-called Laplace-Lagrange theory. Historically, this method was first developed by Laplace and Lagrange around the end of the 18th century to study the orbital evolution of the Solar System planets. As such, this method relies on the smallness of the interacting bodies' eccentricities and inclinations – i.e., e and $I \approx 0$ – but holds for all semimajor axis ratios. Without going into the details, the general outcome of this method is that the

disturbing function now becomes expressed as an infinite series in powers of e and I, with coefficients that encapsulate the ratio of the semimajor axes via the so-called *Laplace coefficients* (their mathematical definition is given by Equation 4.2). This is the approach adopted in this dissertation to e.g. analyse the perturbations of a planetesimal due to a planet. This said, however, it must be noted that the full Laplace-Lagrange expansion of the disturbing function no longer converges for values of eccentricities larger than $e \gtrsim 0.66$: this is known as the Sundman criterion, which arises due to the divergence of the series solution of Kepler's equation (Equation 3.8) beyond this eccentricity value. Finally, while the Laplace-Lagrange expansion does not place any restrictions on the semimajor axis ratios, it becomes unreliable when there is any orbital crossing between the interacting bodies. This is because when two orbits cross at some longitude, $\mathbf{r}_1 = \mathbf{r}_2$ and the disturbing functions of either body – or equivalently, the Laplace coefficients – diverge; see Equation (3.21). This well-known problem is often alleviated in the literature by softening the point-mass Newtonian potential. In Chapter 4, I apply this expansion method to arbitrarily softened potentials and derive the disturbing function due to a massive disc to second order in eccentricities, applicable to both particulate or gaseous discs.

3.2.2.2 Legendre Expansion

The second method for expanding the disturbing function is that based on Legendre polynomials. This approach makes use of a Legendre expansion of the disturbing function in the ratio of semimajor axis, $a_1/a_2 \ll 1$, with no assumptions about the eccentricities and inclinations; which is in contrast with the Laplace-Lagrange expansion. As a result, within this framework, the disturbing function is written as an infinite series in powers of the ratio a_1/a_2, with coefficients that include the eccentricities and inclinations via the so-called *Hansen coefficients*. While this method has a better radius of convergence in terms of e and I when compared to the Laplace-Lagrange expansion, it converges slowly with respect to a_1/a_2. Consequently, the Legendre expansion is better suited for describing the dynamical evolution of *hierarchical* systems where there interacting bodies are well separated, but not for the purposes of this dissertation concerning the dynamics of adjacent massive planetesimals constituting a debris disc.

Needless to say, both the Laplace-Lagrange and Legendre expansions reproduce the very same expressions of the disturbing function upon Taylor expanding the corresponding expressions about the unrestricted orbital parameters; that is, a_1/a_2 in the Laplace-Lagrange expansion, and e and I in the Legendre expansion.

3.2.3 Lagrange's Planetary Equations

The expansion of the disturbing function as given by Equation (3.22) encapsulates the dependence of the perturbing potential on the orbital elements. This can then be employed to determine how the orbital elements of the perturbed body vary in time, which is what one is often interested in. To this end, one can either use a Hamiltonian approach[1] or equivalently, and as is often done in celestial mechanics, *Lagrange's planetary equations*. The complete set of Lagrange's planetary equations is given by[2]:

$$\frac{da}{dt} = \frac{2}{na}\frac{\partial R}{\partial \lambda}, \tag{3.25}$$

$$\frac{de}{dt} = -\frac{\sqrt{1-e^2}}{na^2 e}\left[\frac{\partial R}{\partial \varpi} + (1-\sqrt{1-e^2})\frac{\partial R}{\partial \lambda}\right], \tag{3.26}$$

$$\frac{d\varpi}{dt} = \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial R}{\partial e} + \frac{\tan(I/2)}{na^2\sqrt{1-e^2}}\frac{\partial R}{\partial I}, \tag{3.27}$$

$$\frac{dI}{dt} = -\frac{\tan(I/2)}{na^2\sqrt{1-e^2}}\left(\frac{\partial R}{\partial \lambda} + \frac{\partial R}{\partial \varpi}\right) - \frac{1}{na^2\sqrt{1-e^2}\sin(I)}\frac{\partial R}{\partial \Omega}, \tag{3.28}$$

$$\frac{d\Omega}{dt} = \frac{1}{na^2\sqrt{1-e^2}\sin(I)}\frac{\partial R}{\partial I}, \tag{3.29}$$

$$\frac{d\lambda}{dt} = n - \frac{2}{na}\frac{\partial R}{\partial a} + \frac{\sqrt{1-e^2}(1-\sqrt{1-e^2})}{na^2 e}\frac{\partial R}{\partial e} + \frac{\tan(I/2)}{na^2\sqrt{1-e^2}}\frac{\partial R}{\partial I}. \tag{3.30}$$

In Equations (3.25) – (3.30), $\lambda = M + \varpi$ is known as the *mean longitude* and is customarily introduced to ensure that the equations are well-behaved when $I = 0$ and/or $e = 0$ (akin to the definition of ϖ). Historically, these equations were first developed for the restricted three-body problem — i.e., $M_c \gg m_1 \gg m_2$ so that the orbit of m_1 is not perturbed – and as such, they are normally expressed in terms of a disturbing function with the dimensions of energy per unit mass. Finally, it is important to note that Lagrange's planetary equations are general, in the sense that they can be employed with any disturbing function, and not only that which arises from the (restricted) three-body problem.

[1] This can be easily achieved once the orbital elements are expressed in terms of canonical conjugate variables, e.g., Delaunay or Poincare variables. Equipped with such variables, it is straightforward to employ Hamilton's equations of motion, noting that the disturbing function is the negative of the true potential; see Equation (3.19).

[2] A derivation of Lagrange's planetary equations can be found in e.g. Brouwer & Clemence (1961).

3.2.4 Classification of Terms in the Disturbing Function

As discussed above, the disturbing function contains an infinite number of terms. Thus, a first look at Equation (3.22) may suggest that this is a rather complicated way for describing the orbital evolution of a given system, especially when using analytical methods. However, this is reconciled by the fact that the different terms in the disturbing function can be classified based on the frequencies associated with the cosine terms in the expansion, and thus the timescales over which they cause modulations of the orbital elements. This allows one to significantly limit the number of terms to be considered for a given astrophysical problem. Indeed, the full disturbing function describes different types of perturbations, which take one of three types: secular, resonant, and short-period. Although throughout this dissertation we are mainly interested in the secular terms, a brief discussion of all three types is necessary to better highlight my findings and potential limitations. This is the subject of this section.

3.2.4.1 Secular Perturbations

Secular terms in the disturbing function are defined as those which are independent of the mean anomalies (or mean longitudes). They typically have periods much longer than either orbital period, with typical timescales on the order of $(M_c/m)T$; and fractional amplitudes of order unity in eccentricities and inclinations, as measured relative to resonant and short-period terms. Thus, secular terms describe the long-term effects of perturbations and, generally, dominate the dynamical evolution of a system.

The main effect of secular perturbations is to redistribute angular momentum amongst the interacting bodies, without causing any exchange of orbital energies. This is a restatement of the fact that secular terms are independent of the interacting bodies' mean anomalies, and so their semimajor axes – and hence, individual energies – are conserved quantities; see Equation (3.25). In other words, the effects of secular perturbations boil down to the modulation of a body's orbital eccentricity and orientation over long timescales, but not its semimajor axis.

A commonly used approximation to model secular perturbations is the so-called *Laplace–Lagrange perturbation theory*, which essentially accounts for the secular terms in the disturbing function (3.22) to only second order in eccentricities and inclinations. Obviously, this approximation is only valid as long as the eccentricities and inclinations remain low; however, it is useful to describe the general features of secular interactions. Additionally, within this approximation, the evolution of e and ϖ is known to be

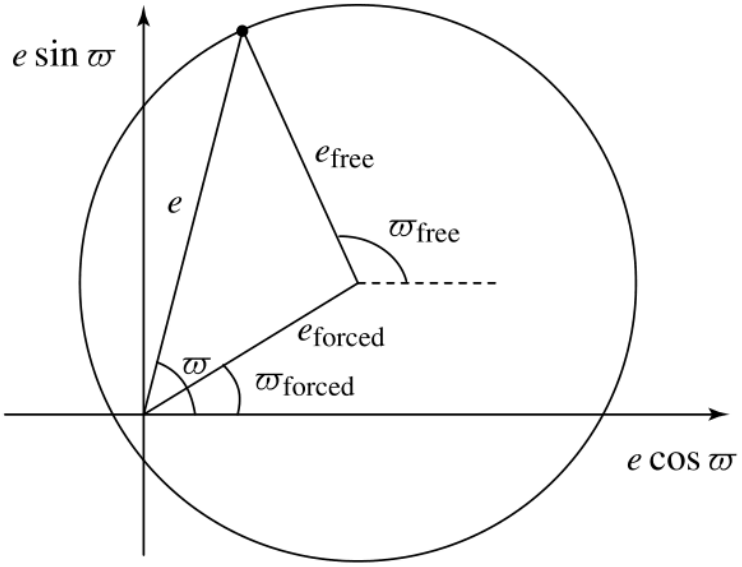

Figure 3.3. Geometrical interpretation of a particle's instantaneous eccentricity e and longitude of pericentre ϖ subjected to secular perturbations; see Section 3.2.4.1 for details.

independent from the evolution of I and Ω; allowing one to study the planar and vertical evolution separately. This is the approach I adopt throughout this dissertation to model the secular perturbations due to both planets and massive discs, and so mathematical details of the relevant disturbing functions, as well as the ensuing dynamical effects, will be presented and explained when called for in the upcoming chapters. Given this, here I only describe the generic solution obtained within the context of Laplace–Lagrange theory.

Laplace–Lagrange theory finds that the secular evolution of a particle can be decomposed into two time-varying components: the *forced* and *free* (or *proper*) elements. For instance, the coupled evolution of the eccentricity e and longitude of pericentre ϖ obey the following relationships:[3]

$$e\cos\varpi = e_{\rm forced}\cos\varpi_{\rm forced} + e_{\rm free}\cos\varpi_{\rm free}, \tag{3.31}$$

$$e\sin\varpi = e_{\rm forced}\sin\varpi_{\rm forced} + e_{\rm free}\sin\varpi_{\rm free}. \tag{3.32}$$

[3]The secular evolution of a particle's inclination and longitude of ascending node can also be described in a similar way; however, however, since I only consider coplanar setups in this thesis, I omit particular details on the subject.

Here, the free eccentricity $e_{\rm free}$ is a constant determined by the particle's initial conditions, and the free longitude of pericentre $\varpi_{\rm free}$ evolves linearly with time,

$$\varpi_{\rm free} = At + \beta_0, \tag{3.33}$$

where A and β_0 are constants; the former determined by the orbital parameters of the massive bodies in the system and the latter by the initial conditions. On the other hand, the forced eccentricity $e_{\rm forced}$ and longitude of pericentre $\varpi_{\rm forced}$ are determined solely by the orbital parameters of the interacting bodies and, depending on the forces at play (e.g., frictional forces), may evolve in time. This decomposition into free and forced components allows for a straightforward geometric interpretation of the particle's secular evolution: in a plane defined by $e(\cos\varpi, \sin\varpi)$, the particle moves around a circle of fixed radius $e_{\rm free}$ at a constant rate A, and the centre of this circle is defined by the forced elements which itself may evolve in time. The vectorial sum of the free and forced elements then yields the instantaneous values of e and ϖ. This is illustrated in Figure 3.3. Note that the particle can move in circles in the $e(\cos\varpi, \sin\varpi)$ plane either in clockwise or anti-clockwise direction: this is determined by the sign of A, which can be positive e.g. for a planetesimal perturbed by a planet, and negative e.g. for a planetesimal embedded in a self-gravitating disc. Further mathematical details on the various parameters appearing in Equations (3.31) – (3.33) – which are system-specific – will be given in subsequent chapters.

The above description suggest that secular perturbations would simply cause the eccentricities (and inclinations) of a perturbed body to vary sinusoidally with time, while the longitudes of pericentres (and ascending nodes) precess on timescales of $\sim A^{-1}$. The strongest effects of secular perturbations, however, manifest at locations where a *secular resonance* occurs. Secular resonances arise when there is an equality amongst the slow frequencies of orbital precession of the interacting bodies[4]. For instance, a *secular apsidal resonance* is said to occur when the apsidal precession rates of two interacting bodies are equal, i.e., $\dot{\varpi}_1 = \dot{\varpi}_2$. At the radial locations where this condition is satisfied, the eccentricity of a body can be driven to arbitrarily large values, i.e., $e \to 1$, as the forced eccentricity formally diverges[5]. I will make use of this fact

[4]Note that secular resonances are different from mean-motion resonances in which case the commensurability occurs between the orbital periods T (rather than the rates at which ϖ and/or Ω precess) – see Section 3.2.4.2.

[5]The divergence of the forced eccentricity is a result implicit to Laplace–Lagrange theory: in reality, including higher order terms in the disturbing function would limit the amplitude of $e_{\rm forced}$ (e.g., Malhotra, 1998).

in Chapters 5 and 6 to introduce a new mechanism for sculpting gaps, or depleted regions, in debris discs.

Finally, we thus far discussed the mathematical concepts behind secular interactions by making use of the disturbing function and ignoring the terms that are dependent on the mean anomalies. However, what is the physical meaning of the secular approximation? An insight into this question can be best obtained based on Gauss' averaging principle. According to this principle, the secular approximation can be thought of as smearing the mass of the perturbing object into a ring – or a wire — along its orbit. This material ring has a linear density which is inversely proportional to the body's instantaneous velocity at each position along its orbit. As such, the ring's density is non-uniform for particles on non-circular orbits, which spend most (least) of their time at the apocentre (pericentre). With this picture in mind, the perturbed body can be thought of as moving under the gravitational force exerted by the perturbing ring of material, and the resulting variations in its orbital elements can be obtained using Gauss' equations (rather than Lagrange's planetary equations). Application of this averaging principle reveals that the secular changes in the orbital elements of a body perturbed by another that is modelled as a ring is identical to that obtained by ignoring the terms in the disturbing function that depend on the mean anomalies. This provides a more physical explanation of secular perturbations, and represents the basic idea upon which Chapter 4 is based, specifically when it comes to computing the orbit-averaged disturbing function due to astrophysical discs.

3.2.4.2 Resonant Perturbations

Resonant terms in the disturbing function, on the other hand, are those which have a period of order the orbital periods T and explicitly depend on the mean anomalies (or mean longitudes) – characteristics in contrast to the secular terms. Additionally, they are characterised with fractional amplitudes on the order of $m/M_c \ll 1$, which are generally smaller than those of the secular terms. As a result, their effects are often negligible, with the secular terms dominating the long-term dynamical evolution. However, resonant terms become important – and even dominant – when the ratio of the orbital periods of the interacting bodies is equal or close to the ratio of two integers, so that the bodies exert periodic gravitational forces onto each other. When this condition is satisfied, the two bodies are said to be in a mean-motion resonance (MMR). Generally, a $(p+q):p$ MMR occurs when

$$\frac{n_2}{n_1} \approx \frac{p}{p+q}, \tag{3.34}$$

where p and q are positive integers, with $p \neq 0$. Here, the integer q represents the order of the resonance, giving a measure of its strength: generally, the higher the order, the weaker the MMR is. This is because the amplitude terms S in the disturbing function are such that $S \propto e^{|q|}$ and $e < 1$, and so higher-order MMRs can be ignored.

Mean-motion resonances can have several effects on the dynamical evolution of the bodies involved, as they mediate the exchange of both angular momentum and energy. The dynamical effects mainly depend on the conjunction of the two bodies, i.e., their point of closest approach to one another. In certain circumstances, mean-motion resonances can help the bodies to remain on stable orbits, shielding them e.g. from scattering events. This happens when although the orbits of the two bodies do cross each other, conjunctions occur at large enough separations that they never come close to one another along their orbits. Examples of such shielding MMRs can be found in the Solar System, with the $3:2$ MMR between Pluto and Neptune, and the $1:1$ MMR – also known as co-orbital resonance – of the Trojan asteroids of Jupiter which share the planet's orbit but are situated $\approx 60^\circ$ from it. In other cases, however, mean-motion resonances may drive the system into dynamical instabilities. This happens when repeated conjunctions at the same point in the orbits of the interacting bodies lead to a significant exchange of angular momentum, gradually putting them onto unstable orbits. A notable example of this can be seen in the Kirkwood gaps of the asteroid belt, which demarcate the locations of MMRs with Jupiter.

Finally, I note that MMRs – unlike secular perturbations – are localised to a narrow range of semimajor axis and have finite radial widths over which they operate. However, when a planetesimal is orbiting very close to a planet, MMRs can become densely packed to the extent that their individual widths overlap each other. In such cases, the planet can drive planetesimals orbiting (or entering) in this region into chaotic motion, ejecting them from the system over timescales ranging from $\sim 10^3$ to 10^6 years. The half-width of the chaotic zone (Δa_p) surrounding the planetary orbit, as a first order approximation, is given by (Wisdom, 1980; Duncan et al., 1989):

$$\Delta a_p \approx 1.3 \left(\frac{m_p}{M_c + m_p} \right)^{2/7} a_p, \tag{3.35}$$

where m_p and a_p are the planet's mass and semimajor axis, respectively, and M_c is the central star's mass.

3.2.4.3 Short-Period Perturbations

Finally, all other terms in the disturbing function that are neither secular nor resonant are classified as short-period terms. Generally speaking, such terms can be safely dropped from the disturbing function as the associated perturbations average out to zero over orbital periods, and so the dynamical evolution will remain well-captured by the secular and/or resonant terms – depending on the physical setup. However, this is not justified in all cases; for instance, in the events of close encounters and scattering.

When two bodies orbiting a central body of mass M_c pass close to each other, their gravitational influence onto each other can become large enough to be comparable to the gravitational pull from the central star. Mathematically, this can be quantified by the so-called Hill radius R_H which, to lowest order in eccentricities, is given by:

$$R_H = \frac{1}{2}\left(\frac{m_1 + m_2}{3M_c}\right)^{1/3}(a_1 + a_2), \tag{3.36}$$

where, as before, m_i and a_i $(i = 1, 2)$ are the masses and semimajor axes of the two bodies, respectively. In such cases, the short-period terms in the disturbing function may become dominant, causing significant changes to each individual body's orbital energy and angular momentum and, as a result, affecting all of the six orbital elements associated with its orbit. This change happens over timescales much shorter than the orbital periods that it can be considered as an instantaneous velocity 'kick'. A series of such velocity kicks could eventually lead to the chaotic diffusion of the scattered particle's orbit over timescales given by (Tremaine, 1993):

$$t_{\text{diff}} \sim 0.01\frac{1}{n_p}\left(\frac{a_p}{a}\right)^{1/2}\left(\frac{m_p}{M_c}\right)^{-2}, \tag{3.37}$$

where quantities sub-scripted with 'p' refer to the planet (the scatterer). Eventually, over long timescales, the particle's eccentricity will increase until it reaches values larger than unity and its orbit is unbound from the system. Such scattering events between a planet and a massive enough swarm of planetesimals can also significantly affect the orbit of the planet, which may migrate as it becomes more circular due to exchange of orbital energy and angular momentum – as is postulated for e.g. Neptune and its interactions with the primordial Kuiper belt.

Chapter 4

Potential Softening and Eccentricity Dynamics in Razor-thin, nearly Keplerian Discs

This chapter presents my own research work, where I investigate the various methods found in the literature for the computation of the long-term, secular effects of astrophysical discs. The work presented here has been published in Monthly Notices of the Royal Astronomical Society as Sefilian & Rafikov (2019).

4.1 Introduction and Problem Statement

Astrophysical discs orbiting a central mass M_c are ubiquitous in a variety of contexts – galactic, stellar, and planetary (Latter et al., 2017). In many instances, masses of such discs M_d are much less than the central object mass, $M_d \ll M_c$. Despite this fact, gravity of such discs can still play an important dynamical role in the orbital evolution of their constituent particles as well as the dynamics of external objects (e.g. Goldreich & Tremaine, 1979; Heppenheimer, 1980; Ward, 1981; Kocsis & Tremaine, 2011; Kazandjian & Touma, 2013; Teyssandier et al., 2013; Meschiari, 2014; Silsbee & Rafikov, 2015a; Petrovich et al., 2019; Sefilian & Touma, 2019). Consequently, characterising the dynamical effects due to disc gravity is important.

Whenever $M_d \ll M_c$, particles perturbed by the disc gravity move on nearly Keplerian orbits that evolve rather slowly. This justifies the use of the so-called *secular approximation* which implies averaging of the fast-evolving dynamical variables over the orbits of particles under consideration (Murray & Dermott, 1999, see also Chapter

3). As already explained in Section 3.2.4.1, the orbit-averaging procedure – also known as Gauss' method – is equivalent to calculating the time-averaged potential due to orbiting point masses by smearing them into massive elliptical "wires" (having shape of their eccentric orbits) with non-uniform linear density proportional to the time spent by an object at a particular phase of its orbit. Such orbit-averaged potential, also known as *secular disturbing function* R_d, fully determines the secular dynamics of the system.

For a test particle with semimajor axis a_p, eccentricity e_p, and apsidal angle ϖ_p due to a co-planar point mass δm_d orbiting with semimajor axis a, eccentricity e_d, and apsidal angle ϖ_d, upon smearing into elliptical rings, the secular disturbing function takes the following form (Murray & Dermott, 1999):

$$\delta R = \frac{G\delta m_d a_p}{a^2}\left[\frac{1}{8}b_{3/2}^{(1)}\left(\frac{a_p}{a}\right)e_p^2 - \frac{1}{4}b_{3/2}^{(2)}\left(\frac{a_p}{a}\right)e_p e_d\cos(\varpi_p - \varpi_d)\right], \tag{4.1}$$

which is valid for both $a > a_p$ and $a < a_p$, as long as particle orbits do not cross. Here $b_s^{(m)}(\alpha)$ is the Laplace coefficient defined by

$$b_s^{(m)}(\alpha) = \frac{2}{\pi}\int_0^\pi \cos(m\theta)\left[1+\alpha^2-2\alpha\cos\theta\right]^{-s}d\theta, \tag{4.2}$$

which obeys $b_s^{(m)}(\alpha^{-1}) = \alpha^{2s}b_s^{(m)}(\alpha)$. Explicit time independence of δR guarantees that the semimajor axes of the secularly interacting objects stay fixed, as already explained in Section 3.2.4.1.

When considering gravitational effects of a razor-thin *continuous* disc with smooth distribution of surface density, a straightforward way to compute the secular disturbing function would be to orbit-average the disc potential (obtained by direct integration over its full surface) along the particle orbit. However, this procedure involves a triple integration – namely, a two-dimensional integral over the disc surface followed by orbit averaging – and is numerically challenging.

A more efficient approach lies in representing the disc as a collection of massive, nested, confocal elliptical 'wires' (also referred to as 'annuli' or 'rings' in this chapter) with fixed semimajor axes (e.g. Touma et al., 2009; Batygin, 2012). For reference, a graphical representation of this approach is provided in Figure 4.1. Within this approach then, given the additive nature of gravity, the disturbing function due to the disc can be represented as a sum of individual contributions in the form of Equation (4.1) produced by all wires. In the continuum limit, this amounts to integration of δR

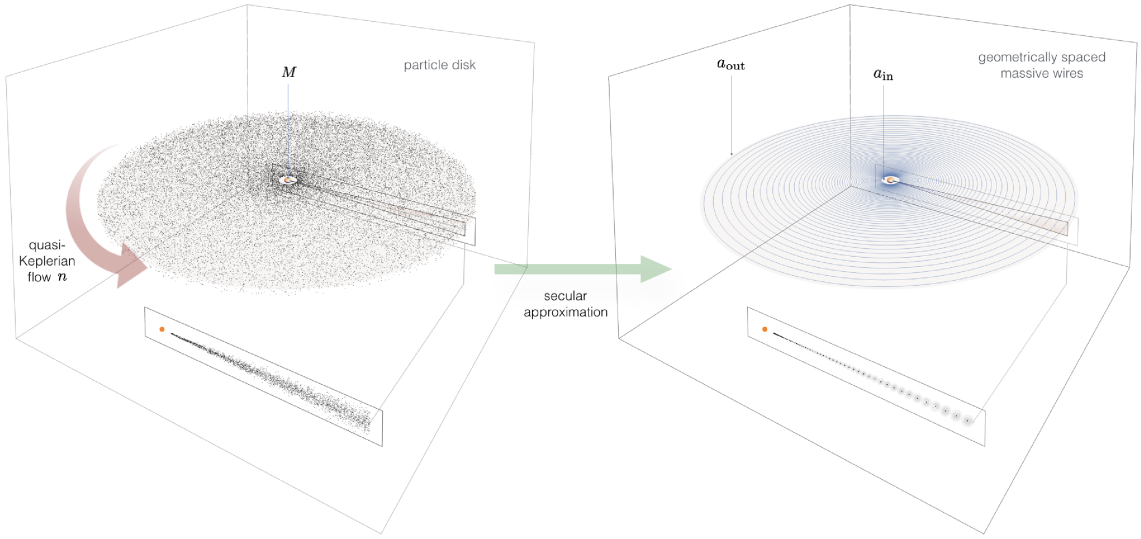

Figure 4.1. Graphical representation of a disc within the secular approximation. A nearly Keplerian disc composed of $N \gg 1$ particles is modelled as a collection of geometrically spaced massive rings (or wires). This diagram has been adapted from Batygin (2018).

(Equation 4.1) over the radial extent of the disc:

$$R_d = \int_{a_{\text{in}}}^{a_{\text{out}}} \delta R, \tag{4.3}$$

where a_{in} and a_{out} are the semimajor axes of the inner and outer disc edges, respectively. In this case, provided that δR is known as a function of a, only a single integration (over the semimajor axes of the rings) is needed, significantly accelerating calculations[1].

Unfortunately, this straightforward procedure is ill-posed from the mathematical point of view. Indeed, it is well known that the Laplace coefficients $b_{3/2}^{(m)}$ featured in Equation (4.1) diverge as $b_{3/2}^{(m)}(\alpha) \to (1-\alpha)^{-2}$ when $\alpha \to 1$. This implies that the radial integration in Equation (4.3) encounters a non-integrable singularity at $a = a_p$. As a result, for a co-planar particle orbiting inside a razor-thin disc, $a_{\text{in}} \leq a_p \leq a_{\text{out}}$, this direct way of computing R_d does not converge to a finite value.

This divergence, as well as the pressing need for having an efficient way of computing R_d (via a one-dimensional integration over a only), have motivated the development of alternative analytic approaches for calculating R_d. These approaches can be generally grouped into two classes. Calculations of one kind are rooted in the derivation of the potential of an axisymmetric disc with power law surface density profile presented in

[1]The Laplace coefficients entering in δR can be easily evaluated, without relying on integration over θ in Equation (4.2), by expressing them through elliptic integrals, see Appendix C.3.

Heppenheimer (1980), which does not suffer from the singularity of Laplace-Lagrange secular theory. A number of subsequent studies used this approach (Ward, 1981) and extended it to the case of eccentric discs, both apsidally aligned (Silsbee & Rafikov, 2015a; Davydenkova & Rafikov, 2018) and misaligned (Davydenkova & Rafikov, in prep.). Higher order (in eccentricity) extensions of this approach have also been developed (Sefilian & Touma, 2019). This framework for treating secular dynamics has been extensively verified using direct orbit integrations under different conditions (Silsbee & Rafikov, 2015a; Fontana & Marzari, 2016; Davydenkova & Rafikov, 2018). In this chapter, I refer to this type of calculation as the unsoftened *Heppenheimer's method*.

Unfortunately, by construction Heppenheimer's method is inapplicable in situations where the disc eccentricity rapidly varies with semimajor axis, potentially resulting in orbit crossings (Davydenkova & Rafikov, 2018). An alternative approach, which avoids this problem, while at the same time alleviating the aforementioned singularity, is to use *softened gravity* by spatially smoothing the Newtonian point-mass potential in various ways – both analytically (e.g. Tremaine, 1998, 2001; Touma, 2002; Hahn, 2003; Touma & Sridhar, 2012; Teyssandier & Ogilvie, 2016) and numerically (e.g. Touma et al., 2009). In these models, the classical Laplace-Lagrange disturbing function (Equation 4.1) is modified by *softening* the interaction potential in some way to circumvent the divergence of R_d as $a \to a_p$. In this method orbit crossing does not lead to problems as long as the softening scale is finite. However, a physical justification for a specific form of softening (absent in the Heppenheimer (1980) approach) often remains unclear, making the introduction of softening rather arbitrary.

The primary goal of this chapter is to assess how well the different calculations relying on potential softening reproduce secular dynamics driven by the gravity of a razor-thin disc. The main metric I use in this exercise is the convergence of the results of such calculations to the true secular evolution (represented by the unsoftened Heppenheimer method) in the limit of vanishing softening, when the limit of Newtonian gravity is recovered. Complementary to this, I develop a general framework for computing the well-behaved secular disturbing function for a broad range of softened gravitational potentials (Appendix A).

The rest of this chapter is organised as follows. First, I describe the general analytical expressions governing the orbit-averaged potential due to a coplanar disc of arbitrary structure and arbitrary softening prescription in Section 4.2. I then provide a brief review of the different softened potentials examined in this work and the unsoftened approach of Heppenheimer in Sections 4.2.1 and 4.2.2, respectively. I

analyse the performance of softening formalisms in reproducing the correct secular dynamics for various disc models in Sections 4.3, 4.4, and 4.5. Finally, I discuss and briefly summarise my results in Sections 4.6 and 4.7, respectively. Technical details of my calculations can be found in Appendices A–D.

4.2 Disturbing Function due to a Disc

Prior to providing the details of different softening prescriptions examined in this chapter in Section 4.2.1, I briefly summarise some of their common features. The ultimate goal of all these prescriptions is the calculation of the disturbing function R_d due to gravity of a (generally eccentric) disc comprised of massive objects (stars, planetesimals, ring particles) or fluid elements (in gaseous discs) moving on Keplerian orbits.

Let us consider the disc to be razor-thin and coplanar. Mass distribution of such a disc can be uniquely characterised by the mass density per unit semimajor axis $\mu_d(a)$, eccentricity $e_d(a)$, and apsidal angle $\varpi_d(a)$ of the trajectories of its constituent elements, as functions of the semimajor axis a. In practice, it is often convenient to use the surface density at periastron $\Sigma_d(a)$ instead of $\mu_d(a)$; its relation to μ_d for arbitrary profiles of e_d and ϖ_d has been established in Statler (2001), Davydenkova & Rafikov (2018) and Davydenkova & Rafikov (in preparation). Constancy of semimajor axis in secular theory implies that $\mu_d(a)$ does not change in time. The same statement is true for $\Sigma_d(a)$ to lowest order in e_d since $\mu_d(a) \approx 2\pi a \Sigma_d(a) + O(e_d)$ (Davydenkova & Rafikov, 2018).

Close inspection of the various softening methods for computing secular disc potential (Section 4.2.1) reveals that all of them arrive at the following general form of the disturbing function for a test particle moving on an orbit with the semimajor axis a_p, eccentricity e_p, and apsidal angle ϖ_p:

$$R_d = n_p a_p^2 \left[\frac{1}{2} \mathsf{A}_d(a_p) \mathbf{e}_p^2 + \mathbf{B}_d(a_p) \cdot \mathbf{e}_p \right]. \tag{4.4}$$

Here, n_p is the test particle mean motion ($n_p^2 = GM_c/a_p^3$), and I have introduced a two-component eccentricity vector for a test particle $\mathbf{e}_p = e_p(\cos\varpi_p, \sin\varpi_p)$, as is customary in celestial mechanics.

The coefficients A_d and $\mathbf{B}_d$ in Equation (4.4) are related to the disc mass (or surface density) and eccentricity profiles in the following fashion:

$$\mathsf{A}_d(a_p) = \frac{2G}{n_p a_p^3} \times \left[\int_{a_{\rm in}}^{a_p} \mu_d(a) \phi_{22}\left(\frac{a}{a_p}\right) da + \int_{a_p}^{a_{\rm out}} \mu_d(a) \frac{a_p}{a} \phi_{11}\left(\frac{a_p}{a}\right) da \right], \tag{4.5}$$

$$\mathbf{B}_d(a_p) = \frac{G}{n_p a_p^3} \times \left[\int_{a_{\rm in}}^{a_p} \mu_d(a) \mathbf{e}_d(a) \phi_{12}\left(\frac{a}{a_p}\right) da + \int_{a_p}^{a_{\rm out}} \mu_d(a) \mathbf{e}_d(a) \frac{a_p}{a} \phi_{12}\left(\frac{a_p}{a}\right) da \right], \tag{4.6}$$

where $\mathbf{e}_d = e_d(a)(\cos\varpi_d(a), \sin\varpi_d(a))$ is the eccentricity vector for an annular disc element[2]. Note that for discs which are apse-aligned, i.e., $d\varpi_d/da = 0$ – as assumed later in this chapter – one has $\mathbf{B}_d = \mathsf{B}_d(\cos\varpi_d, \sin\varpi_d)$.

Functions $\phi_{ij}(\alpha)$, $i, j = 1, 2$, entering these expressions fully characterise the *softened* ring-ring secular interaction, see Equation (4.11). They are unique for each potential softening prescription, with explicit forms for the models that I explore in this chapter specified in Table 4.1. This Table shows that coefficients ϕ_{ij} appearing in the literature are linear combinations of *softened Laplace coefficients* $\mathcal{B}_s^{(m)}$ defined by

$$\mathcal{B}_s^{(m)}(\alpha, \epsilon) = \frac{2}{\pi} \int_0^\pi \cos(m\theta) \left[1 + \alpha^2 - 2\alpha\cos\theta + \epsilon^2(\alpha)\right]^{-s} d\theta. \tag{4.7}$$

The *softening parameter* $\epsilon(\alpha)$ appearing in this definition remains non-zero as $\alpha \to 1$, thus preventing the divergence of the softened Laplace coefficients $\mathcal{B}_s^{(m)}(\alpha, \epsilon)$ at $\alpha = 1$ (unlike the classical $b_s^{(m)}(\alpha)$). The explicit form of $\epsilon(\alpha)$ is different for every softening method considered in this chapter, see Section 4.2.1 and Table 4.1. Appendix C collates some useful relations for softened Laplace coefficients $\mathcal{B}_s^{(m)}(\alpha, \epsilon)$, as well as their approximate asymptotic behaviour and relationships to complete elliptic integrals.

The mathematical structure of R_d given by Equation (4.4) is similar to that of the classical Laplace-Lagrange planetary theory (Murray & Dermott, 1999), see Equation (4.1). Indeed, let us consider mass distribution of a point mass smeared along an elliptical orbit, $\mu_d(a) \to m_{\rm pl}\delta(a - a_{\rm pl})$ (where $\delta(z)$ is the Dirac delta-function), and set softening to zero (so that $\mathcal{B}_s^{(m)}(\alpha, \epsilon \to 0) \to b_s^{(m)}(\alpha)$). Then one finds that R_d reduces to the unsoftened, orbit-averaged potential δR due to a planet with mass $m_{\rm pl}$ and semimajor axis $a_{\rm pl}$, with the *unsoftened* coefficients ϕ_{ij} in the form (Murray & Dermott,

[2] I refer the reader to Heppenheimer (1980); Silsbee & Rafikov (2015a); Davydenkova & Rafikov (2018) for the expressions of A_d and $\mathbf{B}_d$ computed using the unsoftened Heppenheimer method for different disc models.

1999):

$$\phi_{11}^{\rm LL}(\alpha) = \phi_{22}^{\rm LL}(\alpha) = \frac{1}{8}\alpha b_{3/2}^{(1)}(\alpha), \tag{4.8}$$

$$\phi_{12}^{\rm LL}(\alpha) = -\frac{1}{4}\alpha b_{3/2}^{(2)}(\alpha), \tag{4.9}$$

see Equation (4.1).

Accordingly, it is intuitive to think of Equations (4.4)–(4.6) as the continuous version of classical Laplace-Lagrange planetary theory, modified by the introduction of non-zero softening parameter ϵ to avoid the mathematical divergence of the classical disturbing function as $a \to a_p$.

I emphasise that the functional forms of ϕ_{ij} are not simple replacements of $b_s^{(m)}$ appearing in the unsoftened definition (4.8) - (4.9) by $\mathcal{B}_s^{(m)}$. This can be seen in Table 4.1, where I summarise some of the expressions for $\phi_{ij}(\alpha)$ proposed in the literature and analysed in this paper (see Section 4.2.1). Nevertheless, examination of these expressions shows that when $\epsilon^2(\alpha) \to 0$, the coefficients $\phi_{ij}(\alpha)$ do reduce to their unsoftened versions $\phi_{ij}^{\rm LL}(\alpha)$ given by Equations (4.8) – (4.9).

In Appendix A, I show that the form of the disturbing function given by Equations (4.4)–(4.6) is generic for a wide class of softening models (and not just the ones covered in Section 4.2.1), for which the interaction potential between the two masses m_1 and m_2 ($m_i \ll M_c$) located at $\mathbf{r}_1$ and $\mathbf{r}_2$, correspondingly, relative to the central mass, has a form[3]

$$\Phi_i(\mathbf{r}_1, \mathbf{r}_2) = -Gm_j\Big[(\mathbf{r}_1 - \mathbf{r}_2)^2 + \mathcal{F}(r_1, r_2)\Big]^{-1/2}, \tag{4.10}$$

with $i, j = 1, 2$ and $j \neq i$. Here $\mathcal{F}(r_1, r_2)$ represents an arbitrary softening function introduced to cushion the singularity which arises otherwise at null interparticle separations. Note that in general this potential may depend not only on the relative distance between the two masses $\mathbf{r}_1 - \mathbf{r}_2$, but also on their distances to the dominant central mass r_1 and r_2, respectively.

Explicit demonstration of the connection between the potential (4.10) and R_d given by Equation (4.4) represents a stand-alone result of this chapter. In particular, my calculations in Appendix A, which can be skipped at first reading, show that the softening parameter ϵ featured in the definition (4.7) is related to $\mathcal{F}$ via $\epsilon^2 = [\max(a_1, a_2)]^{-2}\mathcal{F}(a_1, a_2)$, where $a_{1,2}$ are the semimajor axes of the interacting particles (see Equation A.21). The most general expressions of ϕ_{ij} entering the arbitrarily

[3]Note that the interparticle force resulting from such potential does not, in general, obey Newton's third law (as long as $\mathcal{F}(r_1, r_2) \neq$ const).

softened ring-ring disturbing function,

$$R_i = \frac{Gm_j}{a_>}\left[\phi_{11}(\alpha)\ e_1^2 + \phi_{22}(\alpha)\ e_2^2 + \phi_{12}(\alpha)\ e_1 e_2 \cos(\varpi_1 - \varpi_2)\right], \tag{4.11}$$

(here $i = 1, 2$ and $j \neq i$) is given by Equations (A.22)–(A.24) in terms of $\mathcal{B}_s^{(m)}(\alpha, \mathcal{F})$. In the above expression, I have defined $a_> = \max(a_1, a_2)$ and $a_< = \min(a_1, a_2)$ such that[4] $\alpha = a_</a_>$.

Note that in equations (4.5) and (4.6) I split integration over a in two parts: over the part of the disc interior to a_p, and exterior to it. I do this because for some softening functions $\mathcal{F}$ the coefficients $\phi_{ij}(\alpha)$ do not obey certain symmetry properties when a/a_p is replaced with a_p/a, see Equation (C.4). Moreover, in general ϕ_{11} and ϕ_{22} are not necessarily identical as in classical Laplace-Lagrange theory (i.e., Equation 4.8); see Table 4.1 and Appendix A for further details.

As to the physical meaning of A_d and $\mathbf{B}_d$, I remind the reader that A_d represents the precession rate of the free eccentricity vector of a test particle in the disc potential, while $\mathbf{B}_d$ characterises the torque exerted on the particle orbit by the non-axisymmetric component of the disc gravity. Corresponding forced eccentricity vector is $\mathbf{e}_{p,f} = -\mathbf{B}_d/A_d$. In particular, test-particles initiated on circular orbits experience eccentricity oscillations of maximum amplitude $e_p^m = 2\,|\mathbf{e}_{p,f}|$.

Finally, note that since $\mathsf{A}_d(a_p)$ and $\mathbf{B}_d(a_p)$ uniquely determine R_d for different forms of softening, comparison of their behaviour in the limit of $\epsilon \to 0$ with that found in the unsoftened Heppenheimer (1980) approach (validated in Silsbee & Rafikov 2015a; Fontana & Marzari 2016; Davydenkova & Rafikov 2018) is sufficient to assess the validity of a particular softening model, see Section 4.3.

4.2.1 Summary of Existing Softening Models

Here I provide a brief description of the four different softening prescriptions that have been previously proposed in the literature. Corresponding expressions for their softening parameters $\epsilon^2(\alpha)$ and coefficients $\phi_{ij}(\alpha)$ are provided in Table 4.1.

[4]Here I clarify that the definitions of $\phi_{11}(\alpha)$ and $\phi_{22}(\alpha)$, even when different (see Table 4.1 and Appendix A), are swapped upon interchanging a_1 with a_2 but keeping, by construction, $\alpha = a_</a_> < 1$ – see Equations (A.22) and (A.23) for details.

4.2.1.1 Formalism of Tremaine (1998) – Tr98

Tremaine (1998) suggested an expression for the secular disturbing function due to a continuous disc, which uses modified Laplace coefficients in the form

$$\mathcal{B}_s^{(m),\mathrm{Tr}} = \frac{2}{\pi}\int_0^\pi \cos(m\theta)\left[1+\alpha^2-2\alpha\cos\theta+\beta_c^2\right]^{-s} d\theta. \tag{4.12}$$

Here β_c^2 is the dimensionless softening parameter, treated as a constant, i.e., independent of distance. The physical interpretation of this manoeuvre is that β_c, inhibiting the formal divergence of R_d as $a \to a_p$, can be viewed as the disc aspect ratio. Within this prescription, it is intuitive to think of the eccentric 'wires' that comprise the disc as having a distance-dependent radius $b = \beta_c \max(a_1, a_2)$. In Tremaine (1998) coefficients $\phi_{ij}(\alpha)$ were expressed as derivatives of $\mathcal{B}_{1/2}^{(m),\mathrm{Tr}}$ with respect to α, see equations (26) of Tremaine (1998). These expressions, along with their versions modified using the recursive relations for Laplace coefficients (see Appendix C.1), can be found in Table 4.1.

4.2.1.2 Formalism of Touma (2002) – T02

Touma (2002) derived the orbit-averaged potential of a disc by assuming individual particles comprising the disc to interact via Plummer potential with a fixed length scale b_c (Binney & Tremaine, 2008). Smearing particles into gravitating eccentric wires, Touma (2002) (see also Touma & Sridhar, 2012) derived the expressions (equations (6) of Touma (2002)) for $\phi_{ij}(\alpha)$ in the form of linear combinations of softened Laplace coefficients $\mathcal{B}_s^{(m),\mathrm{T}}$, similar to those of Tremaine (1998):

$$\mathcal{B}_s^{(m),\mathrm{T}} = \frac{2}{\pi}\int_0^\pi \cos(m\theta)\left[1+\alpha^2-2\alpha\cos\theta+\beta^2\right]^{-s} d\theta. \tag{4.13}$$

However, in Touma (2002) the softening parameter $\epsilon^2(\alpha) = \beta^2$ is no longer a constant but depends on the distance such that $\beta = b_c/\max(a_1, a_2)$. Within this formalism, one can think of a disc as comprised of nested annuli with a constant thickness b_c.

4.2.1.3 Formalism of Hahn (2003) – H03

Hahn (2003) computed the orbit-averaged interaction between two eccentric wires by accounting for their vertical thickness. The vertical extent h of a ring effectively softens its gravitational potential over a dimensionless scale $H \sim h/a$, which was assumed to

be constant in that work (see also Ward, 1989). Hahn (2003) demonstrated that the resultant $\phi_{ij}(\alpha)$ are functions of softened Laplace coefficients

$$\mathcal{B}_s^{(m),\mathrm{H}} = \frac{2}{\pi}\int_0^\pi \cos(m\theta)\Big[1+\alpha^2-2\alpha\cos\theta+H^2(1+\alpha^2)\Big]^{-s}d\theta, \tag{4.14}$$

with constant $H \ll 1$. In other words, the softening parameter is given by $\epsilon^2(\alpha) = H^2(1+\alpha^2)$ in that work. The explicit expressions for $\phi_{ij}(\alpha)$ in terms of $\mathcal{B}_s^{(m),\mathrm{H}}$ are given by equations (17) of Hahn (2003).

4.2.1.4 Formalism of Teyssandier & Ogilvie (2016) – TO16

Teyssandier & Ogilvie (2016) modified the unsoftened expressions (4.8), (4.9) for $\phi_{ij}^{\mathrm{LL}}(\alpha)$ by simply replacing the usual Laplace coefficients $b_s^{(m)}$ with softened versions defined such that

$$\mathcal{B}_s^{(m),\mathrm{TO}} = \frac{2}{\pi}\int_0^\pi \cos(m\theta)\Big[1+\alpha^2-2\alpha\cos\theta+S^2\alpha\Big]^{-s}d\theta. \tag{4.15}$$

Thus, their softening parameter is $\epsilon^2(\alpha) = S^2\alpha$, where S is a dimensionless constant. According to the authors, this substitution approximates the process of vertical averaging over the disc with constant aspect ratio S, and alleviates the classical singularity. The corresponding expressions for $\phi_{ij}(\alpha)$ are given by equations (7)-(9) of Teyssandier & Ogilvie (2016).

The aforementioned softening prescriptions have their softening parameters $\epsilon^2(\alpha)$ controlled by different constants — β_c, b_c, H, and S. For this reason, in what follows – with some abuse of notation – I will collectively refer to these constants as "*softening parameters*" and denote them by ς.

4.2.2 The Unsoftened Heppenheimer Method

A different approach to computing the disturbing function of a razor-thin disc has been developed by Heppenheimer (1980) *without* resorting to any form of softened gravity (see also Ward, 1981). The essence of this method is in computing the potential by direct integration over the disc surface before expanding the integral limits (which involve instantaneous particle position r) in terms of small eccentricity of a test particle[5]. This expansion is followed by time-averaging over the orbit of a test particle.

[5]Note that the order of these procedures is opposite to what is usual in the Laplace-Lagrange treatment (e.g. Murray & Dermott, 1999). For further details, see e.g. Heppenheimer (1980).

The outcome of this procedure is a set of expressions, akin to Equations (4.4)–(4.6), which are convergent throughout the disc, in contrast to the classical Laplace-Lagrange theory. Mathematically, this convergent behaviour is due to the fact that the emergent expressions contain Laplace coefficients $b_{1/2}^{(m)}(\alpha)$ – and not $b_{3/2}^{(m)}$ – which diverge only weakly (logarithmically) as $\alpha \to 1$: $b_{1/2}^{(m)}(\alpha) \propto \log(1-\alpha)$. As a result, upon integrating these expressions over the radial extent of the disc, one obtains a convergent and finite result for R_d. Physically, convergent expression is only natural since the calculation of the disc potential by direct two-dimensional integration over its surface is fully convergent at every point in the disc. The Heppenheimer method simply allows one to properly capture this property, unlike the standard Laplace-Lagrange procedure (when applied to continuous discs).

In his pioneering calculation, Heppenheimer (1980) applied this method to axisymmetric power-law discs to recover the orbit-averaged disc potential to second order in eccentricities. This calculation has been subsequently extended to more general disc structures (Silsbee & Rafikov, 2015a; Davydenkova & Rafikov, 2018) (hereafter, SR15 and DR18, respectively), as well as to higher order in eccentricities (Sefilian & Touma, 2019). This framework has been extensively verified for eccentric discs using direct integrations of test particle orbits in actual disc potentials (e.g. SR15, Fontana & Marzari, 2016, DR18), validating this approach.

Table 4.1. The coefficients $\phi_{ij}(\alpha)$ (with $i, j = 1,\ 2$) of the secular disturbing function with softened gravity featured in Equations (4.5) and (4.6), which govern the individual secular ring-ring interaction (Equation 4.11), adopted from the literature (listed in the first column).

Formalism	$\epsilon^2(\alpha)$	$\phi_{11}(\alpha)$	$\phi_{12}(\alpha)$	$\phi_{22}(\alpha)$
Laplace-Lagrange	–	$\frac{1}{8}\alpha b^{(1)}_{3/2}$	$-\frac{1}{4}\alpha b^{(2)}_{3/2}$	ϕ_{11}
Tremaine (1998)	β_c^2	$\frac{1}{8}\left(2\alpha\frac{d}{d\alpha}+\alpha^2\frac{d^2}{d\alpha^2}\right)\mathcal{B}^{(0),\mathrm{Tr}}_{1/2}$ $=\frac{1}{8}\alpha\left[\mathcal{B}^{(1),\mathrm{Tr}}_{3/2}-3\alpha\beta_c^2\mathcal{B}^{(0),\mathrm{Tr}}_{5/2}\right]$	$\frac{1}{4}\left(2-2\alpha\frac{d}{d\alpha}-\alpha^2\frac{d^2}{d\alpha^2}\right)\mathcal{B}^{(1),\mathrm{Tr}}_{1/2}$ $=-\frac{1}{4}\alpha\left[\mathcal{B}^{(2),\mathrm{Tr}}_{3/2}-3\alpha\beta_c^2\mathcal{B}^{(1),\mathrm{Tr}}_{5/2}\right]$	ϕ_{11}
Touma (2002)	$\beta^2=b_c^2/a_>^2$	$-\frac{5}{8}\alpha\mathcal{B}^{(1),\mathrm{T}}_{3/2}+\frac{3}{16}\alpha^2\mathcal{B}^{(0),\mathrm{T}}_{5/2}+\frac{3}{8}\alpha(1+\alpha^2)\mathcal{B}^{(1),\mathrm{T}}_{5/2}$ $-\frac{15}{16}\alpha^2\mathcal{B}^{(2),\mathrm{T}}_{5/2}-\frac{3}{8}\alpha\beta^2(\alpha\mathcal{B}^{(0),\mathrm{T}}_{5/2}-\mathcal{B}^{(1),\mathrm{T}}_{5/2})$	$\frac{9}{8}\alpha\mathcal{B}^{(0),\mathrm{T}}_{3/2}+\frac{1}{8}\alpha\mathcal{B}^{(2),\mathrm{T}}_{3/2}-\frac{9}{8}\alpha(1+\alpha^2)\mathcal{B}^{(0),\mathrm{T}}_{5/2}$ $+\frac{21}{16}\alpha^2\mathcal{B}^{(1),\mathrm{T}}_{5/2}+\frac{3}{8}\alpha(1+\alpha^2)\mathcal{B}^{(2),\mathrm{T}}_{5/2}+\frac{3}{16}\alpha^2\mathcal{B}^{(3),\mathrm{T}}_{5/2}$	$-\frac{5}{8}\alpha\mathcal{B}^{(1),\mathrm{T}}_{3/2}+\frac{3}{16}\alpha^2\mathcal{B}^{(0),\mathrm{T}}_{5/2}+\frac{3}{8}\alpha(1+\alpha^2)\mathcal{B}^{(1),\mathrm{T}}_{5/2}$ $-\frac{15}{16}\alpha^2\mathcal{B}^{(2),\mathrm{T}}_{5/2}-\frac{3}{8}\beta^2(\mathcal{B}^{(0),\mathrm{T}}_{5/2}-\alpha\mathcal{B}^{(1),\mathrm{T}}_{5/2})$
Hahn (2003)	$H^2(1+\alpha^2)$	$\frac{1}{8}\alpha\left[\mathcal{B}^{(1),\mathrm{H}}_{3/2}-3\alpha H^2(2+H^2)\mathcal{B}^{(0),\mathrm{H}}_{5/2}\right]$	$-\frac{1}{4}\alpha\left[\mathcal{B}^{(2),\mathrm{H}}_{3/2}-3\alpha H^2(2+H^2)\mathcal{B}^{(1),\mathrm{H}}_{5/2}\right]$	ϕ_{11}
Teyssandier & Ogilvie (2016)	$S^2\alpha$	$\frac{1}{8}\alpha\mathcal{B}^{(1),\mathrm{TO}}_{3/2}$	$-\frac{1}{4}\alpha\mathcal{B}^{(2),\mathrm{TO}}_{3/2}$	ϕ_{11}

Note. The ratio of semimajor axes between any two disc rings is given by $\alpha = a_</a_>$, where $a_> = \max(a_1, a_2)$ and $a_< = \min(a_1, a_2)$. The softened interactions under consideration are those of Tremaine (1998), Touma (2002), Hahn (2003) and Teyssandier & Ogilvie (2016) – see Section 4.2.1 for further details. For reference, the expressions of ϕ^{LL}_{ij} corresponding to the (unsoftened) Newtonian ring-ring interaction (i.e., classical Laplace–Lagrange formalism) are also shown in the top row. The Laplace coefficients which are softened by the introduction of a softening parameter $\epsilon^2(\alpha)$ are defined in Equation (4.7); see also Equations (4.15) – (4.13). Note that the expressions of ϕ_{ij} reported in Touma (2002) have been corrected in a subsequent paper of Touma & Sridhar (2012).

4.3 Comparison: Power-Law Discs

My goal is to examine the performance of different softening prescriptions outlined in Section 4.2.1 in comparison with the results obtained using the unsoftened Heppenheimer method (Section 4.2.2).

I start this exercise using a model of apse-aligned (i.e., $d\varpi_d/da = 0$), truncated power-law (hereafter PL) disc as a simple example. I characterise surface density and eccentricity of such a disc by

$$\Sigma_d(a) = \Sigma_0 \left(\frac{a_0}{a}\right)^p, \qquad e_d(a) = e_0 \left(\frac{a_0}{a}\right)^q \tag{4.16}$$

for $a_{\rm in} \leq a \leq a_{\rm out}$, where Σ_0 and e_0 are the pericentric surface density and eccentricity of the disc at some reference semimajor axis a_0.

Plugging this ansatz into Equations (4.4) – (4.6), the secular disturbing function R_d due to PL discs can be simplified to (Silsbee & Rafikov, 2015a)

$$R_d = K\Big[\psi_1 e_p^2 + \psi_2 e_p e_d(a_p)\cos(\varpi_p - \varpi_d)\Big], \tag{4.17}$$

where $K = \pi G \Sigma_0 a_0^p a_p^{1-p}$ and the dimensionless coefficients ψ_1 and ψ_2 are given by

$$\psi_1 = 2\int_{\alpha_1}^{1} \alpha^{1-p}\phi_{22}(\alpha)d\alpha + 2\int_{\alpha_2}^{1} \alpha^{p-2}\phi_{11}(\alpha)d\alpha, \tag{4.18}$$

$$\psi_2 = 2\int_{\alpha_1}^{1} \alpha^{1-p-q}\phi_{12}(\alpha)d\alpha + 2\int_{\alpha_2}^{1} \alpha^{p+q-2}\phi_{12}(\alpha)d\alpha, \tag{4.19}$$

with $\alpha_1 = a_{\rm in}/a_p$ and $\alpha_2 = a_p/a_{\rm out}$.

The coefficients ψ_1 and ψ_2 are functions of the power-law indices (p and q), any softening parameter involved (through ϕ_{ij}), as well as the test particle semimajor axis a_p (through $\alpha_{1,2}$). They are related to A_d and B_d via

$$\mathsf{A}_d(a_p) = \frac{2K}{n_p a_p^2}\psi_1, \qquad \mathsf{B}_d(a_p) = \frac{K}{n_p a_p^2} e_d(a_p)\psi_2. \tag{4.20}$$

As shown in Appendix D, for certain ranges of power-law indices p and q, both ψ_1 and ψ_2 converge to values depending only on p and q and any softening parameter used, provided that the test particle orbit is well separated from the disc boundaries (i.e., in the limit $\alpha_{1,2} \to 0$). For p and q in these ranges (determined in Appendix D

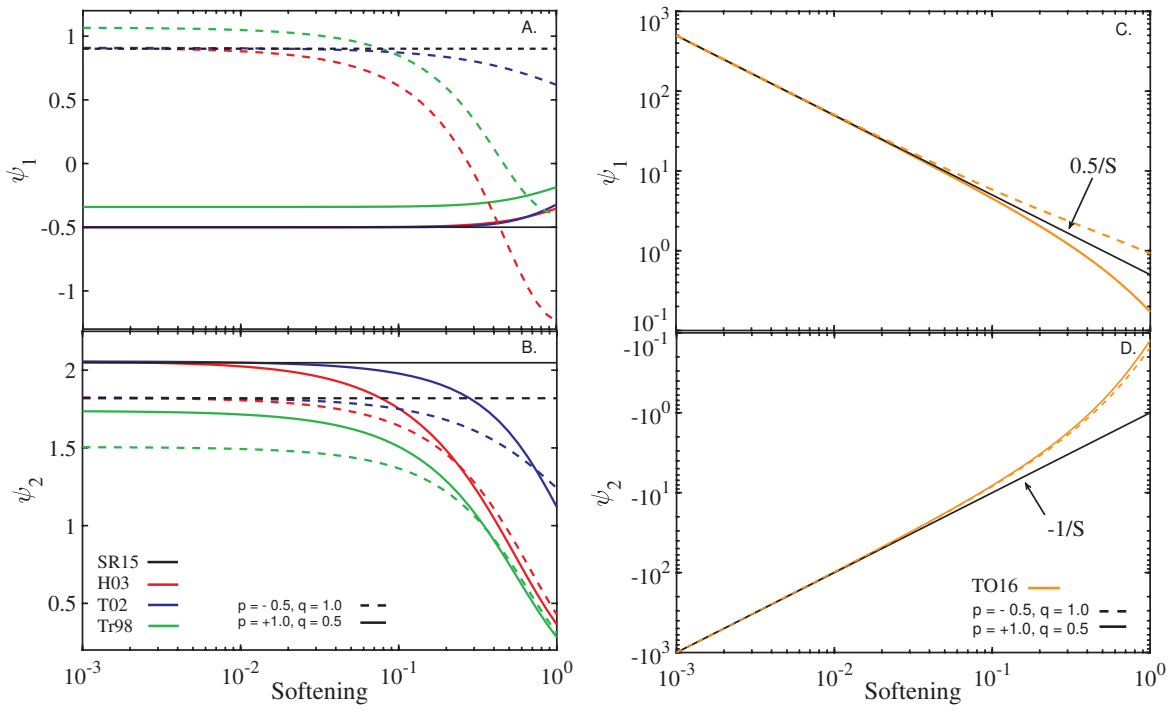

Figure 4.2. Behaviour of the axisymmetric (ψ_1, Equation (4.18), top panels) and non-axisymmetric (ψ_2, Equation (4.19), bottom panels) components of the softened gravitational potential due to an infinite power-law disc as a function of softening ς. The calculations assume two different disc structures specified by the values of p and q shown by different line types as explained in legend. For clarity, the results obtained by the softened formalisms of Tremaine (1998), Touma (2002) and Hahn (2003) are collated in the left-hand panels and those obtained by the softening method of Teyssandier & Ogilvie (2016) are shown in the right-hand panels. The left-hand panels also show the ψ_1 and ψ_2 obtained by SR15 not assuming any softening (black horizontal lines). See the text (Section 4.3.1) for details.

for each of the considered softened formalisms, similar to SR15), the coefficients ψ_1 and ψ_2 are determined by the *local* behaviour of $\Sigma_d(a)$ and $e_d(a)$ in the vicinity of test particle semimajor axis.

Given this, I first focus on infinitely extended ($\alpha_{1,2} \to 0$) PL discs with p and q within these ranges (I defer discussion of secular dynamics near the disc edges to Section 4.5). Then, ψ_1 and ψ_2 become independent of a_p (i.e., functions of p, q, and ς only), making them useful as simple metrics for judging the validity of different models of softening.

4.3.1 Behaviour with respect to Variation of Softening

Figure 4.2 illustrates the behaviour of ψ_1 and ψ_2 predicted by each of the softening formalisms described in Section 4.2.1 for an infinite PL disc, shown as a function of

the corresponding "softening"[6] ς for two different sets of p, q (indicated in panel B). For reference, black horizontal lines show the values of ψ_1 and ψ_2 expected from the calculations of SR15 using the unsoftened Heppenheimer approach[7].

The left-hand panels of Figure 4.2 illustrate the behaviour of the softening models of Tremaine (1998), Touma (2002), and Hahn (2003). They demonstrate that the latter two formalisms predict ψ_1 and ψ_2 in quantitative agreement with the unsoftened calculations of SR15: results of both Touma (2002) (blue) and Hahn (2003) (red) converge to the SR15 results as their corresponding softening ς approaches zero; both the amplitude and sign of ψ_1 and ψ_2 are reproduced. It is also evident that, depending on disc model, ψ_1 and ψ_2 converge to values given by SR15 at different values of softening. Nevertheless, I generally[8] find that $\varsigma \lesssim 10^{-3}$ guarantees the convergence of ψ_1 and ψ_2 to within few per cent of the correct values for all p and q as long as $a_{\rm in} \ll a_p \ll a_{\rm out}$ (see e.g. Figure 4.5).

The same panels also indicate that $\psi_1(\varsigma)$ and $\psi_2(\varsigma)$ predicted by the softened formalism of Tremaine (1998) (in green), while converging to finite values as $\varsigma = \beta_c \to 0$, do not reproduce the SR15 results exactly in this limit. Indeed, one can see that even for the smallest adopted value of $\beta_c = 10^{-3}$, the softening prescription of Tremaine (1998) yields ψ_1 and ψ_2 different by tens of per cent from SR15. It is easy to demonstrate that these quantitative differences do not vanish by further decreasing β_c. For instance, when $p = 1$, the coefficient ψ_1 can be evaluated analytically as

$$\psi_1^{\rm Tr98} = -\frac{1}{2\sqrt{\beta_c^2+1}} + \frac{\mathbf{E}\left(2/\sqrt{\beta_c^2+4}\right)}{\pi\sqrt{\beta_c^2+4}} = -\frac{1}{2} + \frac{1}{2\pi} + \mathcal{O}(\beta_c^2) \qquad (4.21)$$

in agreement with Panel A ($\mathbf{E}(k)$ is the complete elliptic integral of a second kind). At the same time, the unsoftened approach of SR15 predicts $\psi_1 = -1/2$ for $p = 1$ disc. Moreover, close inspection of Figures 4.2(A) and (B) shows that, in the limit of $\beta_c \to 0$, the ψ_1 and ψ_2 curves computed using softening model of Tremaine (1998) are offset vertically from the unsoftened calculations by $1/2\pi$ and $-1/\pi$, respectively, for any $(p,\ q)$ – see also Figure 4.5. I will analyse the reasons for this quantitative discrepancy in Section 4.6.1.

[6] The softening length b_c present in the formulation of Touma (2002) is scaled by the test particle semimajor axis a_p in all the figures where I present results for infinite PL discs. I do this to properly collate the results computed by different softening formalisms in one figure.

[7] Equations (A37) and (A38) in Silsbee & Rafikov (2015a) provide analytic expressions for ψ_1 and ψ_2, respectively, for infinite PL discs.

[8] For particles with orbits near sharp disc edges, I find that smaller values of ς is required to recover the expected dynamics, see Section 4.5.

Right-hand panels of Figure 4.2 show the behaviour of ψ_1 (Panel C) and ψ_2 (Panel D) as a function of "softening", $\varsigma = S$, resulting from the approach of Teyssandier & Ogilvie (2016). There are several features to note here. First, this model predicts $\psi_1 > 0$ for all values of softening S and disc models (i.e., p and q), implying *prograde* free precession. This is in contrast with the other softening prescriptions, as well as SR15, which correctly capture *retrograde* free precession for $p = 1$ and prograde for $p = -0.5$ (see Panel A). Similarly, ψ_2 is always negative, contrary to the expectations (see Panel B). Second, in the limit of $S \to 0$, both ψ_1 and ψ_2 attain values independent of the disc model, which is clearly inconsistent with the dependence on (p, q) seen in Figure 4.2(A) and (B). Thirdly, and most importantly, both ψ_1 and ψ_2 *diverge* as the softening $S \to 0$. Indeed, it suffices to employ the asymptotic expansion of the Laplace coefficients $\mathcal{B}_{3/2}^{(m),\mathrm{TO}}$ in the limit of $\alpha \to 1$ (Equation C.7) to demonstrate that both ψ_1 and ψ_2 (Equations 4.18 – 4.19) behave as

$$\psi_1^{\mathrm{TO16}} \approx \frac{1}{2S} + \mathcal{O}(S), \qquad \psi_2^{\mathrm{TO16}} \approx -\frac{1}{S} + \mathcal{O}(S) \tag{4.22}$$

as $S \to 0$ for all values of p and q. The behaviour shown in Figures 4.2(C) and (D) agrees with these asymptotic expressions.

4.3.2 Details of Convergence of Different Softening Prescriptions

Different softening prescriptions explored in this chapter are designed to modify the behaviour of the integrand in Equations (4.5) and (4.6) primarily in the vicinity of the test particle orbit, i.e., as $a \to a_p$ or $\alpha \to 1$. For this reason, it is interesting to look in more detail on how this modification actually allows each softening model to achieve (or not) the expected results. This exercise also illustrates the contribution of different parts of the disc to secular dynamics.

To this goal, I compute the values of ψ_1 and ψ_2 in an infinitely extended PL disc, like in Section 4.3.1, but now with a narrow clean gap (in semimajor axis) just around the test particle orbit, and explore the effect of varying the width of this gap (Ward, 1981). The inner and outer edges of the gap, in which $\Sigma_d(a)$ is set to zero, are at $a_{d,i} = (1-x)a_p \leq a_p$ and $a_{d,o} = (1-x)^{-1}a_p \geq a_p$, respectively, with a single parameter x controlling the gap width. As $x \to 0$, the width of the gap goes to zero. I compute secular coefficients in such a gapped disc denoted $\tilde{\psi}_1(x)$ and $\tilde{\psi}_2(x)$, by appropriately changing the upper integration limits in Equations (4.18) and (4.19),

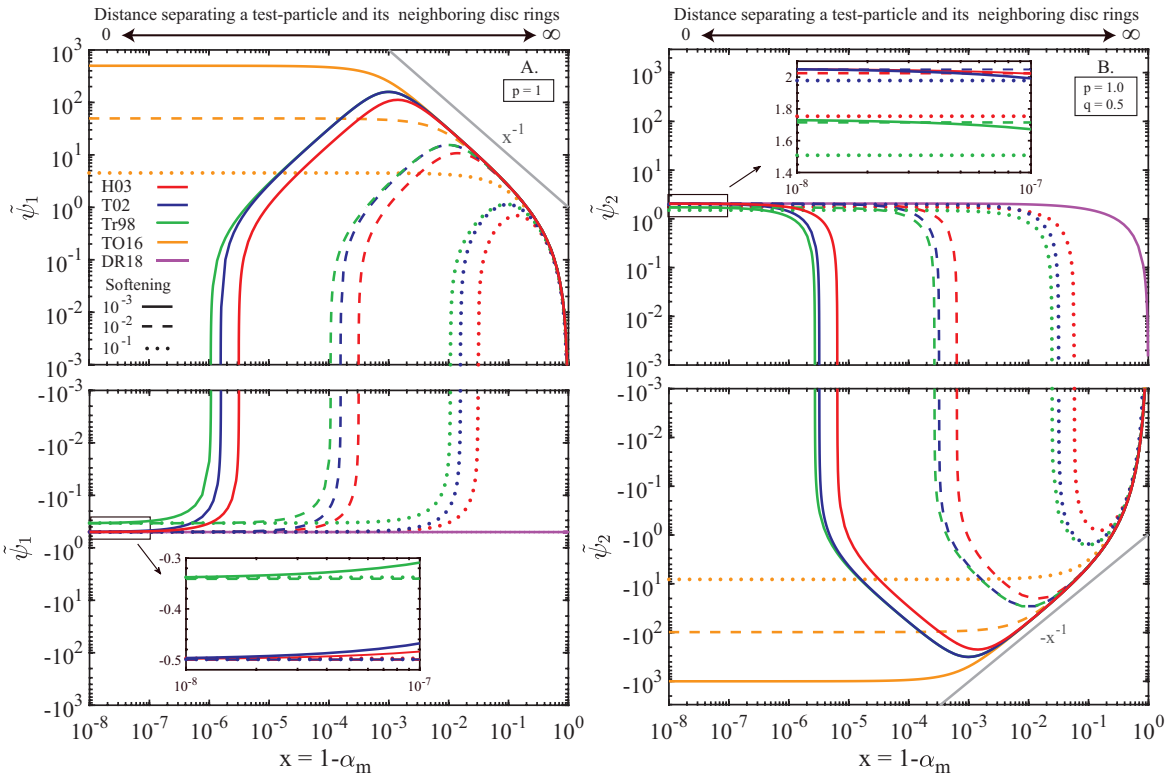

Figure 4.3. Behaviour of the cumulative pre-factors $\tilde{\psi}_1(x)$ (panel A) and $\tilde{\psi}_2(x)$ (panel B) of the disturbing function due to a power-law disc ($p = 1$, $q = 0.5$ and $a_{\rm in} \to 0$, $a_{\rm out} \to \infty$) with softened gravity, shown as a function of x — relative separation between a given test particle orbit and the nearest neighbouring disc rings. Formalisms of Hahn (2003), Touma (2002), Tremaine (1998), and Teyssandier & Ogilvie (2016) are shown by different colours as indicated in panel (A), for different values of softening (shown by different line types). The purple lines represent results obtained by the unsoftened expressions of Davydenkova & Rafikov (2018) (DR18) based on the Heppenheimer method (see Section 4.6.3). Insets illustrate the behaviour as $x \to 0$ for the three convergent softened formalisms — see the text (Section 4.3.2) for more details.

i.e., from 1 to $\alpha_m \equiv 1 - x$. This eliminates gravitational effect of the disc annuli with $a_{d,i}(x) < a < a_{d,o}(x)$.

In Figure 4.3, I display the behaviour of $\tilde{\psi}_1(x)$ (Panel A) and $\tilde{\psi}_2(x)$ (Panel B) as a function of $x = 1 - \sqrt{a_{d,i}/a_{d,o}}$ for various values of softening ς to highlight the effects of different softening prescriptions. The calculations assume a base PL disc model with $p = 1$ and $q = 0.5$ (recall that ψ_1 depends on p, while ψ_2 depends on $p + q$; Equations 4.18 and 4.19). There are several notable features in this figure.

First, when the gap is wider than the characteristic softening length ςa_p, i.e., $\varsigma \lesssim x \leq 1$, the amplitudes of both $\tilde{\psi}_1(x)$ and $\tilde{\psi}_2(x)$ increase from zero at $x = 1$ (infinitely wide gap) to their maximum values reached at $x \sim \varsigma$. In all cases ψ_1

is positive, meaning *prograde* precession of a test particle orbit in a wide gap, in agreement with the unsoftened results of Ward (1981) and Davydenkova & Rafikov (2018) — secular effect of a collection of distant disc 'wires' conforms to expectations of the classical Laplace–Lagrange theory (i.e., prograde precession).

In the range $\varsigma \lesssim x \ll 1$, I find that $\tilde{\psi}_1(x) \sim |\tilde{\psi}_2(x)| \sim x^{-1}$, irrespective of the softening model used; their maximum values are always $\sim \varsigma^{-1}$. This convergent behaviour is easy to understand since for $\varsigma \lesssim x$ the role of softening is negligible, $\mathcal{B}_s^{(m)}(\alpha, \varsigma) \approx b_s^{(m)}(\alpha)$, and all ϕ_{ij} effectively reduce to their classical counterparts $\phi_{ij}^{\rm LL}$ given by Equations (4.8) and (4.9), which can be easily verified using the expressions listed in Table 4.1. The scaling of $\tilde{\psi}_1(x)$ and $|\tilde{\psi}_2(x)|$ with x is simply a result of asymptotic behaviour of $b_{3/2}^{(m)}(\alpha) \to (1-\alpha)^{-2}$ as $\alpha \to 1$, upon radial integration in Equations (4.18) and (4.19).

Secondly, upon reaching their extrema at $x \sim \varsigma$, amplitudes of $\tilde{\psi}_1(x)$ and $\tilde{\psi}_2(x)$ computed using softening prescriptions of Tr98, T02, and H03 start decreasing as x decreases. In the range of semimajor axes corresponding to $x \lesssim \varsigma$, softening significantly modifies the behaviour of $\mathcal{B}_s^{(m)}(\alpha, \varsigma)$ away from the divergent behaviour of $b_s^{(m)}(\alpha)$. The modification is such that the softened interaction with the disc annuli $\lesssim \varsigma a_p$ away from the test particle orbit starts to dynamically *counteract* the contribution of the more distant annuli (with $x \approx 1$). As a result of this compensation, $\tilde{\psi}_1$ and $\tilde{\psi}_2$ cross zero and change sign at some $x = C\varsigma^2$, where $C \sim 1$ is a constant[9].

At the same time, $\tilde{\psi}_1^{\rm TO16}$ and $\tilde{\psi}_2^{\rm TO16}$ calculated according to Teyssandier & Ogilvie (2016) clearly show different behaviour. Instead of decreasing in amplitude as $x \lesssim \varsigma$, they remain essentially constant, having reached their saturated values $\sim \varsigma^{-1}$ at $x \sim \varsigma$. This explains the lack of convergence with S obvious in Figure 4.2C, D, since the values to which $|\tilde{\psi}_1^{\rm TO16}|$ and $|\tilde{\psi}_2^{\rm TO16}|$ converge keeps increasing as $\varsigma \to 0$. Moreover, both coefficients also never change sign, always predicting prograde precession ($\tilde{\psi}_1^{\rm TO16} > 0$). The origin of this difference with other smoothing prescriptions will be addressed in Section 4.6.2.

Upon further decrease of x below ς^2, both $\tilde{\psi}_1$ and $\tilde{\psi}_2$ computed using models of Tr98, T02 and H03 ultimately converge to their corresponding values obtained for a continuous disc (i.e, for $x = 0$, see Figure 4.2) independent of the assumed value of ς.

I note that the opposite contributions to e.g. ψ_1 produced by the distant ($x \gtrsim \varsigma$, positive) and nearby (i.e., with $x \lesssim \varsigma$, negative) disc annuli is not unique to softened

[9]For $p = 1$, $\tilde{\psi}_1$ becomes analytic for the softened formalisms of both H03 and Tr98 allowing me to quantify the value of C. Performing the integral over $d\alpha$ in Equation (4.18) - (4.19), I find $C_{Tr98} = (\pi - 1)/2$ and $C_{H03} = \pi$; in agreement with Figure 4.3. For other values of p and q, for which $\psi_1 < 0$ (c.f. Figure 4.5), I numerically find that C varies by at most a factor of ten.

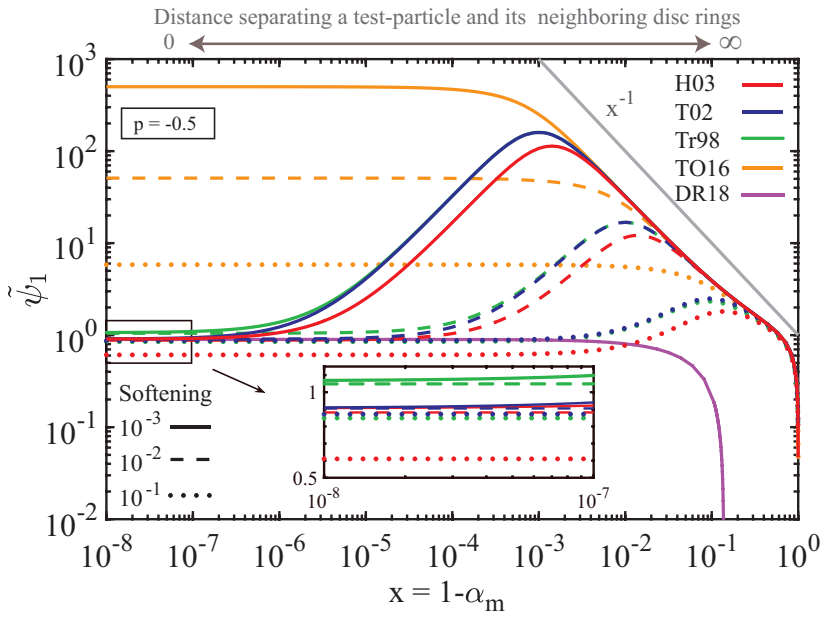

Figure 4.4. Same as Figure 4.3, but now for an axisymmetric power-law disc with $p = -0.5$. Note that for this disc model softened $\tilde{\psi}_1(x)$ does not cross zero and converges to a positive value as $x \to 0$, in agreement with the results in Figure 4.2A.

gravity. Indeed, both Ward (1981) and Davydenkova & Rafikov (2018), using the unsoftened Heppenheimer method, found that a particle orbit fully embedded in a $p = 1$ disc has negative precession rate, whereas a particle orbiting fully in the gap precesses in the positive sense (and at high rate if the gap is narrow). As the gap width is reduced, a smooth transition between the two regimes must occur as the test particle orbit starts crossing the gap edge (i.e., for $x \lesssim e_p$), with the disc annuli crossing the particle orbit giving rise to a negative contribution to $\tilde{\psi}_1$. Eventually, the shrinking of the gap brings $\tilde{\psi}_1$ to a finite negative value (for $p = 1$ disc) as $x \to 0$. This sequence is very similar to the behaviour I find with softened gravity for $x \lesssim \varsigma$.

In Figure 4.4, I show calculations for $\tilde{\psi}_1(x)$ similar to those in Figure 4.3(A) but for a different disc model — axisymmetric PL disc with $p = -0.5$. In this case unsoftened calculations (e.g., SR15) predict that disc gravity should drive *prograde* precession of a test particle in a smooth disc. One can clearly see that many of the features present in Figure 4.3 are reproduced for this model as well: discrepancy between the TO16 model and others, $\tilde{\psi}_1(x) \sim x^{-1}$ scaling for $\varsigma \lesssim x \ll 1$, decay of $\tilde{\psi}_1(x)$ for $\varsigma^2 \lesssim x \lesssim \varsigma$, and ultimate convergence to ψ_1 in a disc with no gap. The only obvious difference is the fact that $\tilde{\psi}_1$ does not cross zero[10] for this disc model with $p = -0.5$.

[10]This is the case for all power-law disc models with $p < 0$ or $p > 3$ for which the expected free precession rate is positive, see Figure 4.5.

To summarise, Figures 4.3 and 4.4 indicate that secular dynamics in softened power-law discs is dictated by the delicate balance of the opposing contributions due to nearby (i.e., with $x \lesssim \varsigma$) and distant disc annuli (i.e., with $x \gtrsim \varsigma$), in qualitative agreement with the *unsoftened* results of Ward (1981). These figures also demonstrate that the softening prescription of TO16 yields inaccurate results due to its inability to capture the dynamical effects of disc annuli adjacent to the test particle orbit (those with $x \lesssim \varsigma$), see Section 4.6.2. I will discuss additional implications of these calculations in Section 4.6.3.

4.3.3 Variation of Disc Model: p and q

I now examine the dependence of ψ_1 and ψ_2 on the specifics of the disc model reflected in power-law indices p and q. Figures 4.5(A) and (B) illustrate the results based on different softening prescriptions[11] assuming a softening value of $\varsigma = 10^{-3}$ (for which Figure 4.2(A) and (B) suggests good convergence of ψ_1 and ψ_2). For reference, black open circles show the expected behaviour of ψ_1 and ψ_2 computed by Silsbee & Rafikov (2015a) using the unsoftened Heppenheimer approach.

It is clear that the softened formalisms of both Touma (2002) and Hahn (2003) perfectly reproduce the expected behaviour of the pre-factors ψ_1 and ψ_2 as a function of p and q (i.e., for various PL disc models). On the other hand, the prescription of Tremaine (1998) predicts a behaviour of ψ_1 and ψ_2 only in qualitative agreement with the expected results: the computed values of secular coefficients deviate by tens of per cent from that of SR15. For all values of p and q, the formalism of Tremaine (1998) yields an additional positive contribution to ψ_1 equal to $1/2\pi$ and a negative contribution to ψ_2 equal to $-1/\pi$ (these offsets are highlighted in Figures 4.5(A) and (B) by scale bars). Although these differences are not very significant, they lead to (1) predicting a wrong sign for the test particle free-precession rate for $p \approx 0$ or $p \approx 3$ (for which SR15 yields $\psi_1 \approx 0$), and (2) a mismatch of tens of per cent between the disc-driven forced eccentricity oscillations, $e_p^m/e_d(a) = |\psi_2/\psi_1|$, and the expectations based on SR15. The latter point is illustrated in Figure 4.5(C).

4.4 Comparison: Non-Power-Law Discs

I now turn my attention to the performance of the different softening prescriptions for more general discs. Namely, I focus on two apse-aligned, non-PL disc models

[11] I do not present results obtained by the method of Teyssandier & Ogilvie (2016).

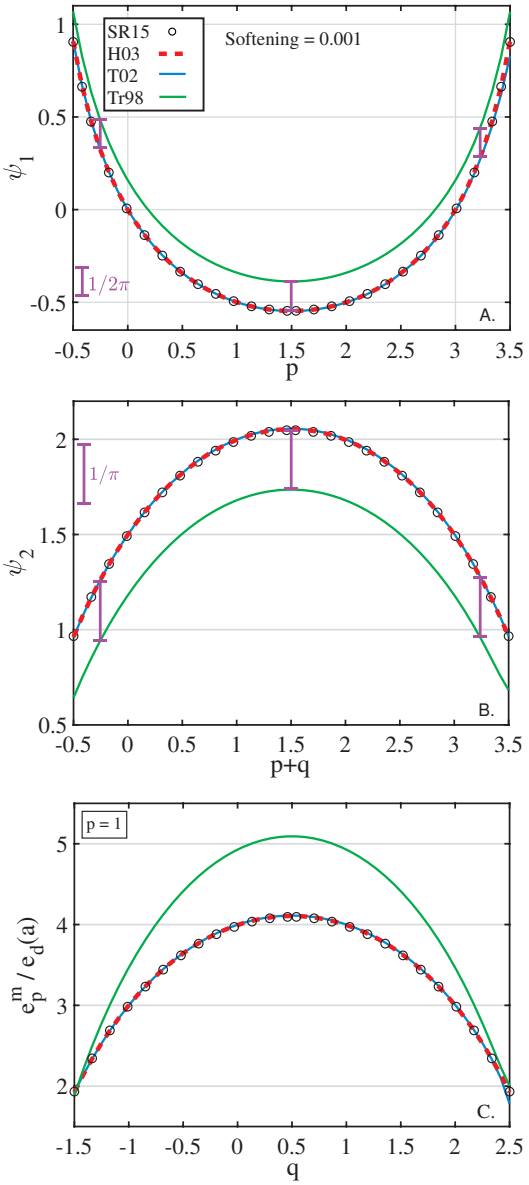

Figure 4.5. Dependence of the coefficients ψ_1 (panel A) and ψ_2 (panel B) on the power-law disc model represented by the indices p and $p+q$, respectively. Panel C shows the amplitude e_p^m of eccentricity oscillations (normalised by disc eccentricity e_d) induced by disc gravity. Results for softened formalisms of Hahn (2003) (in red), Touma (2002) (in blue) and Tremaine (1998) (in green) are computed using softening $\varsigma = 10^{-3}$. Calculations assume infinitely extended disc (i.e., no edge effects). For reference, open black circles show the profiles of ψ_1, ψ_2 and e_p^m as computed by SR15: curves for Hahn (2003) and Touma (2002) fall on top of them, while those for Tremaine (1998) show constant offset in terms of both ψ_1 and ψ_2 (illustrated by scale bars in panels A and B) resulting in deviation between e_p^m curves (panel C).

previously studied by Davydenkova & Rafikov (2018) (DR18) based on the unsoftened Heppenheimer method. The dynamics in such non-PL discs, according to DR18, differs from the PL discs in a very important way: the free-precession of test-particles can naturally change from retrograde to prograde (and vice versa) within such discs. Furthermore, an important feature of the models considered below is that Σ_d smoothly goes to zero at finite radii in a manner that does not give rise to the edge effects; see DR18 and Section 4.5.

4.4.1 Quartic Disc Model

I start by looking at the secular dynamics in the potential of a *Quartic* disc characterised by the surface density

$$\Sigma_d(a) = \tilde{\Sigma}_0 \frac{(a_{\rm out}-a)^2(a_{\rm in}-a)^2}{(a_{\rm out}-a_{\rm in})^4}, \tag{4.23}$$

and linear eccentricity profile in the form

$$e_d(a) = \tilde{e}_0\left(1+\frac{a_{\rm out}-a}{a_{\rm out}-a_{\rm in}}\right) \tag{4.24}$$

for $a_{\rm in} \le a \le a_{\rm out}$ (with $a_{\rm in} = 0.1$ AU, $a_{\rm out} = 5$ AU), where $\tilde{\Sigma}_0 = 1153$ g cm^{-2} and $\tilde{e}_0 = 0.01$ are normalization constants (one of the models in DR18).

Figure 4.6 summarises the salient features of secular dynamics in the potential of such a disc adopting a softening value of $\varsigma = 10^{-3}$. It shows the excellent agreement between the radial profiles of A_d, B_d and e_p^m computed using the unsoftened calculations of Davydenkova & Rafikov (2018) and those computed using softening prescriptions of Touma (2002) and Hahn (2003). Similar to the case of PL discs, I find that the softening prescription of Tremaine (1998) yields results that agree *qualitatively* with the expected results but differ quantitatively. Deviations of A_d and B_d computed using this model from Davydenkova & Rafikov (2018), in particular, modify the locations at which A_d and B_d become zero. This explains the slight shift in the semimajor axes at which $e_p^m = 2B_d/A_d$ goes through zero or diverges; see Figure 4.6.

The difference between the Tremaine (1998) and Touma (2002) calculations illustrated here could be relevant for understanding the quantitative differences between the studies of Tremaine (2001) and Gulati et al. (2012) who analysed the slow ($m = 1$) modes supported by softened Kuzmin discs with softening prescriptions $b \propto r$ and $b = $ const respectively.

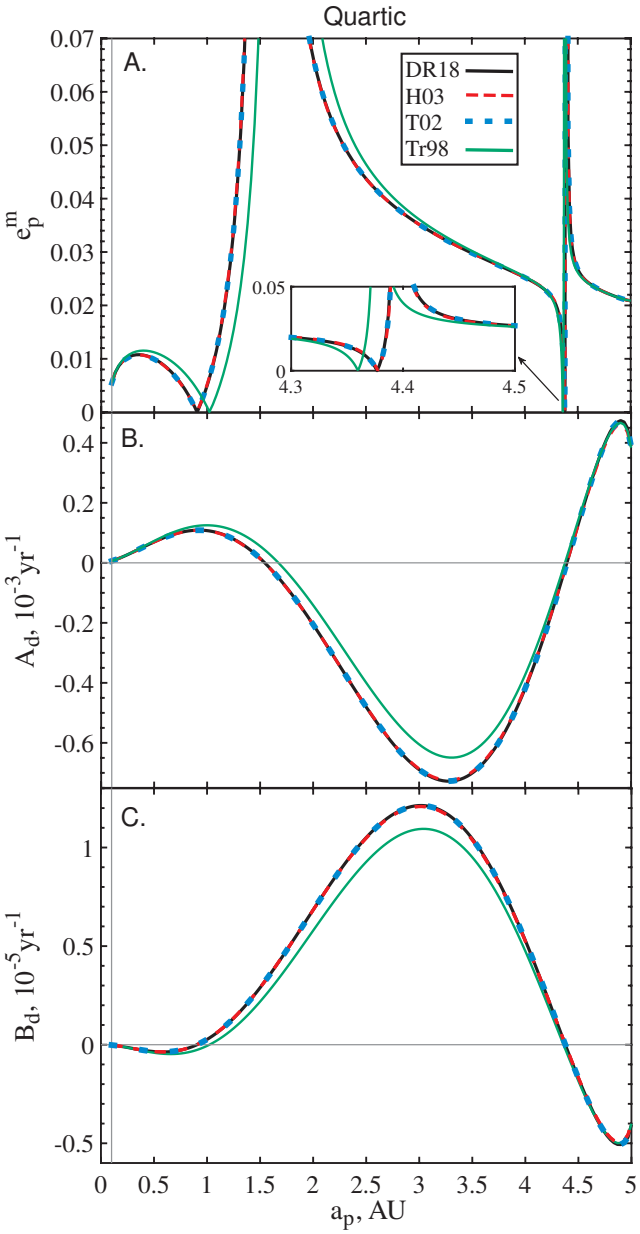

Figure 4.6. Performance of different softening formalisms (different colours) with softening parameter $\varsigma = 10^{-3}$ in the potential of a Quartic disc, see Equation (4.23), with the eccentricity profile (4.24). The disc extends from $a_{\rm in} = 0.1$ AU to $a_{\rm out} = 5$ AU. Shown as a function of semimajor axis a_p are the profiles of (A) the amplitude e_p^m of the disc-induced eccentricity oscillations, (B) the rate of disc-driven free precession A_d, and (C) the coefficient B_d appearing in the non-axisymmetric part of the disturbing function (4.4). The black lines represent the expected unsoftened results as computed by Davydenkova & Rafikov (2018). Curves for Hahn (2003) and Touma (2002) fall on top of the unsoftened results, while the softening method of Tremaine (1998) shows only qualitative agreement.

4.4.2 Gaussian Rings

Next I investigate secular dynamics in the potential of another disc model from DR18 — a Gaussian ring with the surface density profile

$$\Sigma_d(a) = \tilde{\Sigma}_0 \exp\left\{\frac{4 - [(a/a_c) + (a_c/a)]^2}{w_c}\right\} \tag{4.25}$$

centered around $a_c = 1.5$ AU with width $w_c = 0.18$ and surface density $\tilde{\Sigma}_0 = 100$ g cm^{-2} at a_c. The eccentricity profile is still given by Equation (4.24).

In Figure 4.7, I plot the behaviour of the corresponding A_d, B_d and e_p^m for the three (convergent) softened formalisms with $\varsigma = 10^{-3}$, together with those of unsoftened Heppenheimer method (DR18, in black). Once again, the results obtained using the formalisms of Touma (2002) and Hahn (2003) fall on top of the expectations. However, for this disc model the formalism of Tremaine (1998) reproduces the unsoftened calculations of Davydenkova & Rafikov (2018) quite well: the relative deviations are always less than 10 per cent. This improvement will be discussed further in Section 4.6.1.

4.5 Effects of Proximity to the Disc Edge

The disc models explored thus far were either infinitely extended (Section 4.3) or had surface density smoothly petering out to zero at finite radii (Section 4.4). This allowed me to not worry about the effects of sharp disc edges — discontinuous drops of the surface density — on secular dynamics, which are known to be important (Silsbee & Rafikov, 2015a; Davydenkova & Rafikov, 2018).

I now relax this assumption and examine the performance of different softening models in the vicinity of a sharp edge of the disc, where surface density drops discontinuously from a finite value to zero at a finite semimajor axis $a = a_{\rm edge}$. To that effect, I analyse the behaviour of secular coefficient A_d computed using the formalism of Hahn (2003) for different values of softening (results for B_d are very similar) near the disc edge[12]. Figure 4.8 shows the run of A_d near the inner edge $a_{\rm in}$ of the disc for test particles both inside ($a_p < a_{\rm in}$) and within ($a_{\rm in} < a_p < a_{\rm out}$) the disc as predicted by the formalism of Hahn (2003). The calculation assumes circular PL disc with $p = 1$

[12]I verified that softening prescriptions of Touma (2002) and Tremaine (1998) give very similar results in the limit $\varsigma \to 0$.

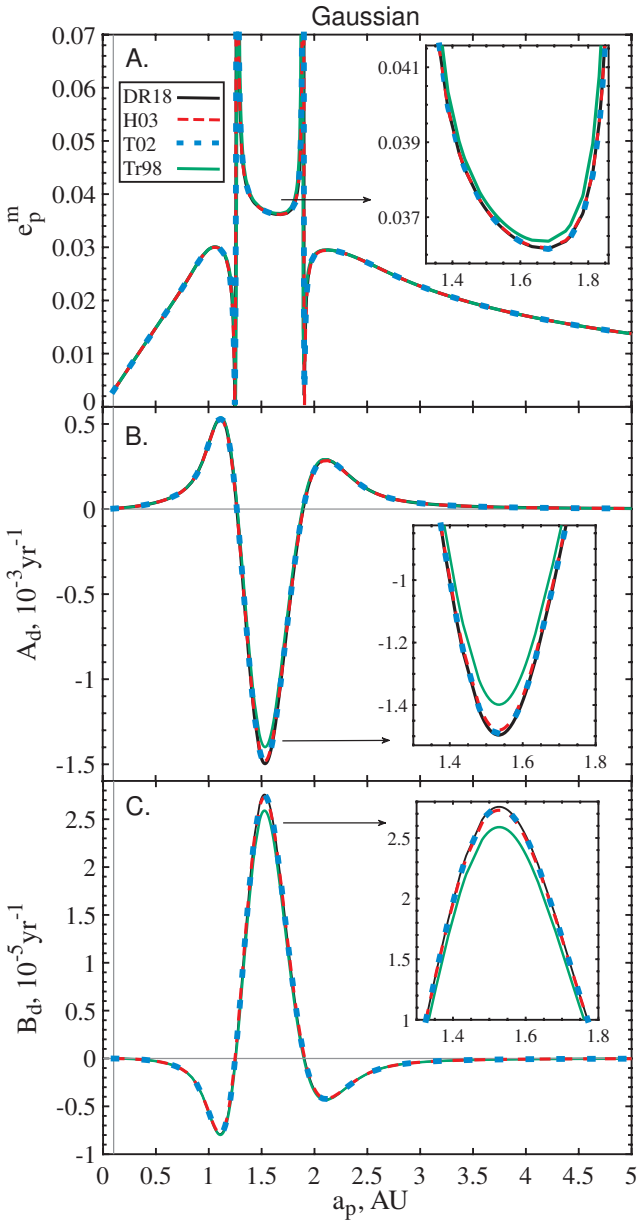

Figure 4.7. Same as Figure 4.6, but now for a Gaussian disc with $\Sigma_d(a)$ and $e_d(a)$ given by Equations (4.25) and (4.24) respectively. Note that for this disc model the formalism of Tremaine (1998) (green) shows quite good agreement with the unsoftened results, even at the quantitative level. See the text (Section 4.4.2) for details.

and $\Sigma_0 = 100$ g cm^{-2} extending between $a_{\rm in} = 1$ AU to $a_{\rm out} = 10$ AU, where I have set $a_0 = a_{\rm out}$ (Equation 4.16).

The unsoftened calculations based on Heppenheimer (1980) invariably predict that the free eccentricity precession rate A_d, as well as B_d, should diverge as the sharp edge of the disc is approached (e.g. SR15, DR18). Tremaine (2001) also found precession

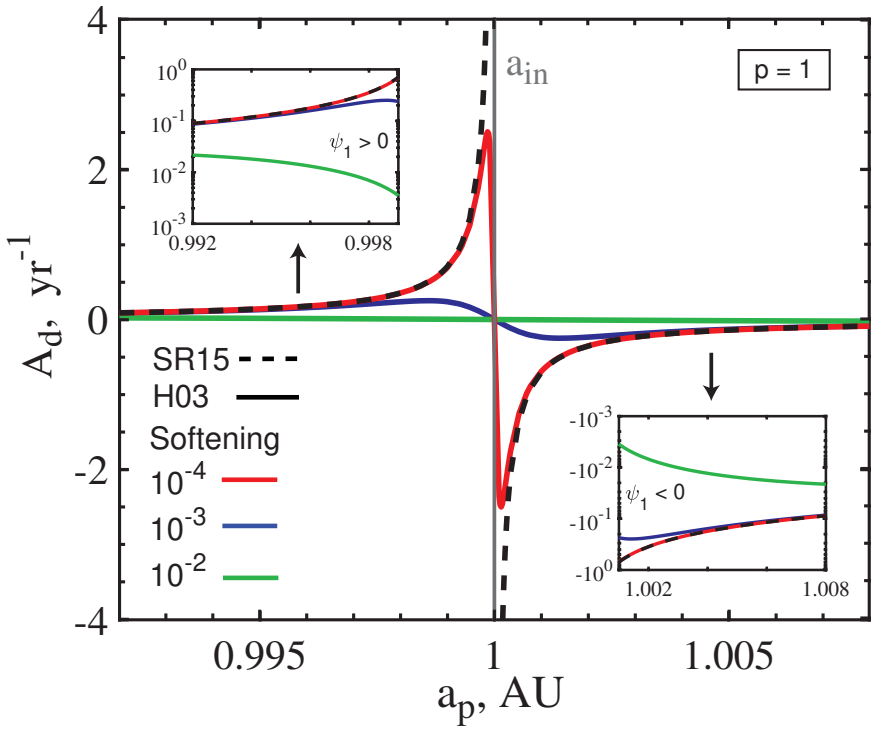

Figure 4.8. The behaviour of the free precession rate A_d near the inner edge $a_{\rm in} = 1$ AU of a circular power-law disc with surface density $\Sigma_d(a) = 100$ g cm^{-2} (10 AU/a) (Equation 4.16). One can see that the expected divergent behaviour of A_d near the disc edge is reproduced by the softening prescription of Hahn (2003) in the limit $\varsigma \to 0$. However, very near the sharp edge of the disc ς has to be very small for quantitative accuracy to be attained. Similar results can be obtained by the softened formalisms of both Touma (2002) and Tremaine (1998).

rate to diverge near the edge of a Jacobs–Sellwood ring (Jacobs & Sellwood, 2001). This is indeed the case as shown by the dashed curve computed using SR15.

The softened calculation using Hahn (2003) does largely reproduce this behaviour. However, I find that very close to the ring edge (at $|a - a_{\rm in}|/a_{\rm in} \sim 10^{-3}$) the agreement is achieved only for $\varsigma \leq 10^{-4}$, which is considerably smaller than the values ($\varsigma \sim 10^{-2}$) required to reproduce the dynamics of particles far from the disc edges, $a_{\rm in} \ll a_p \ll a_{\rm out}$, see Figure 4.2. For instance, the softened calculation with $\varsigma = 10^{-2}$ predicts A_d different from the SR15 results near the disc edge by more than an order of magnitude. Thus, accurately capturing secular dynamics near the sharp edges of discs/rings requires using very small values of softening[13]. This finding could be problematic, for instance,

[13]On the other hand, this condition is relaxed when the edge is not exactly sharp but rather has a finite width Δr over which the disc surface density smoothly peters out to zero; in this case ς only needs to be $\lesssim \Delta r/r$.

for numerical modelling of planetary rings, often found to have very sharp edges (Graps et al., 1995; Tiscareno, 2013).

Note that in Figure 4.8 softened A_d passes through zero exactly at $a_{\rm in}$, showing two sharp peaks of opposite signs just around this radius. Similar behaviour was found by Davydenkova & Rafikov (2018) for zero-thickness discs with Σ_d dropping sharply but continuously near the edge, demonstrating that variation of the sharpness of the edge is akin to softening gravity. In the case of truly zero-thickness disc and no softening (e.g., SR15) the segment of A_d curve connecting the two peaks turns into a vertical line at $a_{\rm in}$.

Similar divergent behaviour of A_d (and B_d) arises also at the outer edge of the disc considered in Figure 4.8 and, in general, at any radius within a disc where $\Sigma_d(a)$ exhibits a discontinuity.

Finally, I note that the dynamics of particles orbiting outside the disc (where $\Sigma_d(a) = 0$) is successfully reproduced by the classical Laplace-Lagrange theory without adopting any softening prescription (e.g., see Petrovich et al., 2019). Indeed, outside the radial extent of the disc semimajor axis overlap (i.e., $a_p = a$) is naturally excluded thus avoiding the classical singularity. Outside the disc the unsoftened calculations based on the Heppenheimer method (e.g., SR15 and DR18) reduce to the classical Laplace–Lagrange theory exactly.

4.6 Discussion

Results of previous sections reveal a diversity of outcomes when different softening models are applied. Two models – those of Touma (2002) and Hahn (2003) – successfully reproduce the unsoftened calculations based on the Heppenheimer method in the limit of zero softening. In the same limit, the formalism of Tremaine (1998) yields convergent results which are, however, different from the unsoftened calculations – typically by tens of per cent. Finally, the softening method of Teyssandier & Ogilvie (2016) does not lead to convergent results in the limit of vanishing softening parameter. Interestingly, the two successful models – i.e., those of Touma (2002) and Hahn (2003) – have been derived using rather different underlying assumptions (see Sections 4.2.1.2 and 4.2.1.3), producing different mathematical expressions for ϕ_{ij} (see Table 4.1), and yet their results are consistent with the unsoftened calculations in the limit of $\varsigma \to 0$.

To understand this variation of outcomes, I developed a general framework for computing the secular coefficients ϕ_{ij} (thus fully determining the softened secular model via Equations (4.4)–(4.6)) given an arbitrary softened two-point interaction potential

in the form given by Equation (4.10). This procedure involves orbit-averaging the softened potential along the particle trajectories; its details are presented in Appendix A. There is also an alternative approach, sketched in Appendix A.4, which assumes the disc to be a continuous entity from the start. Both of them arrive at the same expressions for R_d.

Using these results, I show in Appendix B that the expressions for ϕ_{ij} found by Touma (2002) and Hahn (2003) can be recovered exactly using this general framework if I set $\mathcal{F}(r_1, r_2) = b_c^2$ and $\mathcal{F}(r_1, r_2) = H^2(r_1^2 + r_2^2)$, respectively, in the expression given by Equation (4.10) for the two-point potential. This approach also allows one to address some of the questions raised above, which I do in Sections 4.6.1 and 4.6.2 below.

4.6.1 On the Softening Prescription of Tremaine (1998)

Results of Sections 4.3 and 4.4 indicate that the softening prescription of Tremaine (1998) – unlike that of Touma (2002) and Hahn (2003) – leads to quantitative differences when compared to the unsoftened calculations. I now demonstrate where these differences come from.

The form of the softened Laplace coefficient $\mathcal{B}_s^{(m),\mathrm{Tr}}$ defined by Equation (4.12) suggests an interaction potential (Equation 4.10) with $\mathcal{F}(r_1, r_2) = \beta_c^2 \max(r_1^2, r_2^2)$ for the softening model of Tremaine (1998). In Appendix B, I show that propagating this form of $\mathcal{F}(r_1, r_2)$ through my general framework results in the following expressions for the coefficients ϕ_{ij}:

$$\phi_{11} = \phi_{22} = \frac{\alpha}{8}\left[\mathcal{B}_{3/2}^{(1),\mathrm{Tr}} - 3\alpha\beta_c^2\mathcal{B}_{5/2}^{(0),\mathrm{Tr}} - \delta(\alpha-1)\beta_c^2\mathcal{B}_{3/2}^{(0),\mathrm{Tr}}\right], \tag{4.26}$$

$$\phi_{12} = -\frac{\alpha}{4}\left[\mathcal{B}_{3/2}^{(2),\mathrm{Tr}} - 3\alpha\beta_c^2\mathcal{B}_{5/2}^{(1),\mathrm{Tr}} - \delta(\alpha-1)\beta_c^2\mathcal{B}_{3/2}^{(1),\mathrm{Tr}}\right]. \tag{4.27}$$

These expressions are different from the entries in Table 4.1 for Tremaine (1998) in a single but very important way – the presence of terms involving Dirac delta-function. Such terms arise because the form of $\mathcal{F}(r_1, r_2)$ adopted in Tremaine (1998) is not sufficiently smooth – its first derivative is discontinuous at $r_1 = r_2$, while the calculation of ϕ_{ij} involves second-order derivatives of $\mathcal{F}$; see Equations (A.25)-(A.27), as well as Equation (A.28). Such singular terms do not arise in other types of softening prescriptions examined in this chapter since they all use infinitely differentiable versions of $\mathcal{F}(r_1, r_2)$. Thus, these terms should not be interpreted as representing some kind

of 'self-interaction' within the disc: they merely reflect the mathematical smoothness properties of $\mathcal{F}$ used in Tremaine (1998).

Presence of these terms in Equations (4.26) and (4.27) introduces corrections to coefficients A_d and B_d (Equations 4.5 and 4.6) in apse-aligned discs in the form:

$$\delta A_d(a_p) = -\frac{\pi G}{2n_p a_p}\beta_c^2 \Sigma_d(a_p)\mathcal{B}_{3/2}^{(0),\mathrm{Tr}}\Big|_{\alpha=1}, \tag{4.28}$$

$$\delta B_d(a_p) = +\frac{\pi G}{2n_p a_p}\beta_c^2 \Sigma_d(a_p) e_d(a_p)\mathcal{B}_{3/2}^{(1),\mathrm{Tr}}\Big|_{\alpha=1}. \tag{4.29}$$

Accounting for these corrections, I confirmed that the correct (unsoftened) behaviour of the coefficients of R_d can be reproduced for the non-PL discs – Quartic and Gaussian models, see Section 4.4. Note that $\delta A_d(a_p)$ and $\delta B_d(a_p)$ are proportional to the local disc surface density $\Sigma_d(a_p)$ and $\mathcal{B}_{3/2}^{(m),\mathrm{Tr}}(\alpha = 1) \sim \beta_c^{-2}$, see Equation (C.7). This likely explains the improved agreement between the calculations of Tremaine (1998) and Davydenkova & Rafikov (2018) for Gaussian rings (see Figure 4.7), which feature mass concentration in a narrow range of radii (in contrast to the Quartic model, see Figure 4.6).

For PL discs the terms proportional to Dirac delta-function in Equations (4.26) and (4.27) give rise to the following modifications of the coefficients ψ_1 and ψ_2 defined by Equations (4.18) and (4.19):

$$\delta\psi_1 = -\frac{1}{4}\beta_c^2 \mathcal{B}_{3/2}^{(0),\mathrm{Tr}}\Big|_{\alpha=1} = -\frac{1}{2\pi} + O(\beta_c^2), \tag{4.30}$$

$$\delta\psi_2 = \frac{1}{2}\beta_c^2 \mathcal{B}_{3/2}^{(1),\mathrm{Tr}}\Big|_{\alpha=1} = \frac{1}{\pi} + O(\beta_c^2), \tag{4.31}$$

see Equation (4.20). These corrections exactly match the offsets seen in Figure 4.5 between the calculations of Tremaine (1998) and the unsoftened calculations, thus explaining the origin of these uniform shifts. I also confirm this explanation in Figure 4.9, where I show the convergence of the modified Tremaine (1998) coefficients to the correct unsoftened values as softening is varied for two values of p and q.

To summarise, Equations (4.26) and (4.27) should replace the expressions given by Equation (26) of Tremaine (1998) *in applications to continuous discs*. However, when considering the interaction of two individual annuli with different semimajor axes (like in the classical Laplace–Lagrange theory), one has $\alpha \neq 1$ and terms in Equations (4.26) and (4.27) containing Dirac delta-function naturally vanish, reducing ψ_1 and ψ_2 back to the expressions quoted in Tremaine (1998).

4.6.2 On the Softening Prescription of Teyssandier & Ogilvie (2016)

I now turn my attention to the model of Teyssandier & Ogilvie (2016) trying to understand its distinct (divergent) behaviour. From the expression for $\mathcal{B}_s^{(m),\mathrm{TO}}$ in Equation (4.15), one infers that this model features softening parameter in the form of $\epsilon^2(\alpha) = S^2\alpha$. To soften secular interaction, Teyssandier & Ogilvie (2016) directly substituted $b_{3/2}^{(m)}$ in the classical expressions (4.8) and (4.9) for ϕ_{ij}^{LL} with $\mathcal{B}_{3/2}^{(m),\mathrm{TO}}$, see Section 4.2.1.4. This simple swap of Laplace coefficients has not been justified rigorously.

On the other hand, in Appendix B I show that softening parameter in the form $\epsilon^2(\alpha) = \varsigma^2\alpha$ corresponds to softening function $\mathcal{F}(r_1, r_2) = \varsigma^2 r_1 r_2$ in the two-point potential (Equation 4.10), see Equation (A.21). Propagating such a form of $\mathcal{F}(r_1, r_2)$ through the general framework in Appendix A, I find the following expressions for the coefficients ϕ_{ij} with $\varsigma = S$ (Appendix B):

$$\phi_{11} = \phi_{22} = \frac{\alpha}{8}\left[\mathcal{B}_{3/2}^{(1),\mathrm{TO}} + \frac{1}{2}S^2\mathcal{B}_{3/2}^{(0),\mathrm{TO}} - \frac{3}{4}S^2(2 + 2\alpha^2 + S^2\alpha)\mathcal{B}_{5/2}^{(0),\mathrm{TO}}\right], \tag{4.32}$$

$$\phi_{12} = -\frac{\alpha}{4}\left[\mathcal{B}_{3/2}^{(2),\mathrm{TO}} + \frac{1}{2}S^2\mathcal{B}_{3/2}^{(1),\mathrm{TO}} - \frac{3}{4}S^2(2 + 2\alpha^2 + S^2\alpha)\mathcal{B}_{5/2}^{(1),\mathrm{TO}}\right]. \tag{4.33}$$

Approach of Teyssandier & Ogilvie (2016) accounts for only the first terms in Equations (4.32) and (4.33), with coefficients which are $O(S^0)$, see Table 4.1. However, as I show below, the correct behaviour of ϕ_{ij} as $S \to 0$ is guaranteed only when *all* the terms present in the above expressions are taken into account.

To demonstrate this, in Figure 4.9 I repeat the same convergence study as in Section 4.3.1 but with the modified ϕ_{ij} given by Equations (4.32) and (4.33). One can see that the correct implementation of the softening $\epsilon^2(\alpha) = S^2\alpha$ proposed by Teyssandier & Ogilvie (2016) leads to the recovery of the expected test particle dynamics in infinite PL discs; this is very different from the divergent behaviour obvious in Figures 4.2(C) and (D). Similar to Hahn (2003) and Touma (2002), both ψ_1 and ψ_2 smoothly converge to their expected unsoftened values in the limit of $S \to 0$ for various PL disc models (i.e., p and q). Further tests using other disc models, looking at the edge effects, etc. reinforce this conclusion.

This discussion strongly suggests that for any adopted form of softening, the expansion of the secular disturbing function must be performed following a certain rigorous procedure[14] as done, for instance, in Appendix A. In other words, a direct

[14] An analogous method is to modify the *literal expansion* of the disturbing function (see Chapter 6 in Murray & Dermott, 1999) to account for softened interactions (e.g., Tremaine, 1998; Hahn, 2003;

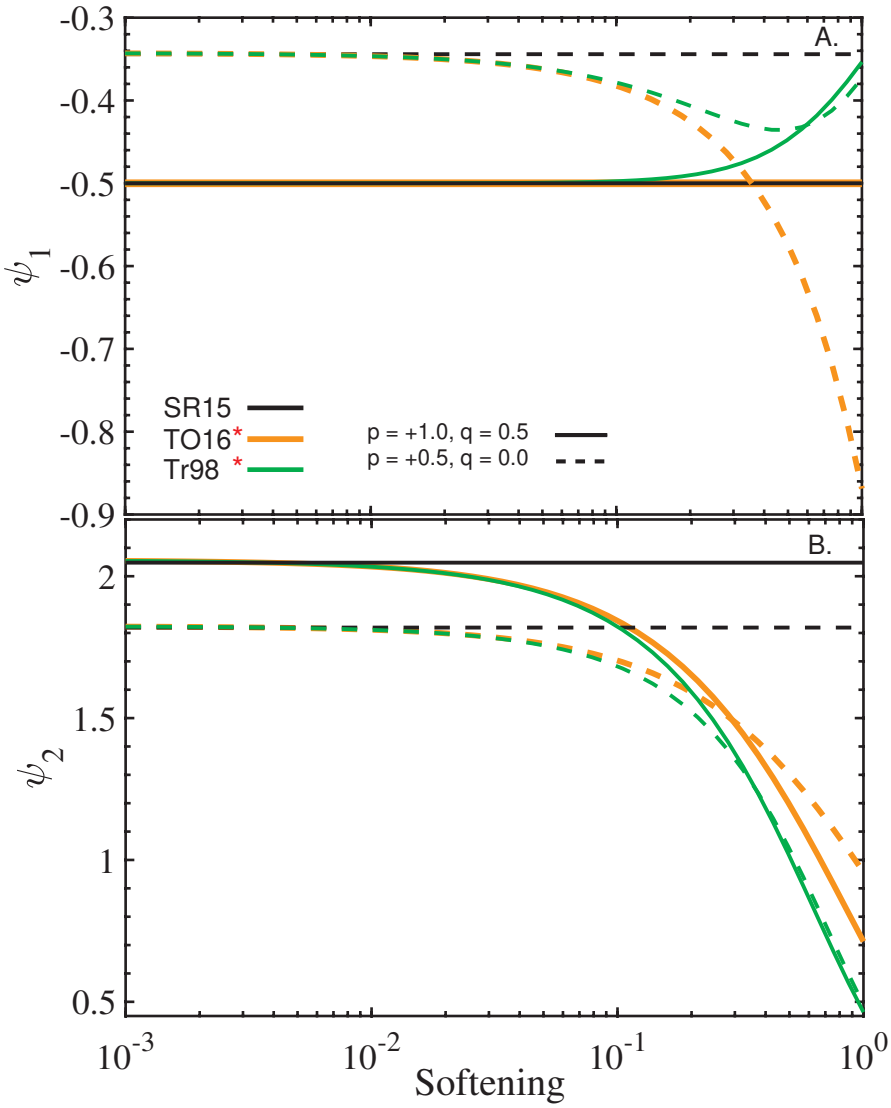

Figure 4.9. Similar to Figure 4.2, but now using the expressions for ϕ_{ij} given by Equations (4.26)–(4.27) and Equations (4.32)–(4.33) obtained by propagating $\mathcal{F}(r_1, r_2) = \varsigma^2 \max(r_1^2, r_2^2)$ of Tremaine (1998) and $\mathcal{F}(r_1, r_2) = \varsigma^2 r_1 r_2$ of Teyssandier & Ogilvie (2016), respectively, through the general framework outlined in Appendix A. Shown as a function of softening ς are ψ_1 (panel A) and ψ_2 (panel B) for two PL disc models specified by p and q indicated in panel A. Black lines represent the expectations based on Silsbee & Rafikov (2015a), to which the new expressions for ψ_1 and ψ_2 successfully converge as $\varsigma \to 0$.

replacement of the classical Laplace coefficients $b_{3/2}^{(m)}$ in Equation (4.1) with their

Lee et al., 2019). This could be done by replacing $b_{1/2}^{(m)}$ with $\mathcal{B}_{1/2}^{(m)}$ in Equation (7.1) of Murray & Dermott (1999) before applying the derivatives with respect to α. I note that this procedure could apply for all $\mathcal{F}(r_1, r_2)$ with continuous first derivatives satisfying $D_1 + D_2 = -1$; see Appendix A.

softened analogues is, evidently, not sufficient for obtaining a well-behaved softened version of Laplace-Lagrange theory for coplanar discs.

4.6.3 Implications for Numerical Applications

In numerical studies of secular dynamics, self-gravitating discs are often treated as a collection of N eccentric annuli (rings), with prescribed spacing (justified by the constancy of the semimajor axis), interacting gravitationally with each other (e.g., Touma et al., 2009; Batygin, 2012). This representation approximates a continuous particulate or fluid disc in the limit of $N \to \infty$ (see e.g. Figure 4.1).

Computational cost associated with the evaluation of mutual ring-ring interactions in this setup, going as $\mathcal{O}(N^2)$, imposes limitations on the number of rings that can be used in practice. This is typically not a problem for the unsoftened calculations, which converge to the expected full disc result even with a relatively coarse radial sampling of the integral contribution to e.g. the free precession rate. Indeed, purple curves in Figures 4.3 and 4.4 demonstrate this by showing the unsoftened $\tilde{\psi}_1(x)$ and $\tilde{\psi}_2(x)$ computed without accounting[15] for the contributions from $a_{d,i} < a_p < a_{d,o}$ (see Section 4.3.2) to the integral terms in the unsoftened expressions of Davydenkova & Rafikov (2018). These curves converge to the correct full disc result without exhibiting large variations in $\tilde{\psi}_1(x)$ and $\tilde{\psi}_2(x)$, typical for softened cases.

On the contrary, the results for the softened gravity presented in Section 4.3.2 do elicit concern about the number of rings N that is needed to accurately capture the eccentricity dynamics of continuous razor-thin discs. Indeed, Figures 4.3 and 4.4 reveal that the expected secular dynamics can be recovered using various softened gravity prescriptions only when one properly accounts for the gravitational effects of all disc annuli, including those very close to the orbit of particle under consideration. Indeed, I demonstrated that to reproduce both the magnitude and the sign of e.g. the free precession rate, the distance Δa separating a given test particle orbit from nearest neighbouring inner and outer disc rings should be quite small, $\Delta a/a_p \lesssim 0.1\varsigma^2$. Only then does the delicate cancellation of large (in magnitude) contributions produced by different parts of the disc recovers the expected (unsoftened) result. Thus, the separation between the modelled disc rings has to be substantially lower than the softening length itself (ςa_p), meaning that N has to be very large, $N \gtrsim 10\varsigma^{-2}$. This could easily make numerical studies of the secular eccentricity dynamics in discs very challenging.

[15]Note that, technically, in the unsoftened case this mathematical procedure is not equivalent to introducing an actual physical gap in the disc, as the latter would result in additional boundary terms.

I further confirmed this expectation by studying the convergence of disc-driven free precession rate in numerically *discretised softened* discs to the precession rate A_d computed exactly for *continuous softened* discs (Equations 4.5 and 4.18). To this end, I represented a given disc model as a collection of N logarithmically spaced rings, and measured the agreement between the radial profiles of theoretical and numerical results for A_d (or ψ_1 for PL discs) by using the following global metric[16]

$$\mathcal{M}(f) = \sqrt{\frac{\int_{a_{\rm in}}^{a_{\rm out}} [f_{\rm theor}(a) - f_{\rm num}(a)]^2 da}{\int_{a_{\rm in}}^{a_{\rm out}} f_{\rm theor}^2(a) da}}. \tag{4.34}$$

Here $f_{\rm num}(a_i)$ is the value of the metric basis (e.g. precession rate A_d) evaluated at the position a_i of ith ring by summing up the contributions of all other rings in the disc, while $f_{\rm theor}(a_i)$ is the analogous quantity computed in the limit of a continuous disc, i.e., as $N \to \infty$ (it is given by the non-discretised version of Equation (4.5) if $f = A_d$, or Equation (4.18) if $f = \psi_1$). Repeating this calculation for various combinations of (N, ς), I can determine the smallest number of rings $N(\varsigma)$ that ensures the desired convergence to within, e.g., ~ 10 per cent (i.e., $\mathcal{M}(f) \sim 0.1$), for a given value of softening ς.

Figure 4.10 depicts a sample of the results obtained using the softening methods of Hahn (2003), Tremaine (1998) and (rectified) Teyssandier & Ogilvie (2016) (see Section 4.6.2) for various axisymmetric disc models as indicated in the legend[17]. Figure 4.10 shows that as $\varsigma \to 0$, the number of rings scales as $N \sim C\varsigma^{-\chi}$ with[18] $C \sim 10$ and $\chi \approx (1.8 - 1.9)$. The only notable exception is the Gaussian ring, for which convergence is faster (i.e., $N \propto \varsigma^{-1.5}$), probably because of mass concentration in a narrow range of radii.

I note that the proportionality constant C in the $N(\varsigma)$ relation is not perfectly defined in the sense that it depends on the (i) desired accuracy (roughly inversely proportional to $\mathcal{M}(f)$), (ii) adopted metric of accuracy (mild dependence), and (iii) softening prescription used — Figure 4.10 shows that discretised calculations using softening model of Hahn (2003) require substantially lower (by ~ 2) number of annuli than those using the models of Teyssandier & Ogilvie (2016) and Tremaine (1998). Nevertheless, these results further reinforce the requirement of large number of rings,

[16] For PL discs, I neglect rings within 10% of disc edges when computing $\mathcal{M}(\psi_1)$.

[17] I exclude the softening method of Touma (2002) from this analysis as it introduces additional complexity due to the nature of softening parameter; $\epsilon^2 = b^2/\max(a_1^2, a_2^2)$, see Section 4.2.1.2.

[18] For example, the curve computed using the (corrected) model of Teyssandier & Ogilvie (2016) has $C = 10.9$ and $\chi = 1.91$, while the one for Quartic disc has $C = 7.2$ and $\chi = 1.75$.

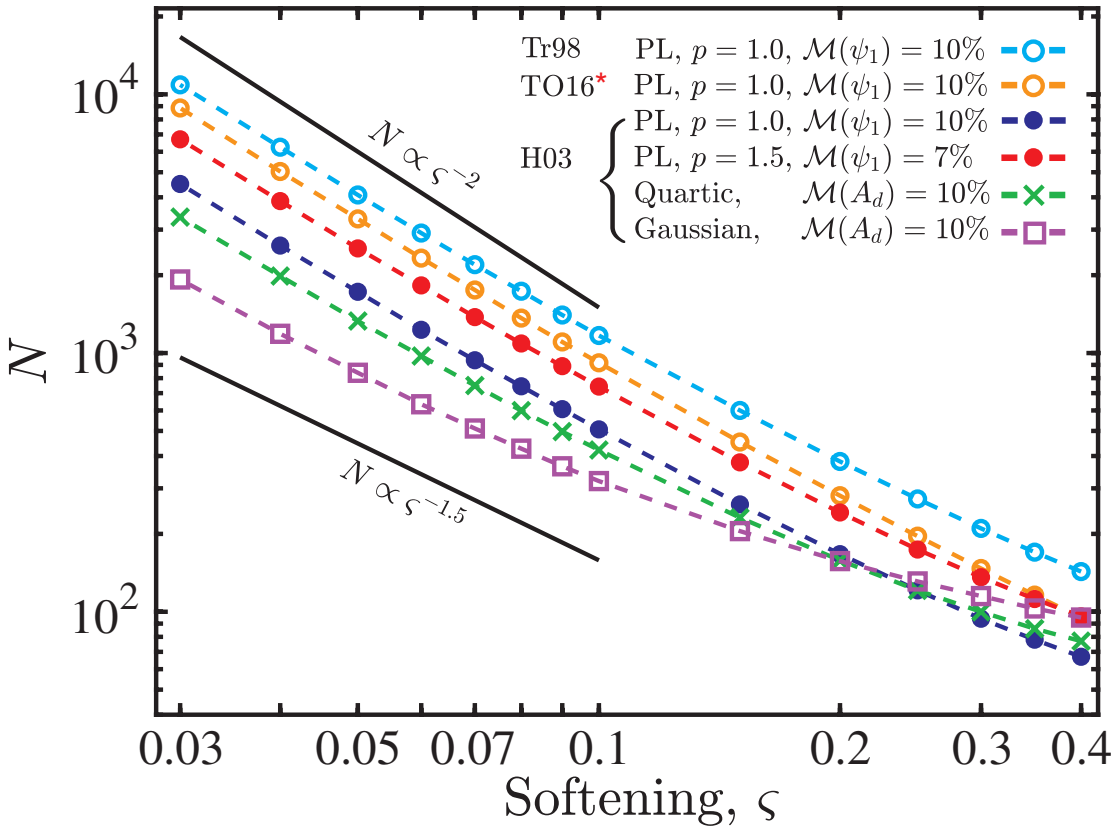

Figure 4.10. Scaling of number of softened annuli (rings) N with softening parameter ς to ensure convergence of disc-driven free precession A_d (or ψ_1) in discretised discs to the expected results in continuous *softened* discs (Equations 4.5 and 4.18). Calculations assume axisymmetric disc models extending from $a_{\rm in} = 0.1$ to $a_{\rm out} = 5$ AU: two PL discs (specified by p), a Quartic disc (same as Figure 4.6) and a Gaussian ring (same as Figure 4.7). I have used the softening methods of Hahn (2003), Tremaine (1998) and (corrected) Teyssandier & Ogilvie (2016), as specified in the panel. Convergence is measured using the metric $\mathcal{M}(f)$ defined by Equation (4.34). One can see that, when $\varsigma \lesssim 0.1$, $N \sim C\varsigma^{-\chi}$, with $C \sim 10$ and $1.5 \lesssim \chi \lesssim 2$. Similar results can be obtained for eccentric discs, and other softening prescriptions. See the text (Section 4.6.3) for details.

with $N \sim \varsigma^{-2}$, to capture the expected secular eccentricity dynamics in nearly Keplerian discs.

Qualitatively similar results were stated in Hahn (2003) who showed that the secular effects of a continuous disc can be recovered only when the disc rings are sufficiently numerous that their radial separation is below the softening length. Although, interestingly, Hahn (2003) and Lee et al. (2019) claimed good convergence of the precession rate to the expected value already for $N \sim \mathcal{O}(\varsigma^{-1})$ (however, note that Lee et al. (2019) also included effects of gas pressure in their calculations, in addition to disc gravity). In my case, the condition on the separation between disc rings motivated by Figures 4.3

and 4.4 (i.e., $\Delta a/a_p \lesssim 0.1\varsigma^2$), along with the results presented in Figure 4.10, indicate that accurate representation of eccentricity dynamics in a cold, razor-thin disc requires a very large number of rings N whenever small values of the softening parameter are used.

As I have shown in Section 4.5, very small values of softening $\varsigma \lesssim 10^{-3}$ are, in fact, necessary to accurately capture eccentricity dynamics near the sharp edges of thin discs. This suggests that N has to be prohibitively large when softened gravity is applied e.g. to study the dynamics of planetary rings (Goldreich & Tremaine, 1979; Chiang & Goldreich, 2000; Pan & Wu, 2016), which are known to have sharp edges.

4.6.4 Further Generalizations and Extensions

All calculations presented in this chapter are based on the expansion of the secular disturbing function R_d due to a coplanar disc — softened and unsoftened — to second order in eccentricities. This approximation may yield inaccurate results when the disc or particle eccentricities are high, e.g., in the vicinity of secular resonances where $A_d(a_p) = 0$ (e.g. Davydenkova & Rafikov, 2018), see Figures 4.6 and 4.7. Such situations may necessitate a higher order extension of the disc potential.

Such an exercise was pursued recently by Sefilian & Touma (2019) who presented a calculation of R_d to fourth order in eccentricities based on the unsoftened method of Heppenheimer (1980). The general framework for calculating R_d with arbitrary softening prescriptions presented in Appendix A can also be extended to higher order in eccentricities in similar way[19]; see e.g. Touma & Sridhar (2012). I expect that conclusions similar to those drawn from my analysis in Sections 4.3–4.5 will also apply to the higher order expansions.

Additionally, although I only analysed coplanar configurations in this chapter, the general framework presented in Appendix A may be extended to account for non-coplanar configurations and study the inclination dynamics.

4.7 Summary

In this chapter, I investigated the applicability of softened gravity for computing the orbit-averaged potential of razor-thin eccentric discs. I compared the disc-driven secular dynamics of coplanar test-particles computed using softening prescriptions available in

[19] Another way to calculate the softened disturbing function for arbitrarily high eccentricities is to numerically compute the ring-ring interaction potential as was done by Touma et al. (2009); see also Sefilian & Touma (2019).

the literature with the calculations based on the unsoftened method of Heppenheimer (1980). My findings can be summarised as follows:

- I confirmed that the softening methods of both Touma (2002) and Hahn (2003) correctly reproduce the eccentricity dynamics of razor-thin discs in the limit of vanishing softening parameter ς for all disc models.
- The softening prescription proposed in Tremaine (1998) yields convergent results as $\varsigma \to 0$. However, quantitative differences of up to $\sim 20 - 30$ per cent from the unsoftened calculations are observed. I demonstrated that these differences arise because of the insufficient smoothness of the interparticle interaction assumed in Tremaine (1998).
- The softening formalism suggested in Teyssandier & Ogilvie (2016) does not result in convergent results in the limit of zero softening.
- Very small values of the (dimensionless) softening parameter are required for correctly reproducing the secular eccentricity dynamics near sharp edges of discs/rings.
- I developed a general analytical framework for computing the secular disturbing function between two coplanar rings with arbitrary interaction potential of rather general form (Equation 4.10). This framework accurately reproduces the orbit-averaged potential of razor-thin discs as $\varsigma \to 0$ for a wide class of softened gravity models.
- Using this general framework, I demonstrated that an accurate implementation of the softened potentials suggested in both Tremaine (1998) and Teyssandier & Ogilvie (2016) leads to the recovery of the expected dynamical behaviour in the limit of small softening.
- My results suggest that the numerical treatments of the secular eccentricity dynamics in softened, nearly Keplerian discs must obey important constraints. Namely, a fine numerical sampling (i.e., large number N of discrete annuli representing the disc, with $N \sim C\varsigma^{-\chi}$ where $C \sim O(10)$ and $1.5 \lesssim \chi \lesssim 2$ is required to ensure that the correct secular behaviour is properly captured by such calculations when ς is small. This finding has important ramifications for numerical treatments of astrophysical discs, especially planetary rings with sharp edges.

In the future, my results for the disc-driven eccentricity dynamics may be extended to higher order in eccentricity, as well as generalised for treating inclination dynamics.

Finally, I close this chapter by pointing out that the results reported here will be useful for my work presented in the forthcoming two chapters (i.e., Chapters 5 and 6), where I investigate the secular interactions between planets and self-gravitating debris discs.

Chapter 5

Planet–Debris Disc Interactions I: A Simplified Model

5.1 Introduction

As already discussed in Chapter 2, spatially resolved images of debris discs frequently reveal complex morphologies such as gaps, spirals, and warps. Most existing models for explaining such morphologies focus on the role of massive perturbers (i.e., planets, stellar companions), ignoring the gravitational effects of the disc itself. Nonetheless, this assumption may not always be justified, especially in view of observations suggesting that debris discs could contain tens of Earth masses in large planetesimals (Wyatt & Dent, 2002; Greaves et al., 2005; Krivov & Wyatt, 2021, see also Section 2.6).

In this and the following chapter, I investigate the long-term, secular interaction between an eccentric planet and an external, *massive* debris disc. The primary aim of this chapter is to present a novel pathway to sculpting gaps, i.e., depleted regions, in broad debris discs. Prior to describing my work, however, it is useful to take a step back and briefly describe the existing mechanisms for sculpting gaps in debris discs, which would also help to put my research into context. The work presented in this chapter has been published in The Astrophysical Journal as Sefilian et al. (2021).

5.1.1 Existing Mechanisms and This Work

As already mentioned in Section 2.3.1, at least four debris discs are known to exhibit double-belt structures that are separated by depleted gaps in their dust distribution as traced by ALMA: HD 107146 (Ricci et al., 2015; Marino et al., 2018a), HD 92945 (Marino et al., 2019), HD 15115 (MacGregor et al., 2019), and HD 206893 (Marino

et al., 2020; Nederlander et al., 2021). These systems (except HD 206893) have no known companions or planets to date, and the discs are gas-poor. In this chapter, I mainly focus on the nearly face-on disc of HD 107146, a nearby $\sim 80-200$ Myr old G2V star (Williams et al., 2004). This disc, extending from ∼30 au to ∼150 au, features a circular ∼40 au wide gap centred at around $70-80$ au in which the continuum emission drops by $\sim 50\%$ (Ricci et al., 2015; Marino et al., 2018a); see Figures 2.6 and 2.7.

Various mechanisms have been explored for explaining the origin of gaps in debris discs. In analogy with the asteroid and Kuiper belts, the most popular scenario involves the presence of single or multiple planets orbiting within the depleted region, which are either stationary or migrating (e.g., Schüppler et al., 2016; Shannon et al., 2016; Zheng et al., 2017; Morrison & Kratter, 2018; Friebe et al., 2022). For instance, it has been suggested that multiple stationary planets or a single but migrating planet of few tens of Earth masses on a near-circular orbit at $\sim 70-80$ au could reproduce HD 107146's gap (e.g., see Ricci et al., 2015; Marino et al., 2018a).

Other scenarios involving planets interior to the disc, rather than embedded within, have also been considered. For instance, Tabeshian & Wiegert (2016) showed that a low-eccentricity planet can carve a gap at its external 2:1 mean motion resonance. On the other hand, Pearce & Wyatt (2015) demonstrated that HD 107146–like discs could be produced as a result of secular interactions and scattering events between a massive ($\sim 10-100 M_{\oplus}$) planetesimal disc and an initially high-eccentricity (~ 0.5) planet of comparable mass to the disc. In the course of evolution, the planetary orbit is then circularised due to scattering events. However, Pearce & Wyatt (2015) consider only the back reaction of the disc on the planet (and vice versa) in their simulations, neglecting the disc self-gravity.

Finally, Yelverton & Kennedy (2018) considered a scenario whereby two coplanar planets carve a gap through their secular resonances within an external debris disc, which was assumed to be *massless*. In their model, the secular resonances occur at sites where the precession rates of the planets (i.e., system's eigenfrequencies) match that of the planetesimals in the disc (due to planetary perturbations). They find that at and around *one* of the two resonant sites, planetesimal eccentricities are excited, triggering a depletion in the disc surface density of the kind seen in HD 107146.

The model proposed by Yelverton & Kennedy (2018) requires (at least) two planets to ensure that their orbits are precessing due to planet-planet interactions, a condition necessary for establishing secular resonances. However, another mechanism that may drive planetary precession is the secular perturbation due to the disc, which was ignored

by Yelverton & Kennedy (2018). This motivates my investigation into whether gaps could be carved in *self-gravitating* debris discs via secular resonances when perturbed by single rather than multiple inner planets. A related scenario was studied by Zheng et al. (2017), who showed that a single planet embedded *within* a decaying gaseous disc (i.e., transitional disc) could carve a wide gap around its orbit via sweeping secular resonances assisted by the waning disc gravity.

In this chapter, I propose that double-ringed structures – akin to that of HD 107146 – could be explained as the aftermath of secular resonances in systems hosting a *single* eccentric planet and an external self-gravitating debris disc. The mechanism I invoke here is different from those of Pearce & Wyatt (2015) and Yelverton & Kennedy (2018). My mechanism is realised through a secular resonance between the apsidal precession rate of planetesimals due to *both* the disc and planet, and that of the planet due to *disc gravity* (see, Yelverton & Kennedy, 2018). Additionally, my mechanism does not require scattering events between the planet and disc particles (see, Pearce & Wyatt, 2015). As I show below, this mechanism is robust over a wide range of parameters; particularly when the disc is less massive than the planet.

The rest of this chapter is organised as follows. In Section 5.2, I first describe my model system and then present the equations governing the secular evolution of the system. In Section 5.3, I describe the essential features of planetesimal dynamics in the combined planet–disc potential. I then characterise the features of the secular resonances over a wide range of parameter space in Section 5.4. In Section 5.5 I apply these considerations to HD 107146, and identify the planet–disc parameters that could reproduce the observed gap. In Section 5.6, using some of these parameters, I investigate the evolution of planet–disc systems and present my main results. I discuss my results along with their observational and theoretical implications in Section 5.7, where I also consider the application of my results to other systems; namely, HD 92945 and HD 206893. Finally, I critically asses the limitations of my model in Section 5.8, and summarise my findings in Section 5.9.

5.2 Analytical Model

I now describe a simple model to analyse the long-term dynamical evolution of planetesimals embedded within a massive debris disc in a single-planet system. In my notation, a planetesimal orbit is characterised by its semimajor axis a, eccentricity e, and longitude of pericentre ϖ. Orbital elements subscripted with 'p' and 'd' refer to the planet and the disc, respectively.

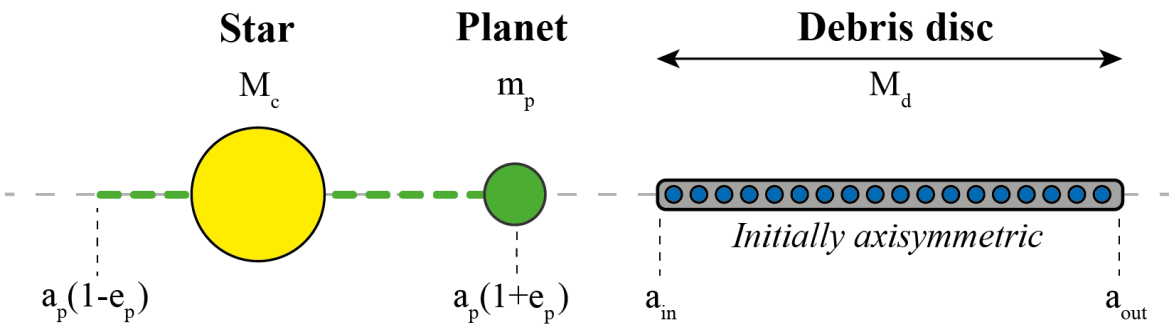

Figure 5.1. A sketch of the initial setup of the model system considered in Chapter 5. A central star of mass M_c is orbited by a planet of mass m_p and an external debris disc of mass M_d ($M_d, m_p \ll M_c$). The planetary orbit (dashed green line), characterised by its semimajor axis a_p, is slightly eccentric ($e_p \leq 0.1$) such that it does not cross the disc along its orbit. The disc, extending from $a_{\rm in}$ to $a_{\rm out} \gg a_{\rm in}$, is razor-thin and axisymmetric characterised by a surface density profile given by Equation (5.1). The entire star-planet-disc system is coplanar.

5.2.1 Model System

My model system consists of a broad debris disc of mass M_d orbiting the host star M_c exterior to, and coplanar with, a planet of mass m_p ($M_d, m_p \ll M_c$). I assume that the planet is initially on a low-eccentricity orbit ($e_p \leq 0.1$) and that it does not intersect the disc along its orbit. I consider the debris disc to be razor-thin and initially axisymmetric. Figure 5.1 illustrates a sketch of the model system's initial configuration. The disc surface density is characterised by a truncated power-law profile given by

$$\Sigma_d(a) = \Sigma_0 \left(\frac{a_{\rm out}}{a}\right)^p \tag{5.1}$$

for $a_{\rm in} \leq a \leq a_{\rm out}$, and $\Sigma_d(a) = 0$ elsewhere. Here, $a_{\rm in}$ and $a_{\rm out}$ are the semimajor axes of the inner and outer disc edges, respectively. Defining $\delta \equiv a_{\rm out}/a_{\rm in} > 1$, the total mass M_d of such a disc can be written as

$$M_d = \frac{2\pi}{2-p}\Sigma_0 a_{\rm out}^2 \left(1 - \delta^{p-2}\right), \tag{5.2}$$

which allows me to express Σ_d in terms of M_d. This setup is very similar to that explored in Rafikov (2013) and Silsbee & Rafikov (2015b) in the context of planetesimal dynamics in circumbinary protoplanetary discs.

In this chapter, unless otherwise stated, I adopt a fiducial disc model with $p = 1$, $a_{\rm in} = 30$ au, and $a_{\rm out} = 150$ au (i.e., $\delta = 5$). This choice of p corresponds to a disc with a constant amount of mass per unit semimajor axis.

5.2.2 Secular Gravitational Effects

I am primarily interested in the long-term dynamics of large (i.e., ∼kilometre-sized) planetesimals. Since the latter are effectively insensitive to radiative non-gravitational forces (Section 2.2), I focus purely on gravitational effects[1] accounting for perturbations due to *both* (1) the debris disc and (2) the planet. For simplicity, however, the non-axisymmetric component of the disc gravity is ignored in this chapter, although, as we will see later (Section 5.6), the disc naturally develops non-axisymmetry (this limitation will be relaxed in Chapter 6). I perform calculations within the framework of secular (orbit-averaged) perturbation theory to second order in eccentricities (Murray & Dermott, 1999, see also Chapter 3).

5.2.2.1 Effects of the Disc and Planet on Planetesimals

The secular dynamics of planetesimals is described by the disturbing function R, which consists of contributions due to the planet R_p and due to the disc R_d. An analytic expression for the disturbing function R_d due to an axisymmetric disc with surface density of the form given by Equation (5.1) has been previously derived in Silsbee & Rafikov (2015a); see also Chapter 4 and the references therein. Combining R_d with the contribution R_p due to the planet (e.g. Murray & Dermott, 1999, Equation (7.7)), the total disturbing function $R = R_d + R_p$ to second order in eccentricities reads as:

$$R = na^2 \left[\frac{1}{2} A e^2 + B_p e \cos\left(\varpi - \varpi_p\right) \right], \tag{5.3}$$

where $n = \sqrt{GM_c/a^3}$ is the planetesimal mean motion, and the meaning of different constants is explained below.

In Equation (5.3), $A = A_d + A_p$ is the precession rate of the free eccentricity vector of a planetesimal. It has contributions from both the gravity of the disc (A_d) and the planet (A_p). The contribution of the planet is (Murray & Dermott, 1999)

$$\begin{aligned} A_p &= \frac{1}{4} n \frac{m_p}{M_c} \frac{a_p}{a} b^{(1)}_{3/2}(a_p/a), \\ &\approx 35.5 \times 10^{-2}\ \mathrm{Myr}^{-1} \frac{m_p}{0.6 M_J} a^2_{p,20}\ a_{70}^{-7/2} M_{c,1.09}^{-1/2}, \end{aligned} \tag{5.4}$$

[1] Note that while non-gravitational thermal effects such as Yarkovsky and YORP may also influence the evolution of ∼km-sized planetesimals (Section 2.2.4), these can be neglected for the purposes of my study. This is because (i) the Yarkovsky effect induces e.g. a semimajor axis drift of only $\lesssim 10^{-5} - 10^{-6}$ au/Myr for large planetesimals at $\sim 10 - 100$ au (e.g., Vokrouhlický et al., 2015; Xu et al., 2020, and references therein), while (ii) the YORP effect modifies the rotation rates and obliquities of bodies, which I do not model here.

where $a_{p,20} \equiv a_p/(20 \text{ au})$, $a_{70} \equiv a/(70 \text{ au})$, $M_{c,1.09} \equiv M_c/(1.09 M_\odot)$, $b_s^{(m)}(\alpha)$ is the Laplace coefficient given by Equation (4.2), and the numerical estimate in Equation (5.4) assumes $a_p/a \ll 1$ so that $b_{3/2}^{(1)}(\alpha) \approx 3\alpha$. The contribution of the disc to the free precession is (Silsbee & Rafikov, 2015a, see also Equation (4.20))

$$\begin{aligned} A_d &= 2\pi \frac{G\Sigma_d(a)}{na}\psi_1 = (2-p)n\frac{M_d}{M_c}\left(\frac{a}{a_{\text{out}}}\right)^{2-p}\frac{\psi_1}{1-\delta^{p-2}} \\ &\approx -14.4\times 10^{-2}\text{ Myr}^{-1}\frac{M_d}{20M_\oplus}a_{70}^{-1/2}\frac{M_{c,1.09}^{-1/2}}{a_{\text{out},150}}\frac{\psi_1}{-0.5}, \end{aligned} \tag{5.5}$$

where $a_{\text{out},150} \equiv a_{\text{out}}/(150 \text{ au})$, and the numerical estimate is for $p = 1$ and $\delta \gg 1$ such that $\psi_1 \approx -0.5$ (e.g. Figure 4.5).

In general, the coefficient ψ_1 in Equation (5.5) depends on the power-law index p as well as the planetesimal semimajor axis with respect to the disc edges (Chapter 4, see also Equation (A33) in Silsbee & Rafikov (2015a)). As the sharp edges of the disc are approached, ψ_1 formally diverges (see Figure 4.8). However, when the planetesimal is well separated from the edges (i.e., $a_{\text{in}} \ll a \ll a_{\text{out}}$), ψ_1 is effectively a constant of order unity (depending on p, see Figure 4.5), which can be well approximated by Equation (A37) in Silsbee & Rafikov (2015a). It is very important to note that the disc and the planet drive planetesimal precession in opposite directions, $A_p > 0$ and $A_d < 0$, with $A_p(a)$ falling off more rapidly with a than $|A_d(a)|$.

The term B_p in Equation (5.3) represents the excitation of planetesimal eccentricity due to the non-axisymmetric component of the planetary potential. It is given by (Murray & Dermott, 1999)

$$B_p = -\frac{1}{4}n\frac{m_p}{M_c}\frac{a_p}{a}b_{3/2}^{(2)}(a_p/a)e_p. \tag{5.6}$$

Note that the analogous term due to the disc is absent in Equation (5.3), since I have neglected the non-axisymmetric component of the disc self-gravity.

5.2.2.2 Effect of the Disc on Planet

Next I consider the effect of the disc on the planet. Since the disc is taken to be axisymmetric, it simply causes the planetary apsidal angle to advance linearly in time such that $\varpi_p(t) = A_{d,p}t + \varpi_p(0)$, i.e., $\dot{\varpi}_p = A_{d,p}$, without exchanging its angular momentum with the planet. In this chapter, without loss of generality, I set $\varpi_p(0) = 0$. In Appendix E, I show that the planetary precession rate $A_{d,p}$ due to the disc with surface density of the form given by Equation (5.1) can be written as (see also, Petrovich

et al., 2019):

$$\begin{aligned} A_{d,p} &= \frac{3}{4} n_p \frac{2-p}{p+1} \frac{M_d}{M_c} \left(\frac{a_p}{a_{\rm out}}\right)^3 \frac{\delta^{p+1}-1}{1-\delta^{p-2}} \phi_1^c \\ &\approx 19.4 \times 10^{-2}\ {\rm Myr}^{-1} \frac{M_d}{20 M_\oplus} \frac{a_{p,20}^{3/2}}{a_{\rm out,150}\ a_{\rm in,30}^2} M_{c,1.09}^{-1/2}, \end{aligned} \tag{5.7}$$

where $n_p = \sqrt{GM_c/a_p^3}$ is the planetary mean motion, $a_{\rm in,30} \equiv a_{\rm in}/(30\ {\rm au})$, and the numerical estimate is for $p = 1$ and $a_p = 20$ au such that $\phi_1^c \approx 1.8$. Here $\phi_1^c = \phi_1^c(a_p/a_{\rm in}, p, \delta)$ is a factor of order unity accounting for contributions of the disc annuli close to the planet (Equation E.7). I show its behaviour as a function of $a_p/a_{\rm in}$ and for various disc models (i.e., p, δ) in Figure E.1. For $a_p/a_{\rm in} \ll 1$, one has $\phi_1^c \approx 1$ regardless of $(p,\ \delta)$.

5.2.2.3 Combined Planet–Disc Effects

The fact that the planet is precessing renders the forcing term in R (Equation 5.3) time-dependent. This time dependence could be eliminated upon transferring to a frame precessing with the planetary orbit, i.e., by subtracting $\Phi A_{d,p}$ from Equation (5.3) where $\Phi = na^2\left(1-\sqrt{1-e^2}\right) \approx na^2e^2/2$ is the action conjugate to the angle $\Delta\varpi \equiv \varpi - \varpi_p$. As a result, I obtain the following expression:

$$R = na^2 \left[\frac{1}{2}\left(A - A_{d,p}\right) e^2 + B_p e \cos\Delta\varpi\right]. \tag{5.8}$$

This completes my development of the disturbing function.

It is worthwhile to note that for the particular set of parameters in equations (5.4), (5.5), and (5.7), the planetesimal free precession rate A at $a = 70$ au is comparable to that of the planetary orbit, $A_{d,p}$. In Figure 5.2 I show the radial behaviour of $A = A_d + A_p$, together with the curve for $A_{d,p}$. The fact that $A(a) = A_{d,p}$ at certain semimajor axes has very important implications for planetesimal dynamics; see Section 5.3.

5.2.3 Evolution Equations and Their Solution

The secular evolution of a planetesimal orbit in the combined potential of the planet and the disc can be determined by Lagrange's planetary equations (Murray & Dermott, 1999, see also Section 3.2.3). Introducing the eccentricity vector $\mathbf{e} = (K, H) = e(\cos\Delta\varpi, \sin\Delta\varpi)$, convenient for describing the dynamics in the frame corotating

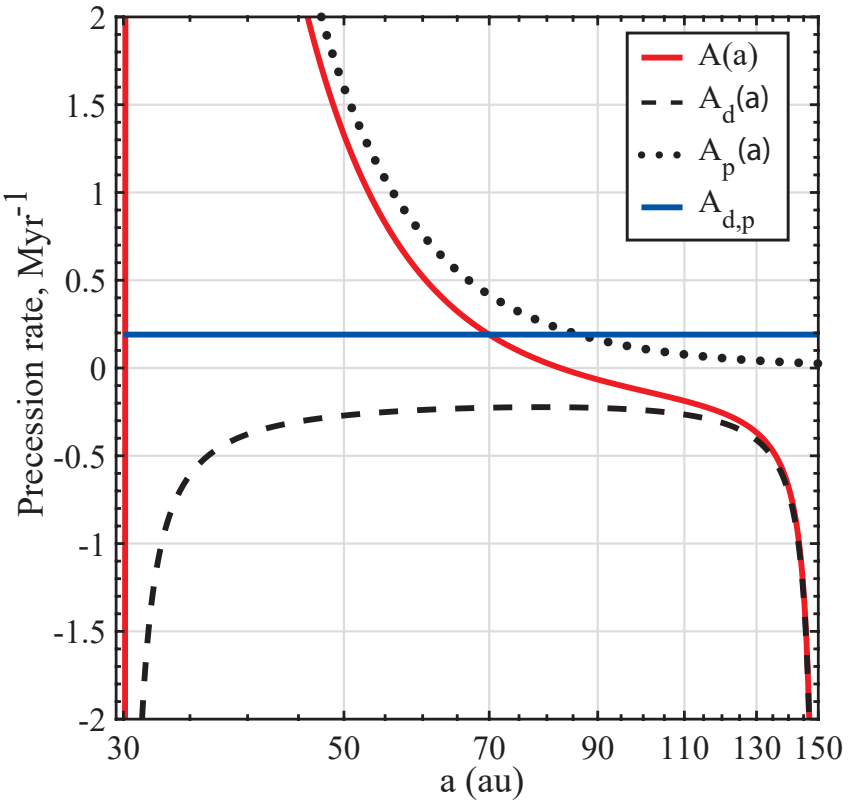

Figure 5.2. Planetesimal free precession rate $A = A_d + A_p$ due to both the planet and the disc as a function of semimajor axis (red curve). Dotted and dashed curves represent $A_p(a)$ and $A_d(a)$, respectively. The blue line represents the rate of planetary precession $A_{d,p}$ due to the disc. Calculations assume a $20M_\oplus$ disc with $p = 1$ extending from $a_{\rm in} = 30$ au to $a_{\rm out} = 150$ au, and a $0.6M_J$ planet at $a_p = 20$ au around a $1.09M_\odot$ star (`Model A`, Table 5.1). Note that $A(a) = A_{d,p}$ at two locations: at 70 au and at $\simeq a_{\rm in}$.

with the planet (e.g. Heppenheimer, 1980), I find that:

$$\begin{aligned} \frac{dK}{dt} &\approx \frac{-1}{na^2}\frac{\partial R}{\partial H} = -(A - A_{d,p})H, \\ \frac{dH}{dt} &\approx \frac{1}{na^2}\frac{\partial R}{\partial K} = (A - A_{d,p})K + B_p. \end{aligned} \tag{5.9}$$

Note that in the case of a massless disc ($A_{d,p} = 0, A = A_p$), one recovers the evolution equations due to a non-precessing perturbing planet (e.g. Murray & Dermott, 1999).

The system of equations (5.9) admits a general solution given by the superposition of the 'free' and 'forced' eccentricity vectors, $\mathbf{e}(t) = \mathbf{e}_{\rm free}(t) + \mathbf{e}_{\rm forced}(t)$ (Murray & Dermott, 1999, see also Section 3.2.4.1). In particular, when planetesimals are initiated on circular orbits, $K(0) = H(0) = 0$, one has $e_{\rm free} = e_{\rm forced}$ and the evolution of

planetesimal orbits is described by:

$$e(t) = 2\left|e_{\rm forced}\sin\left(\frac{A-A_{d,p}}{2}t\right)\right|, \tag{5.10}$$

$$\tan\Delta\varpi(t) = \tan\left(\frac{A-A_{d,p}}{2}t-\frac{\pi}{2}\right), \tag{5.11}$$

where $\Delta\varpi$ stays in the range $[-\pi,\pi]$, and the forced eccentricity is given by

$$e_{\rm forced}(a)=\frac{-B_p(a)}{A(a)-A_{d,p}}=\frac{-B_p(a)}{A_d(a)+A_p(a)-A_{d,p}}. \tag{5.12}$$

Equations (5.10)–(5.12) represent the key solutions needed for my work. I remark that this framework has been previously verified against direct orbit integrations of test particles in discs (e.g., Chapter 4; Silsbee & Rafikov, 2015a; Fontana & Marzari, 2016; Davydenkova & Rafikov, 2018).

For illustrative purposes, in Figure 5.3 I show the radial profiles of instantaneous eccentricities (left panels) and longitudes of pericentre (relative to the planet, right panels) of planetesimals computed using Equations (5.10) and (5.11) (i.e., for $e(0)=0$) at different times, as indicated in each panel. The calculations assume the same disc–planet parameters as in Figure 5.2, and I have taken $e_p=0.05$ – the parameters of the fiducial disc–planet model (`Model A`, Table 5.1) that I consider in detail later in this chapter (Section 5.6). Furthermore, here I have sampled secular evolution using $N=5000$ planetesimals with semimajor axes distributed logarithmically between $a_{\rm in}$ and $a_{\rm out}$, i.e., with a ratio of spacing $\beta=(a_{\rm out}/a_{\rm in})^{1/N}\approx 1.0003$, each of which is represented by a blue dot in Figure 5.3. I note that, as is typical for secular evolution, the eccentricity oscillation at a given semimajor axis is bounded between the initial value of 0 and $e_m(a)=2|e_{\rm forced}(a)|$ (the red lines in left panels of Figure 5.3). Moreover, as expected, the period of each eccentricity oscillation in the frame corotating with the planet is given by $\tau_{\rm sec}=2\pi/(A-A_{d,p})$.

5.3 Planetesimal Eccentricity Behaviour and Secular Resonances

I now describe the essential features of planetesimal dynamics in the combined disc–planet potential[2]. In general, planetesimal orbits evolve differently depending on their

[2]For a detailed summary of the dynamics in an analogous setup (in application to planetesimal dynamics in circumbinary protoplanetary discs), see Rafikov (2013) and Silsbee & Rafikov (2015b).

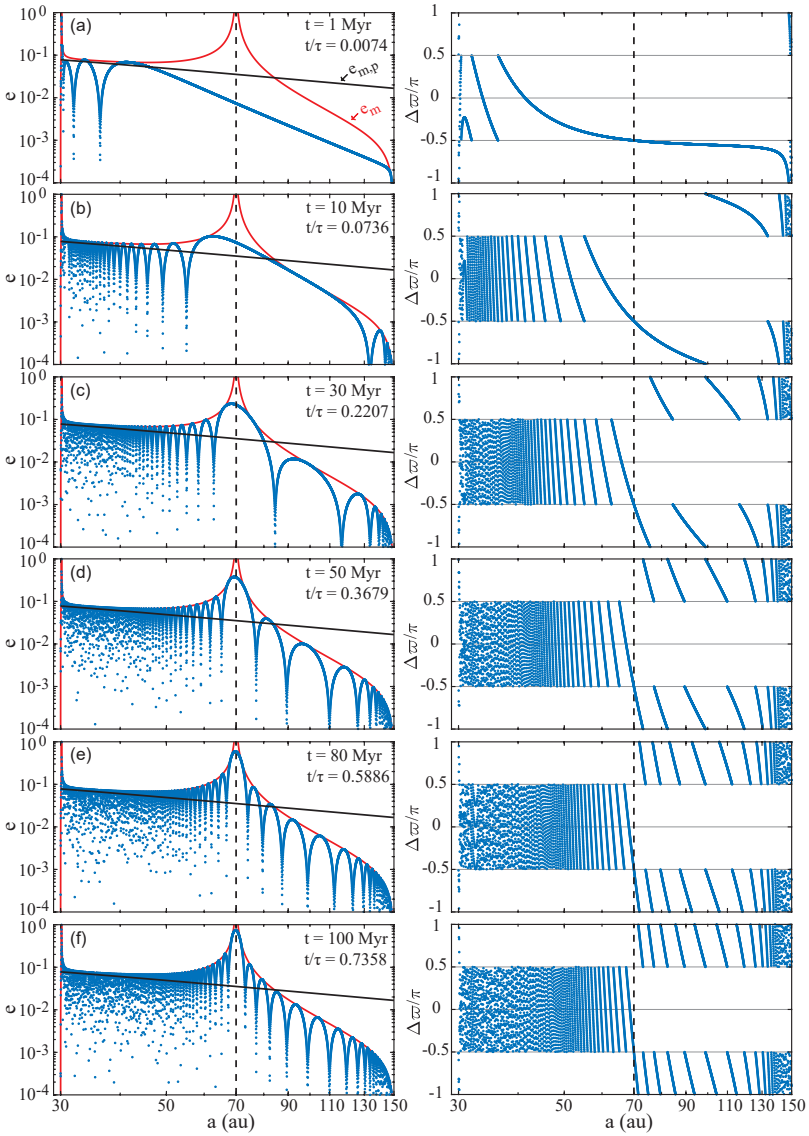

Figure 5.3. Snapshots of the planetesimal eccentricities e (left panels) and apsidal angles $\Delta\varpi$ (right panels, measured relative to that of the precessing planet) as a function of semimajor axis a after $t = 1, 10, 30, 50, 80$, and 100 Myr of evolution (top to bottom). The time is also indicated relative to $\tau \approx 135$ Myr, Equation (5.15). The planetesimals were initiated on circular orbits in the fiducial disc–planet model (`Model A`, Table 5.1). The maximum of eccentricity oscillations $e_m = 2|e_{\rm forced}|$ (Equation 5.12) is shown by the red lines. For reference, the solid black lines show the maximum planetesimal eccentricities driven by the planet in the absence of the disc ($e_{m,p}$, Equation 5.13). The dashed vertical lines show the secular resonance location ($a_{\rm res} = 70$ au), where eccentricities diverge in the course of evolution. One can clearly see that at the resonance $\Delta\varpi = -\pi/2$ at all times. Note also the resonance near the disc inner edge. An animated version of this figure has been made available online.

free precession rate $A(a)$ relative to that of the planet $A_{d,p}$, i.e., for $A(a) > A_{d,p}$ or $A(a) < A_{d,p}$ – see Equations (5.10) and (5.11).

For the particular set of parameters in Figures 5.2 and 5.3, I find that the regime $A(a) > A_{d,p}$ is realised at small separations from the planet, where the precession rate of planetesimals is dominated by the planet so that $A \approx A_p$ (except near $a_{\rm in}$ where A_d diverges due to disc edge effects; see e.g. Chapter 4 and Silsbee & Rafikov (2015a)); see also Equations (5.4) and (5.5). In this *planet-dominated* regime, planetesimal orbits precess in the same direction as the planet (i.e., prograde, see Equation 5.11 and right panels of Figure 5.3), and one has $e_{\rm forced} > 0$ (Equation 5.12). Thus, as planetesimal orbits evolve, the apsidal angles $\Delta\varpi$ remain constrained within $[-\pi/2, \pi/2]$ at all times. Moreover, planetesimals attain their maximum eccentricity when their orbits are aligned with that of the planet, i.e., when $\Delta\varpi = 0$; see Equation (5.10) and Figure 5.3. Assuming $A_p \gtrsim A_{d,p}$, the maximum planetesimal eccentricity in this regime is $e_{m,p} \approx |2e_{\rm forced,p}|$ with (e.g. Murray & Dermott, 1999)

$$e_{\rm forced,p} = \frac{-B_p}{A_p} = \frac{b^{(2)}_{3/2}(a_p/a)}{b^{(1)}_{3/2}(a_p/a)} e_p \approx \frac{5}{4}\frac{a_p}{a} e_p, \approx 1.8\times10^{-2}\frac{a_{p,20}}{a_{70}}\frac{e_p}{0.05}, \tag{5.13}$$

see Equation (5.12), where I have used the approximations $b^{(1)}_{3/2}(\alpha) \approx 3\alpha$ and $b^{(2)}_{3/2}(\alpha) \approx (15/4)\alpha^2$ valid for small α. This is the limit of a *massless* disc, a configuration most often adopted in studies of debris discs (see e.g. Sections 2.4 and 2.6). In the course of evolution, planetesimals in this regime will form an eccentric structure largely aligned with the planetary orbit (e.g. Wyatt et al., 1999).

In the opposite *disc-dominated* limit, far from the planet (and for $a \approx a_{\rm in}$, which I discuss later), Figure 5.2 shows that the precession rate of planetesimals is dominated by the disc so that $A \approx -|A_d| \lesssim A_{d,p}$. In this regime, planetesimal orbits undergo retrograde free precession (see Equation 5.11 and the right panels of Figure 5.3), and one has $e_{\rm forced} < 0$. Thus, the apsidal angles $\Delta\varpi$ are confined within the range $\pm[\pi/2, \pi]$ at all times. Moreover, planetesimals attain their maximum eccentricity when their orbits are *anti-aligned* with the planetary orbit, i.e., when $|\Delta\varpi| = \pi$; see Equation (5.10). Assuming $A_{d,p} \to 0$ for simplicity, the maximum eccentricity in this regime is

$e_{m,d} \approx |2e_{\rm forced,d}|$ with

$$|e_{\rm forced,d}| = \left|\frac{B_p}{A_d}\right| \approx \frac{15e_p}{16|(2-p)\psi_1|}\frac{m_p}{M_d}\left(\frac{a_p}{a}\right)^3\left(\frac{a_{\rm out}}{a}\right)^{2-p}, \quad (5.14)$$
$$\approx 4.7\times10^{-3}\frac{m_p}{M_d}\frac{e_p}{0.05}\frac{a_{p,20}^3 a_{\rm out,150}}{a_{70}^4},$$

where the numerical estimate assumes $p = 1$ and $a_{\rm in} \ll a \ll a_{\rm out}$ so that $\psi_1 \approx -0.5$. Equation (5.14) shows that planetesimal eccentricities in the disc-dominated regime decline more rapidly with a than in the planet-dominated regime, and their magnitude is suppressed – an effect pointed out in Rafikov (2013). In the course of evolution, planetesimals in this regime will form an eccentric structure anti-aligned with the planetary orbit.

5.3.1 Main Secular Resonance

More importantly, one can clearly see that the transition between planet- and disc-dominated regimes occurs via a *secular eccentricity resonance* where $A(a) = A_{d,p}$; see Figure 5.2 (see also Rafikov, 2013; Silsbee & Rafikov, 2015b). This resonance emerges because the relative precession between the planetesimal orbits and the planetary orbit vanishes, while the torque exerted by the non-axisymmetric component of the planet is nonzero. At and around the locations of secular resonances, $a = a_{\rm res}$, planetesimal eccentricities are forced to arbitrarily large values (in linear approximation); see the left panels of Figure 5.3. This is because the denominator in Equation (5.12) becomes small, introducing a singularity into the secular solution[3] (Rafikov, 2013). By taking a limit $A(a_{\rm res}) \to A_{d,p}$ in Equation (5.12), I find that the growth of eccentricity at the resonance occurs linearly in time, $e(t) = t/\tau$, with a characteristic timescale given by

$$\tau = \frac{1}{|B_p(a_{\rm res})|} \approx 158\ {\rm Myr}\frac{0.6M_J}{m_p}\frac{0.05}{e_p}\frac{a_{\rm res,70}^{9/2}}{a_{p,20}^3}M_{c,1.09}^{1/2}, \quad (5.15)$$

where the approximation is valid for $a_p \ll a_{\rm res}$. Equation (5.15) also explains why the eccentricities at the resonance near the disc inner edge are pumped up more quickly than at the resonance at 70 au, see the left panels of Figure 5.3.

Moreover, one can see from the right panels of Figure 5.3 that at the resonance, $\Delta\varpi$ remains fixed at $-\pi/2$ as expected from Equation (5.11); see the right panels of Figure

[3]Including higher-order terms (in eccentricities) of the disturbing function of Equation (5.8) imposes a finite upper limit on the amplitude of $e_{\rm forced}$ at secular resonance (Malhotra, 1998; Ward & Hahn, 1998a).

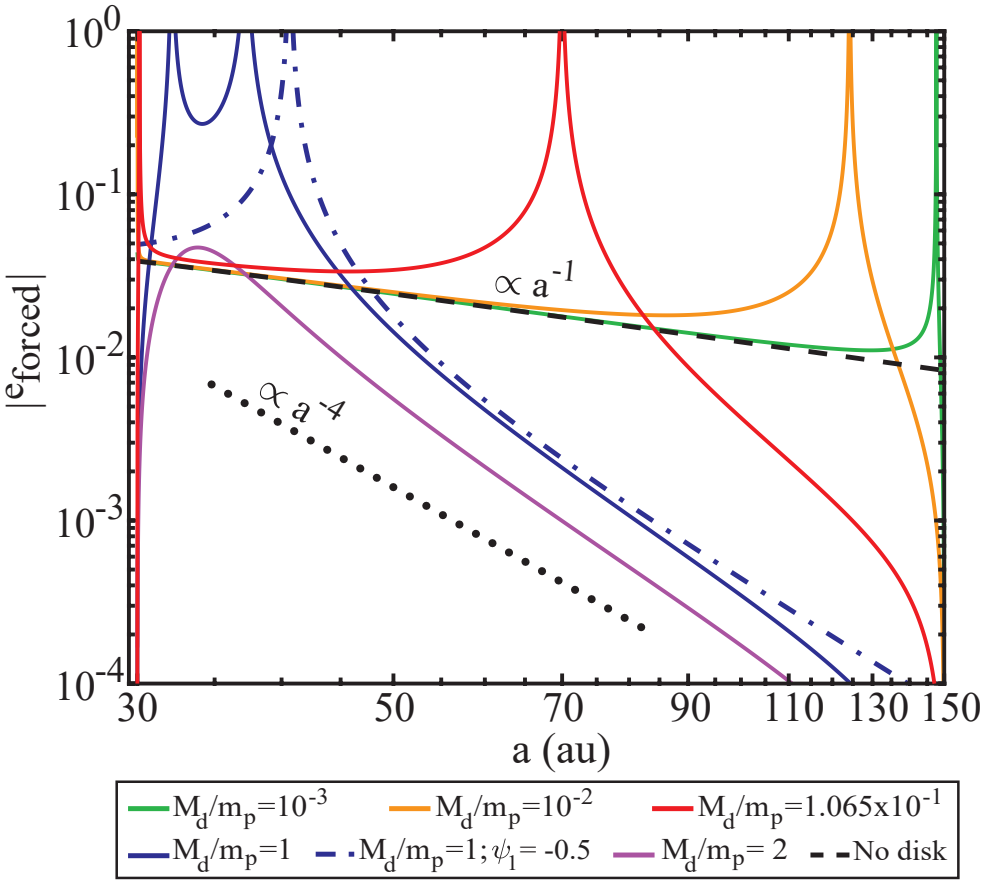

Figure 5.4. Forced eccentricities of planetesimals as a function of their semimajor axis a, computed for different values of M_d/m_p (with fixed $m_p = 0.6M_J$). The calculations assume all other system parameters are the same as in Figures 5.2 and 5.3. All of these curves scale linearly with the planetary eccentricity e_p, which I have taken to be 0.05 in this calculation. For reference, the black dashed line shows forced eccentricity in the case of a massless disc $e_{\rm forced,p}$ (Equation 5.13), and the dotted line illustrates the asymptotic behaviour of eccentricity given by $e_{\rm forced,d}$ (Equation 5.14). Note the occurrence of two secular resonances for $10^{-3} \leq M_d/m_p \leq 1$, with one of them being near the inner disc edge. See the text (Section 5.3) for details.

5.3. In Section 5.4.1 I will show that such secular resonances are generic: they occur for a large range of disc-to-planet mass ratios, $10^{-4} \lesssim M_d/m_p \lesssim 2$, for all $a_p \lesssim a_{\rm in}$.

To further illustrate the analysis above, Figure 5.4 shows the radial profiles of planetesimal forced eccentricities computed for different values of disc mass. The calculations are done for the same planetary parameters as in Figures 5.2 and 5.3. The most pronounced feature in Figure 5.4 is the occurrence of a secular resonance within the disc (apart from the one very close to $a_{\rm in}$, see below) for $10^{-3} \leq M_d/m_p \leq 1$, where $e_{\rm forced}$ diverges. At the same time, $e_{\rm forced}$ asymptotically approaches $e_{\rm forced,p}$ inward of

the resonance, i.e., where $A \gtrsim A_{d,p}$, whereas $e_{\text{forced}} \to e_{\text{forced,d}}$ external to it, i.e., where $A \lesssim A_{d,p}$ (which is, of course, possible only if $a_{\text{in}} \lesssim a_{\text{res}} \lesssim a_{\text{out}}$). At the highest disc mass, $M_d/m_p = 2$, there are no secular resonances as the disc dominates planetesimal precession throughout the whole disc.

I note that in the region where the dynamics is dominated by the disc, $e_{\text{forced}}(a)$ does not follow the simple power-law profile $\propto a^{-4}$ given by Equation (5.14). By and large, this is because the disc edge effects neglected in computing Equation (5.14) render $\psi_1 = \psi_1(a)$ in a nontrivial manner, even when $a_{\text{in}} \lesssim a \lesssim a_{\text{out}}$ (Silsbee & Rafikov, 2015a). For instance, it is evident in Figure 5.2 that $A_d(a)$ behaves more like a constant for $a_{\text{in}} \ll a \ll a_{\text{out}}$ rather than as $A_d \propto a^{-1/2}$ (Equation 5.5), implying that $|\psi_1| \propto a^{1/2}$ for the employed disc model. This will be important in Section 5.4.1. As a matter of fact, ψ_1 becomes independent of semimajor axis only in discs of infinite radial extent (Chapter 4; Silsbee & Rafikov, 2015a), whereas the radial range of the disc that I have adopted is finite with $\delta = a_{\text{out}}/a_{\text{in}} = 5$ (Section 5.2.1).

5.3.2 Secular Resonance at a_{in}

Finally, I clarify that the origin of the resonance at $\approx a_{\text{in}}$ (apart from the one at $\gtrsim a_{\text{in}}$) lies in the fact that $A_d \propto -|\psi_1|$ diverges as the sharp edges of a razor-thin disc are approached; see the black dashed lines in Figure 5.2. This makes $|A_d(a)| \sim A_p(a)$ as $a \to a_{\text{in}}$, even for a modest value of disc mass. However, it is also known that discs with Σ_d dropping continuously near the edges rather than discontinuously, or discs with small but nonzero thickness, should exhibit finite A_d near the edges (Chapter 4; Davydenkova & Rafikov, 2018); different from my disc model. Thus, in such more realistic discs, only a single resonance – rather than two – will occur, as we shall see later in Chapter 6. For now, this is portrayed in Figure 5.4 for $M_d/m_p = 1$ by artificially stipulating $\psi_1(a) = -0.5$, i.e. by ignoring the edge effects (Chapter 4; Silsbee & Rafikov, 2015a).

5.3.3 Secular Resonances and Gaps in Debris Discs

To summarise, the analysis presented in this section elucidates that the disc gravity can have a considerable impact on the secular evolution of planetesimals. In the remainder of this chapter, I exploit the feasibility of the discussed secular resonance as the basis of a mechanism for sculpting depleted regions, i.e., gaps, in debris discs.

The emergence of a gap could be understood as follows. Planetesimals on eccentric orbits spend most of their time near their apocentre, farther away from their orbital

semimajor axes. Thus, provided that a secular resonance occurs within the disc, I expect the surface density of planetesimals to be depleted around the resonance location where planetesimal eccentricities grow without bound. This reasoning, in essence, is similar to that presented by Yelverton & Kennedy (2018) where the authors show that two planets could carve a gap in an external *massless* debris disc through their secular resonances. Additionally, given that generally planetesimals in the inner disc parts tend to apsidally align with the planet while those in the outer parts tend to anti-align, I expect the depleted region to have a non-axisymmetric shape. This effect has been previously pointed out by Pearce & Wyatt (2015) in the context of secular interaction between a debris disc and an interior, precessing planet.

5.4 Characterisation of Secular Resonances

I now investigate how the characteristics of the secular resonances – i.e., their locations, their associated timescales for exciting eccentricities, and their widths – depend on the properties of the disc and the planet. This will guide me in putting constraints on the possible disc–planet parameters that could reproduce the structure of an observed debris disc featuring a gap, as will be done in Section 5.5 for HD 107146.

5.4.1 Location of Secular Resonances

As mentioned in Section 5.3, secular resonances occur at semimajor axes $a = a_{\rm res}$ where the apsidal precession rates of both the planet and planetesimals are commensurate,

$$A_d(a_{\rm res}) + A_p(a_{\rm res}) = \dot{\varpi}_p \equiv A_{d,p}. \tag{5.16}$$

Using Equations (5.4), (5.5), and (5.7), I can express the resonance condition of Equation (5.16) in terms of the disc-to-planet mass ratio M_d/m_p and the relevant semimajor axes, i.e., $a_{\rm res}$, a_p, and $a_{\rm out}$, scaled by $a_{\rm in}$:

$$C_1\psi_1\frac{M_d}{m_p}\left(\frac{a_{\rm res}}{a_{\rm in}}\right)^{2-p} + \frac{1}{4}\frac{a_p}{a_{\rm res}}b^{(1)}_{3/2}\left(\frac{a_p}{a_{\rm res}}\right) = \frac{3}{4}C_2\phi_1^c\frac{M_d}{m_p}\left(\frac{a_p}{a_{\rm in}}\right)^3\left(\frac{a_p}{a_{\rm res}}\right)^{-3/2}. \tag{5.17}$$

Here $C_1 = (2-p)/(\delta^{2-p}-1)$ and $C_2 = C_1(1-\delta^{-p-1})/(p+1)$ are constants that depend on the disc model. It follows from Equation (5.17) that the locations of secular resonances can be computed *relative* to the disc inner edge as functions of $a_p/a_{\rm in}$ and M_d/m_p. This is illustrated in Figure 5.5, where I plot the contours of M_d/m_p in the

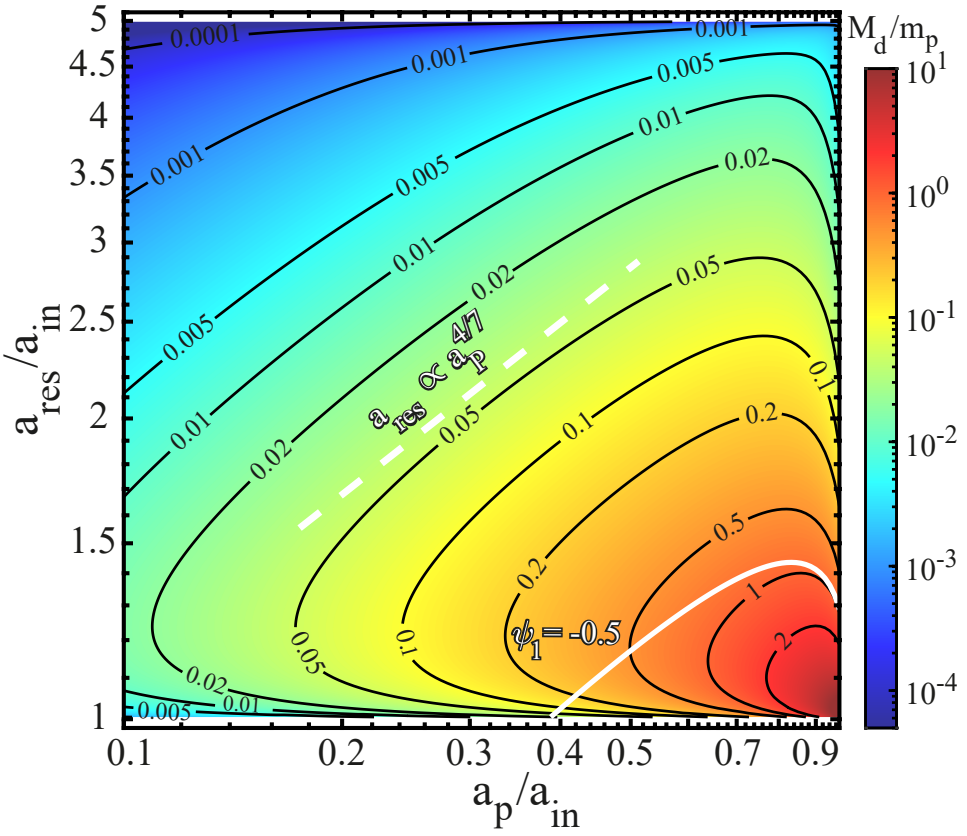

Figure 5.5. Location of secular resonances relative to the disc inner edge $a_{\rm res}/a_{\rm in}$ as functions of $a_p/a_{\rm in}$ and M_d/m_p. Calculations assume a power-law disc model with $p = 1$ and $\delta \equiv a_{\rm out}/a_{\rm in} = 5$. The solid white line represents the contour for $M_d/m_p = 1$ obtained by ignoring disc edge effects, i.e., $\psi_1 = -0.5$. The dashed white line shows the scaling of $a_{\rm res}$ with a_p for fixed M_d/m_p, Equation (5.18). See the text (Section 5.4.1) for details.

$(a_p/a_{\rm in},\ a_{\rm res}/a_{\rm in})$ plane computed using my fiducial disc model, i.e., $p = 1$ and $\delta = 5$ (Section 5.2.1).

Figure 5.5 shows that for any given planet, two or no secular resonances occur within the disc provided that $10^{-4} \lesssim M_d/m_p \lesssim 2$. Additionally, one can see that for any $a_p/a_{\rm in}$, one of the resonances always occurs in the vicinity of the disc inner edge as described in Section 5.3.2, i.e., $a_{\rm res,1} \simeq a_{\rm in}$, and its location varies weakly with M_d/m_p. On the other hand, the second resonance occurs at semimajor axis $a_{\rm res,2} \gtrsim a_{\rm res,1}$ whose location changes significantly with varying M_d/m_p. Indeed, with increasing M_d/m_p (at fixed $a_p/a_{\rm in}$), this resonance is pushed inwards from $\simeq a_{\rm out}$ toward the inner resonance at $\simeq a_{\rm in}$ until both resonances "merge," i.e., the distance between them approaches zero. Figure 5.4 provides a complementary view of this behaviour. Looking at Figure

5.5 once can also see that, for planets closer to the disc, larger M_d/m_p is necessary to maintain the resonance at a given semimajor axis.

I remind that the existence of the inner resonance is mainly due to the disc edge effects. That is, the divergence of $A_d(a) \propto -|\psi_1(a)|$ as $a \to a_{\rm in}$ allows the resonance condition of Equation (5.16) to be satisfied around $\approx a_{\rm in}$, even for relatively small values of M_d (Section 5.3). This explains why for a given $a_p/a_{\rm in}$, the resonance at $a_{\rm res,1}$ is constrained to be very close to $\simeq a_{\rm in}$ irrespective of M_d/m_p. In the absence of edge effects, this inner resonance will not exist, resulting in a single resonance for fixed system parameters rather than two. This is illustrated in Figure 5.5 for $M_d/m_p = 1$ by setting $\psi_1(a) = -0.5$ (solid white line).

The behaviour of the resonance locations can be explained analytically. Consider the approximate form of the resonance condition, Equation (5.16), in the limit of $a_p/a_{\rm in} \to 0$ so that $A_{d,p}$ is negligible and one can use the asymptotic limit of $b^{(1)}_{3/2}$, and the two terms on the left-hand side of Equation (5.17) balance each other (recall that $\psi_1 < 0$). It is then easy to demonstrate that for a resonance to occur at $a_{\rm in} \lesssim a_{\rm res} \lesssim a_{\rm out}$, the disc mass must be given by

$$\begin{aligned} \frac{M_d}{m_p} &\approx \frac{3\delta^{2-p}}{4|(2-p)\psi_1(a_{\rm res})|}\left(\frac{a_p}{a_{\rm in}}\right)^2\left(\frac{a_{\rm res}}{a_{\rm in}}\right)^{p-4}, \\ &\approx 0.15\ a^2_{p,20}\ a^{-3.5}_{\rm res,70}, \end{aligned} \tag{5.18}$$

where the numerical estimate is obtained for the adopted fiducial disc model ($p = 1$, $\delta = 5$), for which $|\psi_1(a)| \propto a^{1/2}$ when $a_{\rm in} \ll a \ll a_{\rm out}$, see Section 5.3.[4] Fixing M_d/m_p in Equation (5.18) then approximates the slopes of the contours in Figure 5.5 reasonably well – see the white dashed line. As expected, the numerical results deviate from the scaling in Equation (5.18) both as $a_{\rm res} \to a_{\rm in}$ or $a_{\rm out}$, where ψ_1 diverges, and as $a_p \to a_{\rm in}$, since $A_{d,p}$ becomes non-negligible.

5.4.2 Timescale for Eccentricity Excitation

I now consider how the eccentricity excitation timescale varies as a function of model parameters. To this end, I make use of the definition of τ given by Equation (5.15), which quantifies the time it takes for initially circular orbits to reach $e = 1$ at the resonance. I note that τ is a strong function of the resonance location, and it explicitly depends on the parameters of the planet but not the disc. This is because the

[4] In an infinitely extending disc, i.e., as $\delta \to \infty$, ψ_1 becomes independent of semimajor axis, e.g., $\psi_1(a) = -0.5$ for $p = 1$. In this case, Equation (5.18) would read as $M_d/m_p \approx 0.26\ a^2_{p,20}\ a^{-3}_{\rm res,70}$.

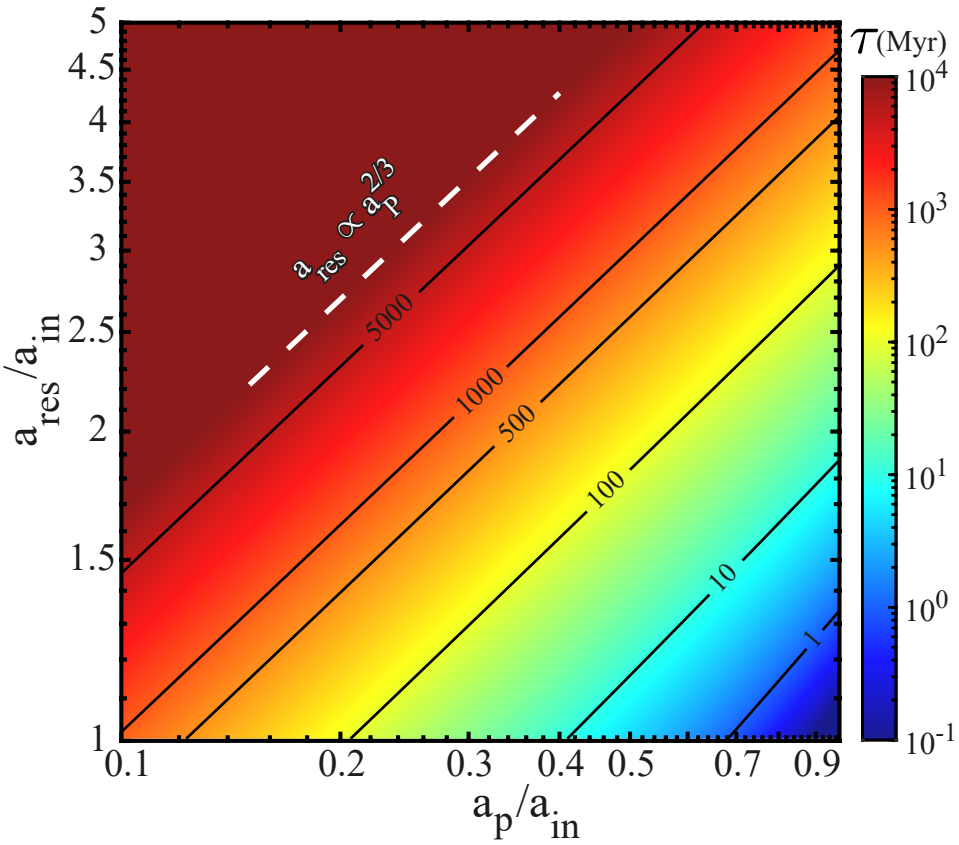

Figure 5.6. Contour plot of the timescale τ for exciting planetesimal eccentricities by the secular resonance (Equation 5.15), in the space of $a_p/a_{\rm in}$ and $a_{\rm res}/a_{\rm in}$. The calculations assume a planet with $m_p = 100 M_\oplus$ and $e_p = 0.1$ around a solar-mass star. The white dashed line shows the scaling of $a_{\rm res}$ with a_p for a fixed value of τ. See the text (Section 5.4.2) for details.

disc, assumed to be axisymmetric in my model (Section 5.2), does not contribute to eccentricity excitation.

In Figure 5.6 I plot the contours of τ in the $(a_p/a_{\rm in},\ a_{\rm res}/a_{\rm in})$ plane for a particular choice of planetary mass and eccentricity, $m_p = 100 M_\oplus$ and $e_p = 0.1$, assuming a solar-mass star. It is evident that the timescales are shorter when the planet and the resonance location are closer together, i.e., in the lower-right corner of parameter space where $a_{\rm res}/a_p \to 1$. Note that for the adopted planetary parameters, over a broad range of parameter space, the timescales range from ~ 10 Myr to a few gigayears; this is comparable to the ages of observed debris discs. Moreover, the slopes of the contours in Figure 5.6 can be explained by setting τ to a constant in Equation (5.15): this yields the scaling $a_{\rm res} \propto a_p^{2/3}$ illustrated by the white dashed line in Figure 5.6.

Finally, Equation (5.15) shows that τ is inversely proportional to both the planetary mass and eccentricity. Thus, more massive or eccentric planets exert larger torque and excite planetesimal eccentricities more quickly, shortening the timescale τ when $a_p/a_{\rm in}$ and $a_{\rm res}/a_{\rm in}$ are kept fixed. This means that in Figure 5.6 the contours of τ will be shifted to the left (right) when the product of m_p and e_p is increased (decreased).

5.4.3 Resonance Width

I now quantify the range of semimajor axes w over which resonances act to significantly excite planetesimal eccentricities. To this end, I follow[5] Yelverton & Kennedy (2018) and calculate the distance over which the forced planetesimal eccentricities $e_{\rm forced}(a)$ exceed a *constant threshold value* $\tilde{e}$. That is, I define w as the difference (in absolute values) between the two values of semimajor axis a_i ($i = 1,\ 2$) satisfying

$$\tilde{e} = |e_{\rm forced}(a_i)| = \left| \frac{-B_p(a_i)}{A(a_i) - A_{d,p}} \right| \tag{5.19}$$

in the vicinity of a given resonance. Here, I clarify that this definition serves as a proxy for the significance of a given resonance, and it does not necessarily correspond to the actual widths of gaps that I expect to observe.[6]

In Equation (5.19), the planetary and disc masses appear only through their ratio M_d/m_p, and the two relevant semimajor axes – a_i and a_p – could be expressed relative to $a_{\rm in}$; see Equations (5.4) – (5.7). Furthermore, the ratio M_d/m_p could be related to $a_p/a_{\rm in}$ and $a_{\rm res}/a_{\rm in}$ by using the condition for secular resonance; see Equations (5.16) and (5.17). Thus, I can compute the resonance width w relative to $a_{\rm in}$ as functions of $a_p/a_{\rm in}$ and $a_{\rm res}/a_{\rm in}$ only, once $\tilde{e}$ and e_p are specified (recall that $B_p \propto e_p$, Equation (5.6)).

The threshold eccentricity $\tilde{e}$ in Equation (5.19) represents an ad hoc parameter, necessitating a physical justification for a particular choice of its value. To this end, I note that the presence of a physical gap within the disc is subject to the condition that planetesimal eccentricities are larger around the resonances than elsewhere. Away from the resonances, the forced planetesimal eccentricity is maximised near the disc inner edge where, approximately, $e_{\rm forced}(a_{\rm in}) \to e_{\rm forced,p}(a_{\rm in})$, which cannot exceed e_p; see Equation (5.13) and Figure 5.4. Based on this reasoning, I adopt $\tilde{e} = e_p$ in what follows, unless stated otherwise.

[5] For an alternative method, see Levison & Agnor (2003).

[6] This is not least because the actual widths of gaps depend non-trivially on the spatial distribution of planetesimals, i.e., the profiles (and gradients) of both $e(a)$ and $\varpi(a)$ (Statler, 2001).

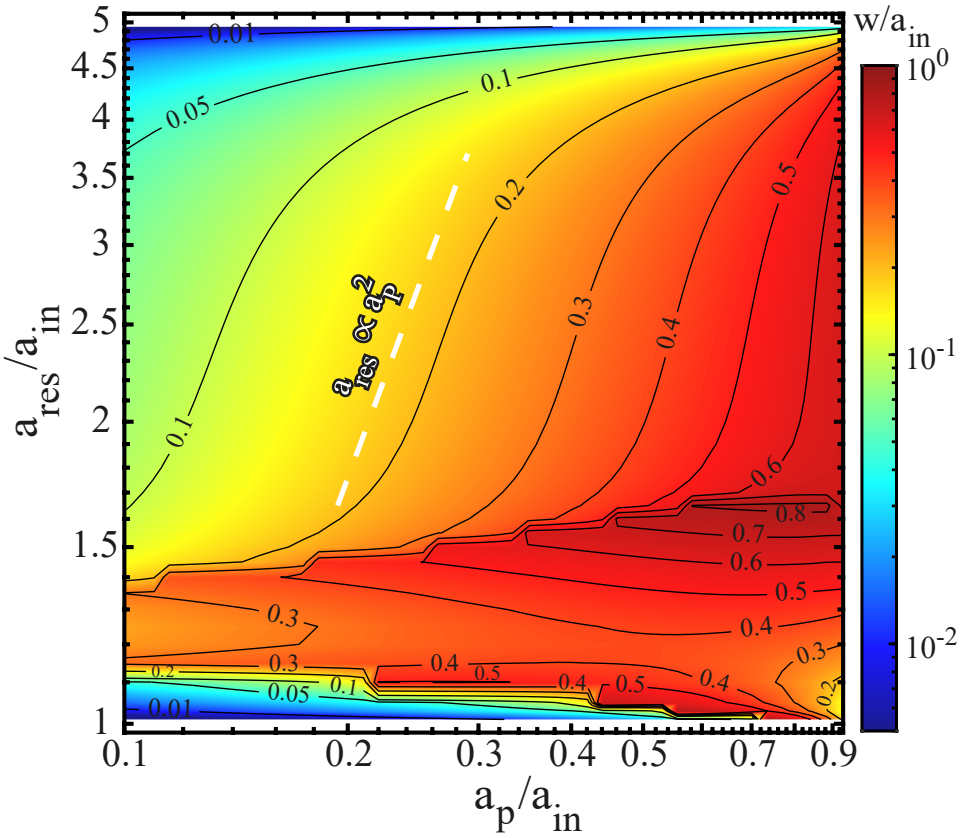

Figure 5.7. Contour plot of the resonance width w relative to $a_{\rm in}$ (Equation (5.19)) in the space of $a_p/a_{\rm in}$ and $a_{\rm res}/a_{\rm in}$, computed using $\tilde{e} = e_p$ and the same disc parameters as in Figure 5.5. The white dashed line shows the scaling of $a_{\rm res}$ with a_p for a fixed value of $w/a_{\rm in}$; Equation (5.20). See the text (Section 5.4.3) for details.

In Figure 5.7 I plot the contours of $w/a_{\rm in}$ in the $(a_p/a_{\rm in},\ a_{\rm res}/a_{\rm in})$ plane for the fiducial disc model with $p = 1$ and $\delta = 5$ (see Section 5.2.1), assuming $\tilde{e} = e_p$. Looking at Figure 5.7, one can see that increasing the planetary semimajor axis for a fixed $a_{\rm in}$ tends to generally broaden the width of a given resonance. This is, though, less obvious in the range $1.1 \lesssim a_{\rm res}/a_{\rm in} \lesssim 1.5$, as the width there is a weaker function of $a_p/a_{\rm in}$. Second (and relatedly), I find that for a given planetary semimajor axis, resonances occurring closer to the disc inner edge generally have larger widths compared to resonances farther away; see also Figure 5.4. The exception to this is if $a_{\rm res}/a_{\rm in} \simeq 1$, where the values of $w/a_{\rm in}$ are comparatively smaller, particularly in the lower-left corner of Figure 5.7.

To understand this behaviour, I remind the reader that for a given $a_p/a_{\rm in}$, my disc model with sharp edges has two resonance sites: one *always* at $a_{\rm res,1} \simeq a_{\rm in}$ and another farther away at $a_{\rm res,1} \lesssim a_{\rm res,2} \lesssim a_{\rm out}$; see Section 5.4.1. In terms of Figure 5.7,

this means that for a given $a_p/a_{\rm in}$ (and M_d/m_p, see Figure 5.5), if the resonances are well separated from each other, i.e., $a_{\rm res,1} \ll a_{\rm res,2}$, the inner resonance will be much narrower than the other. This behaviour could be understood for instance by looking at the curves in Figure 5.4 for $M_d/m_p = 10^{-3}, 10^{-2}$, or 10^{-1}, which show that the inner resonance width is insignificant.

On the other hand, for fixed $(a_p/a_{\rm in}, M_d/m_p)$, if the resonances are close to each other such that $a_{\rm res,2}/a_{\rm in} \lesssim 1.5$ and $a_{\rm res,1} \simeq a_{\rm in}$ (see Figure 5.5), the resonances "merge" together yielding relatively large values of $w/a_{\rm in}$. What I mean by "merging" here is that $e_{\rm forced}(a)$ in-between the resonances stays larger than $\tilde{e}$, and my definition of w does not disentangle the two resonances.[7] This could be understood, for instance, by looking at the curve for $M_d/m_p = 1$ in Figure 5.4. These considerations explain why the contours of constant $w/a_{\rm in}$ in Figure 5.7 behave differently for $a_{\rm res}/a_{\rm in} \lesssim 1.5$ compared to $a_{\rm res}/a_{\rm in} \gtrsim 1.5$.

To better understand the behaviour of $w/a_{\rm in}$, in Appendix F I derive an analytic expression for the resonance widths showing that, to a good approximation,

$$\frac{w}{a_{\rm in}} \approx \frac{2}{a_{\rm in}} \left| \frac{B_p(a)/\tilde{e}}{dA/da} \right|_{a_{\rm res}} \propto \frac{e_p}{\tilde{e}} \frac{a_p}{a_{\rm in}} \left(\frac{a_{\rm res}}{a_{\rm in}} \right)^{-1/2}, \tag{5.20}$$

where the scaling holds for $p = 1$ in the limits of $a_p/a_{\rm in} \to 0$ and $a_{\rm in} \ll a_{\rm res} \ll a_{\rm out}$. First, Equation (5.20) shows that the width is inversely proportional to the gradient of A at $a_{\rm res}$. This explains why resonances in proximity of the disc edges are relatively narrow: in the limit of $a_{\rm res} \to a_{\rm in}, a_{\rm out}$, one has $A \to A_d$, which diverges due to edge effects (Figure 5.2), and dA/da is very large. Second, Equation (5.20) indicates that the resonance width is directly proportional to $B_p \propto e_p$; this makes intuitive sense since e_p controls the amplitude of planetesimal eccentricities (Equation 5.12). It follows that more eccentric planets tend to produce wider resonances, provided that $\tilde{e}$ can be chosen independently from e_p (though this is not clear a priori). Third, and more importantly, the scaling of Equation (5.20) adequately explains the slopes of the $w/a_{\rm in}$ contours: setting $w/a_{\rm in}$ to a constant in Equation (5.20) yields the scaling $a_{\rm res} \propto a_p^2$, which is obvious in Figure 5.7 (see the white dashed line). Indeed, by fitting the numerical results in Figure 5.7 with the functional form of Equation (5.20), I find that the following expression

$$w \approx 15.3 \text{ au } a_{p,20}\ a_{\rm res,70}^{-1/2}\ a_{\rm in,30}^{1/2}\ (e_p/\tilde{e}) \tag{5.21}$$

[7] Adopting larger $\tilde{e}$ at fixed e_p could modify this behaviour. However, it is not clear a priori what value must be assigned to $\tilde{e}$, not least because $e_{\rm forced}(a) \propto e_p$ could stay well above unity in-between the resonances in linear Laplace–Lagrange theory.

provides an acceptable approximation of the resonance widths for my fiducial disc model (Section 5.2.1).

5.5 Example: Application to HD 107146

For a given debris disc exhibiting a depletion in its surface density, I can hypothesise that this depletion is due to eccentricity excitation by secular resonances mediated by the gravity of the disc and an unseen planet. I can then employ the characteristics of the secular resonances analysed in Section 5.4 to constrain the disc–planet parameters that could configure the secular resonances appropriately and produce a depletion similar to the observations. In this section, as an exemplary case, I apply these considerations to the HD 107146 disc and identify the "allowed" parameter space subject to observational constraints. The detailed investigation of the dynamical evolution in models chosen from the allowed parameter space is carried out in the next section (i.e., Section 5.6).

5.5.1 Constraints from Gap Location

As noted in Section 5.1.1, ALMA observations show that the HD 107146 disc, spanning from $a_{\rm in} \sim 30$ au to $a_{\rm out} \sim 150$ au, features a gap centred at $a_g \sim 70 - 80$ au (Ricci et al., 2015; Marino et al., 2018a). Thus, I must choose the disc–planet parameters such that a secular resonance occurs within the depleted region. Here I opt to fix the resonance location at $a_{\rm res} = 70$ au. The analysis in Section 5.4.1 then allows me to uniquely determine the ratio M_d/m_p as a function of $a_p/a_{\rm in}$, see also Equation (5.18). In other words, for a given disc mass, I can deduce the planetary mass and semimajor axis that configure the resonance location appropriately (or vice versa). This is displayed by the black solid lines in Figure 5.8 for various values of disc mass (in units of $M_\oplus$).

However, the disc mass cannot be arbitrarily large and must be constrained. To this end, I note that observations of HD 107146 have detected around $0.25 M_\oplus$ of dust at millimetre wavelengths (Ricci et al., 2015; Marino et al., 2018a). By extrapolating this up to planetesimals of ~ 100 km in diameter, the estimated total disc mass is $M_d \sim 100 - 300 M_\oplus$ (assuming a size distribution with an exponent of -3.5; Ricci et al. (2015); Marino et al. (2018a)). Here I choose to take $100 M_\oplus$ as the upper limit of the disc mass. Based on this, I exclude regions in the (a_p, m_p) parameter space that require more massive discs – see the grey shaded area in the upper part of Figure 5.8.

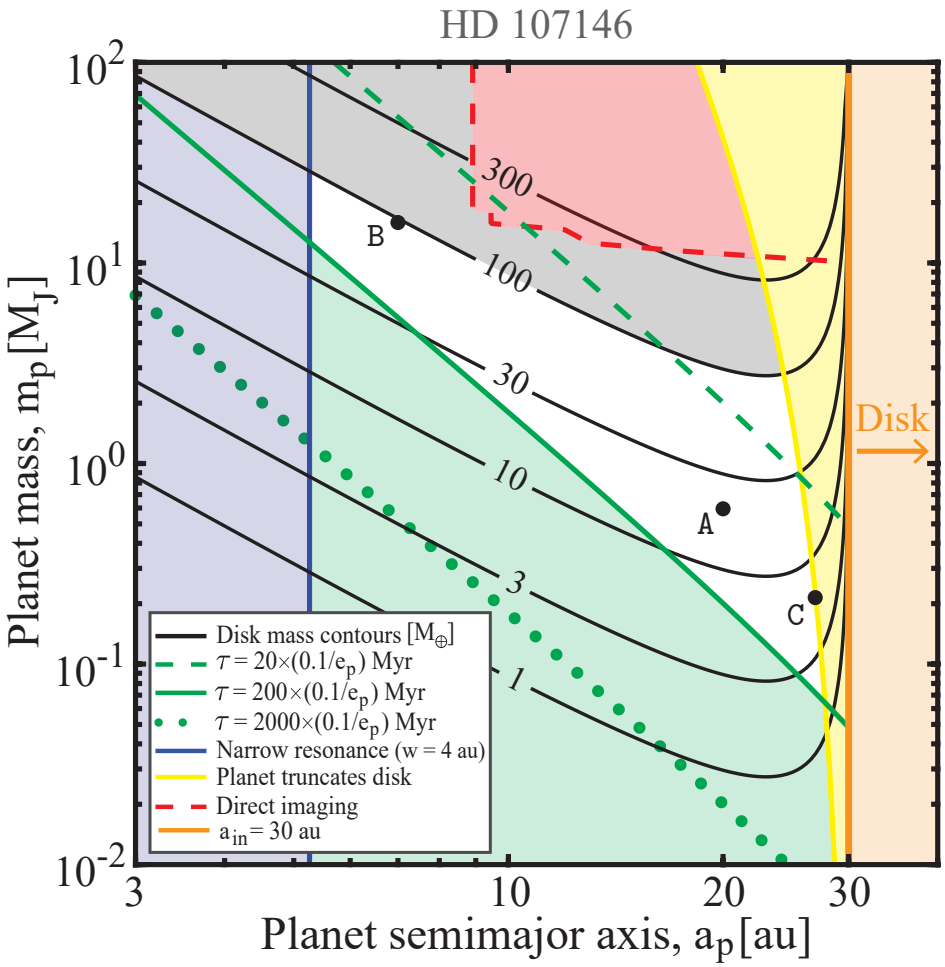

Figure 5.8. Combinations of the planet (m_p) and disc masses (M_d) as a function of planet semimajor axis a_p that are expected to produce a gap in an HD 107146–like disc at 70 au. The curves of constant M_d are shown by the black contours. The grey region is ruled out, as the disc would be too massive. The green region shows the excluded region where the eccentricity excitation timescales are much longer than the stellar age. The blue region is ruled out, as the resulting resonance width would be much narrower than the observed gap. A planet close to the disc inner edge is ruled out (yellow region) by considerations of overlapping mean motion resonances. The red region is ruled out by direct imaging. The remaining white area represents the region where the disc–planet parameters meet all of the above conditions. The lettered points represent the model parameters discussed in Sections 5.6.1 and 5.6.2.1, and are listed in Table 5.1 for reference. See the text (Section 5.5) for details.

5.5.2 Constraints from Stellar Age and Disc Asymmetry

I can further constrain the parameter space by considering the age of HD 107146, which is estimated to be $t_{\rm age} \sim 80 - 200$ Myr (Williams et al., 2004). Specifically, I require the timescale for eccentricity excitation at the resonance τ to be less than

around the age of the system, i.e., $\tau \lesssim t_{\rm age}$. From Section 5.4.2, however, I know that τ depends not only on the planet's mass and semimajor axis but also on its eccentricity; see Equation (5.15). To this end, I note that ALMA observations have found that the HD 107146 disc is roughly axisymmetric, with a 2σ upper limit of ~ 0.03 for the global disc eccentricity (Marino et al., 2018a). This suggests that the invoked planet must be of relatively low eccentricity. Thus, in what follows, I limit myself to $e_p \leq 0.1$.

The green curves in Figure 5.8 show contours along which the excitation timescale τ is 20, 200, and 2000 Myr (dashed, solid, and dotted lines, respectively) at $a_{\rm res} = 70$ au. The calculations assume $e_p = 0.1$ – the maximum value of e_p that I consider in my subsequent calculations – and use the stellar mass of HD 107146, namely $M_c = 1.09 M_\odot$ (Watson et al., 2011). I first note that by definition, $\tau \propto 1/e_p$ (Equation (5.15)); thus, for less eccentric planets the contours shown in Figure 5.8 will correspond to longer timescales. Second, recall that τ is a measure of the time within which initially circular planetesimal orbits become radial, $e \to 1$ (Section 5.4.2). Thus, even if $\tau \gtrsim t_{\rm age}$ for a given planet (such that $e(t_{\rm age}) \lesssim 1$), one might still expect sufficient eccentricity excitation for depletion to be apparent at the resonance within the stellar lifetime. Given these considerations and the uncertainty on the age of the system, I exclude the region in (a_p, m_p) parameter space corresponding to $\tau > 200\,(0.1/e_p)$ Myr. This is illustrated by the green shaded region in Figure 5.8.

5.5.3 Constraints from Gap Width

As noted in Section 5.1.1, the gap width in the HD 107146 disc is estimated to be $w_{\rm obs} \approx 40$ au (Marino et al., 2018a). Given this, the planet's semimajor axis could, in principle, be constrained by using the analysis of resonance widths w in Section 5.4.3 (recall that $w \propto a_p$, Equation 5.21). However, I remind the reader that the resonance widths as defined in Section 5.4.3 do not necessarily correspond to the physical width of gaps that I expect to form. Nevertheless, I could still use the definition of w to rule out the range of planetary semimajor axes for which the resonance widths would be negligible, i.e, $w/w_{\rm obs} \ll 1$. Here I consider resonance widths to be negligible if $w/w_{\rm obs} \leq 0.1$ (this choice is somewhat arbitrary). The blue solid line in Figure 5.8 corresponds to $w/w_{\rm obs} = 0.1$; planetary semimajor axes to the left of this line are ruled out (blue shaded region).

5.5.4 Considerations of Mean-motion Resonances

Finally, I note that the planet cannot be arbitrarily close to the disc. This is because the planetary orbit is surrounded by an annular "chaotic zone" wherein particles will be quickly ejected from the system due to overlapping first-order mean motion resonances (MMRs); see Section 3.2.4.2. Moreover, the secular approximation of Section 5.2 would break down within this zone (e.g. Murray & Dermott, 1999). The half-width Δa_p of the chaotic zone on either side of the planetary orbit depends on the planet's mass (Wisdom, 1980; Duncan et al., 1989) such that, to lowest order[8], it is given by Equation (3.35). I thereby can rule out the region in the (a_p, m_p) parameter space wherein the planet's chaotic zone would lie within the disc, i.e., $a_p + \Delta a_p > a_{\rm in}$. This is illustrated by the yellow shaded region near the right boundary of Figure 5.8. Planetary parameters lying along the yellow solid line correspond to $a_p + \Delta a_p = a_{\rm in}$; thus, they could be responsible for setting the inner disc edge (e.g. Quillen, 2006) at $a_{\rm in} = 30$ au (orange line).

* * *

I have now identified the "allowed" range of disc–planet parameters that can produce an HD 107146–like disc structure. This is represented by the white (unshaded) region in Figure 5.8, and roughly defined by a_p in the range $\sim 5 - 27$ au, m_p between ~ 0.1 and $25 M_J$, and $3 \lesssim M_d/M_\oplus \lesssim 100$. Note that the allowed combinations of m_p and a_p are consistent with the limits placed by direct imaging of HD 107146 (Apai et al., 2008); see the dashed red curve in Figure 5.8. For reference, the combinations of m_p, a_p and M_d that I consider later in this chapter are labelled as `Models A`, `B`, and `C` in Figure 5.8; see also Table 5.1. Note that each of these configurations correspond to $\tau \approx 135 \times (0.05/e_p)$Myr, and model `A` represents the fiducial configuration considered next in Section 5.6.1.

I remark that in the above discussion, I have implicitly ignored the occurrence of an *inner* secular resonance at $\simeq a_{\rm in}$; apart from the one already fixed at $a_{\rm res} = 70$ au in Figure 5.8; see Section 5.4.1. This can be justified on the grounds that the inner resonance is of very narrow width except if the two resonances are close to each other, which is not the case here (Section 5.4.3). As a result, and as I will show next, the inner resonance is irrelevant and does not have any observable effect.

[8]Strictly speaking, Equation (3.35) is valid for circular orbits in the absence of collisions. The chaotic zone is known to broaden with both increasing eccentricity (Mustill & Wyatt, 2012) and due to collisional effects (Nesvold & Kuchner, 2015a). For simplicity, I have ignored these effects.

Table 5.1. Parameters of the Disc–Planet Systems Considered in Section 5.6.

Model	$M_d(M_\oplus)$	$m_p(M_J)$	$a_p(au)$	M_d/m_p	e_p	$\varpi_p(0)$	$\tau_{\rm sec}(Myr)$
`A`	20	0.6	20	1.05×10^{-1}	0.05	0	33
`A-Loep`	...	...	...	...	0.025	...	...
`A-Hiep`	...	...	....	...	0.1	...	...
`B`	95	15.8	7	1.89×10^{-2}	0.05	...	56
`C`	6	0.2	26.93	9.44×10^{-2}	...	...	26

Note. The combinations of M_d, m_p, and a_p (Columns 2–4) chosen from the allowed region in Figure 5.8. Column 5 presents the disc-to-planet mass ratio. Columns 6–7 present the planet's eccentricity and initial apsidal angle, whose precession period is given in Column 8. (a) `Model A` is the fiducial configuration adopted in this chapter. (b) Each of the considered models have $\tau \approx 135 \times (0.05/e_p)$ Myr.

Finally, I point out that Equations (5.15), (5.18), and (5.21), combined with Equation (3.35), can be applied to generate an approximate version of Figure 5.8 for any other observed debris disc with a gap.

5.6 Evolution of the Disc Morphology

In the previous section, I identified the combinations of the "allowed" disc–planet parameters that could reproduce the observed depletion in the HD 107146 disc; see Figure 5.8. I now investigate the dynamical evolution of disc–planet systems using some of these parameters. My specific aims here are two-fold: to illustrate how secular resonances sculpt depleted regions, and to analyse more fully the disc and gap morphology in the course of secular evolution.

5.6.1 A Fiducial Configuration

I begin by presenting results showing the evolution of the disc surface density in the fiducial configuration, i.e., model `A` (see Table 5.1). I remind the reader that model `A` is the configuration that was considered in Section 5.3, where I discussed the temporal evolution of planetesimal eccentricities and apsidal angles as a function of semimajor axis – see Figure 5.3. To this end, I convert the orbital element distributions of planetesimals shown in Figure 5.3 – which, I remind, were determined *analytically*

using Equations (5.10) and (5.11) – into surface density distributions. Technical details about this procedure can be found in Appendix G, and may be skipped by the reader at first reading. However, to avoid confusion, I remark that the results presented in this and in subsequent sections are obtained by the analytical model described in Section 5.2 and not by direct N-body simulations, which is beyond the scope of this chapter.

The resulting maps of the (normalised) disc surface density Σ at times corresponding to those in Figure 5.3 are shown in Figure 5.9. For reference, in this figure I also show the planet's orbit and its pericentre position, which precesses with a period of $\tau_{\rm sec} \equiv 2\pi/A_{d,p} \approx 33$ Myr (Equation (5.7)). To facilitate the interpretation of my results, in Figure 5.10 I also show the profiles of the azimuthally averaged disc surface density $\langle\Sigma\rangle$ as a function of radial distance r at the same times as in Figure 5.9. Below I provide a detailed description of the different evolutionary stages that I identified.

Stage 1 ($0 \leq t \lesssim \tau_{\rm sec}$): At early times, the disc quickly evolves away from its initial axisymmetric state by developing a trailing spiral structure (see Figures 5.9(a), (b)). This spiral structure initially starts off at the inner disc edge and propagates radially outwards with time as it wraps around the star; see also the animated version of Figure 5.9. For instance, by 1 Myr at least two windings are noticeable (Figure 5.9(a)), with the outermost prominent spiral arm occurring at ~ 40 au. This arm moves out to ~ 60 au by 10 Myr (Figure 5.9(b)). A complementary view of this behaviour is provided by Figures 5.10(a),(b).

I note that the outermost portion of the spiral is associated with planetesimal orbits that have attained their maximum eccentricity, i.e., have completed half a precession period – see Figure 5.3. Interior to this, the spirals become difficult to discern, since planetesimals in this region have completed more than one precession period and their orbits are phase-mixed; i.e., $\Delta\varpi(a)$ spans the range $[-\pi/2, \pi/2]$ – see Figures 5.3(a), (b). As a result, the surface density distribution interior to the outermost spiral looks roughly axisymmetric; see, e.g., panel (b) of Figure 5.9. I also find that the spiral propagates outwards at a slower rate as it extends to larger radii; see panels (a)–(c) of Figure 5.9 and its animated version. This follows from the fact that the planetesimal precession rate is a decreasing function of the semimajor axis; see Figure 5.2.

I remark that the behaviour described thus far shows some parallels with the findings of Wyatt (2005b), who showed that an eccentric planet launches a spiral wave that propagates throughout a *massless* disc. The main difference is that, in my setup, the spiral wave extends out to only about a radius of 70 au and not to the outer disc edge (as would happen in a massless disc); see Figure 5.9. This is to be expected, since

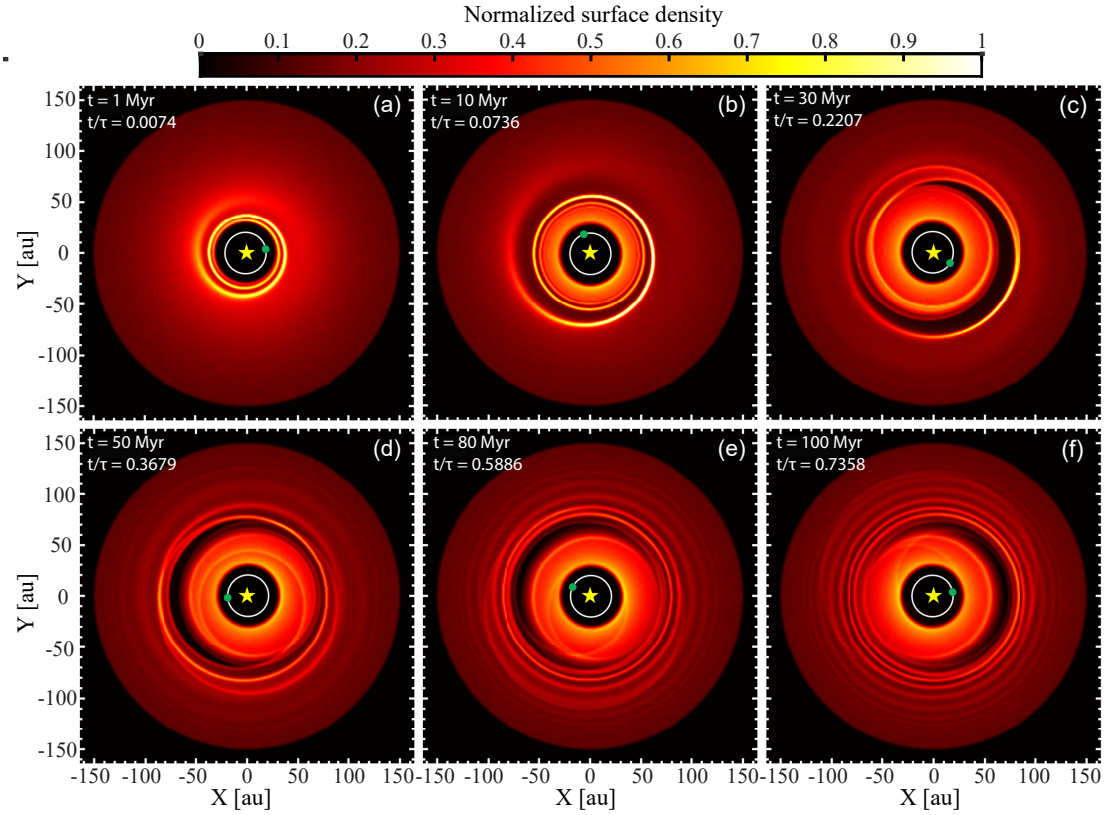

Figure 5.9. Series of two-dimensional snapshots showing the evolution of the (normalised) disc surface density Σ in the fiducial model (`Model A`, Table 5.1), as derived from the analytically computed dynamical state of planetesimals shown in Figure 5.3. The snapshots correspond to the same moments of time t as in Figure 5.3, and are indicated in each panel for reference. The time is also indicated relative to $\tau \approx 135$ Myr, Equation (5.15). All panels have 400×400 pixels and share the same surface density scale (and normalization constant) as shown in the colour bar. In each panel, the stellar position is marked by the yellow star, while the planet's orbit and its pericentre position are shown by the white solid line and green circle, respectively. To enhance the resolution of the images, the orbit of each planetesimal ($N = 5000$ in number) has been populated with 10^4 particles with the same orbital elements but with randomly distributed mean anomalies (see Appendix G). Note the launching of a trailing spiral wave at the inner disc edge $a_{\rm in}$ at early times (panel (a)), its subsequent wrapping around the star (panel (b)), and the formation of a crescent-shaped gap around the secular resonance at $a_{\rm res} = 70$ au (panels (c)–(f)). One can also see a wound spiral pattern beyond the gap at late times (panels (d)–(f)). It is also clear that no gap forms around the secular resonance at $\sim a_{\rm in}$. See the text (Section 5.6.1) for more details. An animated version of this figure has been made available online.

in my model, planetesimal dynamics is dominated by the planet only within ≈ 70 au, beyond which the disc gravity becomes important – see Figure 5.2 and Section 5.3.

Stage 2 ($t \sim \tau_{\rm sec}$): By the time the planet has nearly completed its first precession cycle, the disc develops a clear depletion in its surface density, which effectively splits the disc into an internal and an external part (Figures 5.9(c), 5.10(c)). The depletion occurs around the location of the secular resonance, i.e., at $a_{\rm res} = 70$ au, where the system was designed to emplace one – see Section 5.5. The appearance of the gap is evidently correlated with the excitation of planetesimal eccentricities at and around $a_{\rm res}$, where $e = t/\tau \approx 0.22$ by 30 Myr (Figure 5.3(c)).

An interesting feature of the gap is that it is of a crescent shape, which points in the direction of the planet's pericentre (Figure 5.9(c)). In other words, the gap is asymmetric in the azimuthal direction such that it is wider and deeper toward the planetary pericentre. This asymmetry is associated with the inner and outer disc components being offset relative to the star in opposite directions (Figure 5.9(c)). Indeed, the inner part forms an eccentric structure that is apsidally aligned with the planet while the outer part is anti-aligned (see also Section 5.3) – the latter though is difficult to discern in Figure 5.9 due to the smaller eccentricities in the outer parts (Figure 5.3). Nevertheless, by simply looking at the azimuthally averaged density profile, one can see that the gap has a radial width of ~ 20 au (measured relative to the initial density profile, Figure 5.10(c)). Looking at Figure 5.10(c), it is also clear that this region is not depleted fully but only partially – by about a factor of two relative to the initial density distribution.

Finally, I note that the gap is surrounded by narrow overdense regions, with the one just exterior to the gap being sharper than that interior to it (see Figures 5.9(c), 5.10(c)). These over-densities correspond to the apocentric positions of planetesimals with semimajor axes in the depleted region. The contrast between the sharpness of the over-densities is mainly due to the apsidal angles of planetesimals at $a \lesssim a_{\rm res}$ being more phase-mixed than at $a \gtrsim a_{\rm res}$ (Figure 5.3(c)). This also justifies why these sharp over-densities are transients: they taper with time as planetesimal orbits around the resonance are perturbed further (see panels (d)–(f) in Figures 5.3, 5.9, and 5.10).

Stage 3 ($\tau_{\rm sec} \lesssim t \lesssim \tau$): Further into the evolution, the structure of the gap practically remains invariant without being significantly affected by the continued growth of eccentricity around $a_{\rm res} = 70$ au (see panels (d)–(f) in Figures 5.3 and 5.9). Indeed, the gap maintains its crescent shape along with its alignment with the planet as it co-precesses with the planet's apsidal line.

At the same time, since the inner component of the disc precesses much faster than the outer component (Figure 5.2), the degree of offset between them varies as the system evolves. This causes the gap width w_g to fluctuate in time, see, e.g., Figures

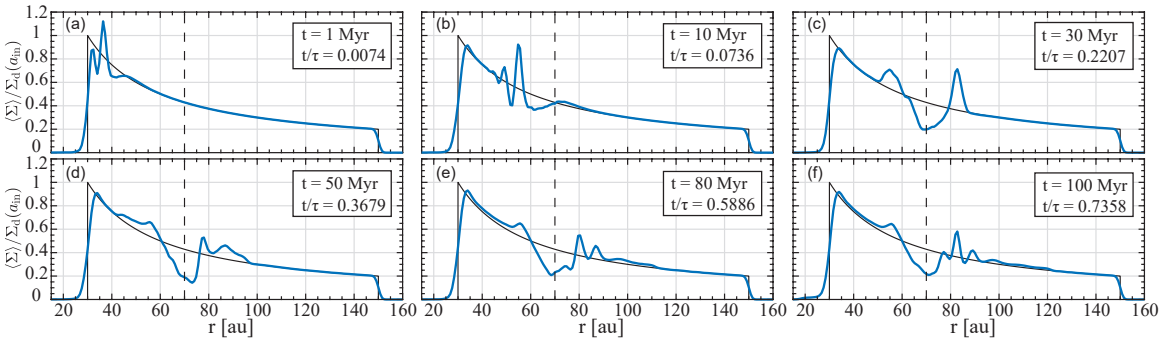

Figure 5.10. The azimuthally averaged surface density of the disc $\langle\Sigma\rangle$ as a function of radial distance r from the star (solid blue lines). Each panel corresponds to a snapshot of the fiducial configuration (Model `A`, Table 5.1) shown in Figure 5.9. The time t of each snapshot is marked in each panel, which is also shown relative to $\tau \approx 135$ Myr for reference. The results are obtained by splitting the disc into 200 annular bins (Appendix G), and are all normalised with respect to the initial analytic surface density $\Sigma_d(a)$ (Equation 5.1 with $p = 1$) at the inner disc edge, $a = a_{\rm in}$. For reference, the normalised profile of the initial $\Sigma_d(a)$ is shown in each panel with the solid black lines. At early times (panels (a), (b)), the overall shape of $\langle\Sigma\rangle$ is similar to the initial profile, but with some peak features around ~ 40 au at 1 Myr and ~ 60 au at 10 Myr, respectively. At all times after 30 Myr (panels (c)–(f)), a clear depletion in the surface density is evident around the location of the secular resonance ($a_{\rm res} = 70$ au, dashed vertical lines). One can see that the width and the depth of the depletion are effectively constant in time (panels (c)–(f)). Note also the peak structure in the density just exterior to the depletion in panels (c)–(f). See the text (Section 5.6.1) for more details. An animated version of this figure has been made available online.

5.10(d)–(f), with a time-averaged value of $w_g \approx 18.13 \pm 1.04$ au. Looking at Figures 5.10(d)–(f), it is also clear that the gap depth remains roughly constant such that, in a time-averaged sense, about $50\% \pm 3\%$ of the initial density is depleted at the resonance.

Note that, at this stage, i.e., at $t \gtrsim \tau_{\rm sec}$, at least one secular period has elapsed for planetesimals interior to the depletion, causing them to settle into a lopsided, precessing coherent structure (Figures 5.9(d)–(f)). It is also noticeable that this structure reveals little or no evidence for surface density asymmetry between its apocentre and pericentre directions, as would have otherwise been the case if the disc were *massless* (i.e., pericentre or apocentre glow; see Wyatt et al., 1999; Wyatt, 2005b; Pan et al., 2016). This can be understood by noting that in this region, although planetesimal dynamics is dominated by the planet, the disc gravity renders the forced eccentricity to be more of a constant with semimajor axis rather than scaling as $1/a$ (see Figures 5.2 and 5.3). This hinders the occurrence of a pericentre or apocentre glow (for a more detailed discussion, see Section 2.4 in Wyatt (2005b)).

On the other hand, planetesimal orbits exterior to the depletion have not yet had the time to be randomly populated in phase (Figure 5.3). Hence, a spiral pattern develops in this region as planetesimals undergo eccentricity oscillations. The spirals appear to wrap almost entirely around the star, and these are more noticeable closer to the depletion than to the outer disc edge (Figures 5.9(d)–(f)). This can also be seen in Figures 5.10(d)–(f) as a series of narrow peaks in the radial profile of $\langle\Sigma\rangle$. This behaviour can be understood by noting that planetesimals closer to the outer disc edge have smaller eccentricities (e.g., Figure 5.3) and that their orbits are quickly phase-mixed as a result of their rapid orbital precession due to disc edge effects, particularly at $a \gtrsim 130$ au (e.g., Figure 5.2, Section 5.2.2). Relatedly, if I were to evolve the system for a longer period of time, planetesimals exterior to the depletion would become phase-mixed, and the spiral structure would fade away. I note that, depending on the resolution of the observations, the spirals in this region may or may not be visible.

Before moving on, I note that already by 1 Myr into the evolution, planetesimal eccentricities around the inner secular resonance (i.e., $a_{\rm res} \approx a_{\rm in}$) are excited to ≈ 1; see, e.g., Figure 5.3(a). Evidently, however, this occurs over such a narrow radial range that it does not lead to the emergence of a gap (see Figures 5.9 and 5.10), in agreement with my expectations from Section 5.4.3. This also justifies my assertion in Section 5.5 to ignore the occurrence of an inner secular resonance for the purposes of Figure 5.8.

5.6.2 Parameter Variation

I now analyse the variation of the disc morphology associated with varying the disc–planet parameters relative to the fiducial values (`Model A`). Below is a detailed account of what I learned, supplemented by an appropriate figure, when called for.

5.6.2.1 Variation of the Planetary Semimajor Axis a_p

I first consider the effects of varying the planetary semimajor axis a_p which, I remind the reader, all else being kept the same, is equivalent to changing the ratio M_d/m_p (Sections 5.4.1 and 5.5). For ease of comparison, I choose the combinations of a_p, m_p, and M_d from Figure 5.8 such that they yield the same eccentricity excitation timescale at the secular resonance τ as in model `A`. The parameters of the chosen models, which I label as `B` and `C`, are listed in Table 5.1 and are marked on Figure 5.8. Note that the planet in `Model C` could be responsible for truncating the disc at $a_{\rm in} = 30$ au; see Section 5.5.4.

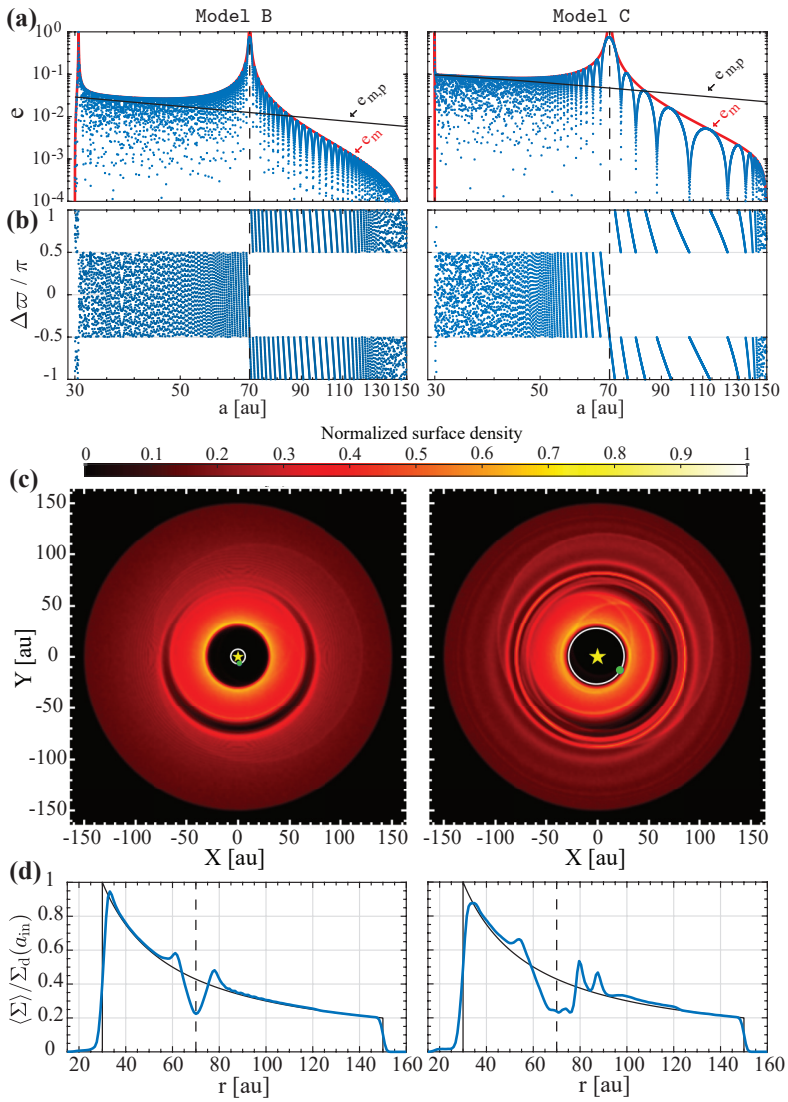

Figure 5.11. Summary of results for `Model B` (small a_p, left column) and `Model C` (large a_p, right column), see Table 5.1. The results are shown after 100 Myr of evolution, corresponding to $t/\tau \approx 0.74$ for both models. Rows (a) and (b) show the planetesimal eccentricities and apsidal angles (relative to that of the planet) as a function of semimajor axis, respectively, which are determined analytically using Equations (5.10) and (5.11). The corresponding snapshots of the disc surface density and radial profiles of the azimuthally averaged surface density are shown in the rows (c) and (d), respectively – see Appendix G for details. All other notations are the same as in Figures 5.3, 5.9, and 5.10. One can see that wider gaps are carved around the secular resonance at $a_{res} = 70$ au when the planet is closer to the disc inner edge than to the star. It is also evident that the resultant gaps are asymmetric and of approximately the same depth in both models. See the text (Section 5.6.2.1) for more details.

Generally, I find that the evolution of the disc morphology in each of models `B` and `C` proceeds in a similar manner as in the fiducial model (i.e., stages 1–3 in Section 5.6.1).

Indeed, I observe the same qualitative behaviour: the launching of a spiral arm at $a_{\rm in}$ and its outward propagation in time, the sculpting of a crescent-shaped gap around $a_{\rm res} = 70$ au by $\sim \tau_{\rm sec}$, the development of a spiral pattern exterior to the depletion at $t \gtrsim \tau_{\rm sec}$ and its subsequent potential disappearance at late times (depending on the period of secular precession at $a \gtrsim a_{\rm res}$).

Figure 5.11 summarises the snapshots of models `B` and `C` at 100 Myr (i.e., $t/\tau \approx 0.74$) into their evolution. A comparison of the results shown in this figure with those of `Model A` (Figures 5.9(f), 5.10(f)) indicate that the only obvious difference is in terms of the radial width of the gaps w_g. Indeed, the gap is radially narrower when the planet is closer to the star than to the inner disc edge: for $a_p = 7$ au (i.e., `Model B`), on time-average, $w_g \approx 11.32 \pm 0.05$ au, while for $a_p = 26.93$ au (i.e., `Model C`), I find that $w_g \approx 20 \pm 2$ au. Note that this dependence is in qualitative agreement with my expectation from Section 5.4.3 regarding the resonance widths; see Equation (5.20). Finally, I note that the gap depth is not affected by variations in planetary semimajor axis: on average, about a half of the initial density is depleted around the secular resonance regardless of a_p.

5.6.2.2 Variation of the Planetary Eccentricity e_p

The models presented thus far assumed the same planetary eccentricity of $e_p = 0.05$. To examine its effect on the disc morphology, I considered the evolution in otherwise identical setups but differing in the value of e_p by a factor of two from model `A`. These are referred to as models `A-Loep` (with $e_p = 0.025$) and `A-Hiep` (with $e_p = 0.1$) in Table 5.1.

Once again, I found that the evolution of the disc morphology qualitatively follows the same stages outlined in Section 5.6.1, but on a shorter timescale when the planet is more eccentric (recall that $\tau \propto 1/e_p$, Equation (5.15)). Additionally, I identified subtle differences in the structure of the spiral arms with increasing e_p. First, the spiral initially launched at $a_{\rm in}$ by the planet became more open for larger e_p – in agreement with the results of Wyatt (2005b). Second, and relatedly, the spirals beyond the gap became more prominent with increasing e_p due to the higher forced eccentricities in that region.

More importantly, however, I found that more eccentric planets give rise to wider gaps – in qualitative agreement with my expectations from Section 5.4.3; see Equation (5.20). Indeed, on time-average, I find that $w_g \approx 12.8 \pm 0.2$ au when $e_p = 0.025$, and $w_g \approx 24.6 \pm 2.8$ au when $e_p = 0.10$. This can be seen in Figure 5.12, where I summarise the results for models `A-Loep` and `A-Hiep`. Note that, for ease of comparison, the

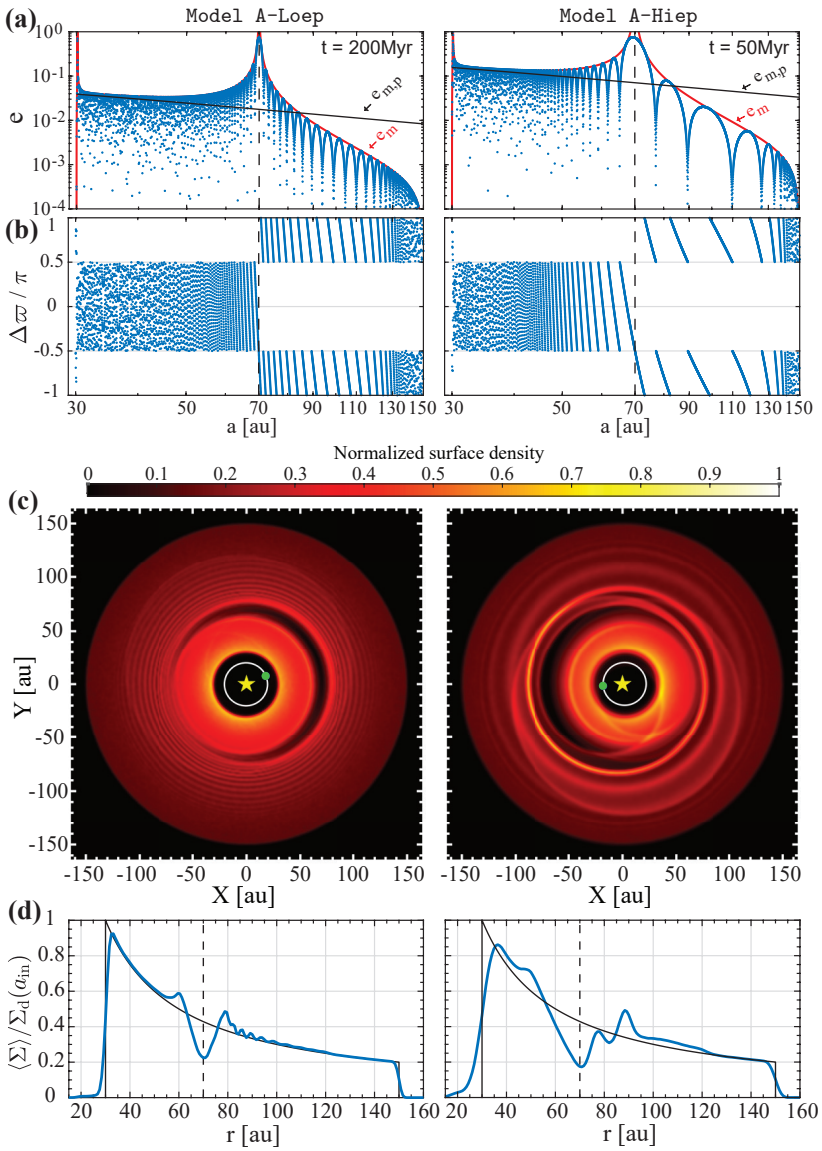

Figure 5.12. Similar to Figure 5.11, but for models `A-Loep` (left panels) and `A-Hiep` (right panels); see Table 5.1. Models `A-Loep` and `A-Hiep` are identical to the fiducial model `A`, except that they are initiated with planets with eccentricities that are lower and higher by a factor of two than in model `A` (i.e., e_p of 0.025 and 0.10), respectively. For ease of comparison, results for each model are shown at different times (as indicated in the top panels) such that they both correspond to $t/\tau \approx 0.74$. One can see that increasing e_p leads to a wider gap around the secular resonance at $a_{\rm res} = 70$ au, without significantly affecting the asymmetric shape of the gap and its depth. See the text (Section 5.6.2.2) for more details.

results are shown at different times such that $t/\tau(e_p) \approx 0.74$ for both models – the results must be compared with those of model `A` at 100 Myr (Figures 5.9 and 5.10). Looking at Figure 5.12, it is also evident that variations in e_p do not significantly affect

the fractional depth of the gap. Note also that, while planets with lower e_p reduce the offset of the inner disc component, the gap retains its non-axisymmetric feature. This is largely related to the fact that, for narrower gaps, a smaller offset suffices for the inner component to occupy about the same fraction of the gap.

5.6.2.3 Variations with Disc and Planet Masses

I now discuss the effects of varying the disc and planet masses while keeping other parameters unchanged. To begin with, I first remind the reader that this requires varying both M_d and m_p simultaneously, i.e., while keeping M_d/m_p constant, to ensure that the secular resonance location where a gap is expected to form remains the same (i.e., $a_{\rm res} = 70$ au); see Sections 5.4.1 and 5.5.1. In Figure 5.8, this is equivalent to moving vertically up or down relative to any of the simulation setups I have considered thus far.

As we know from Section 5.2, the secular precession rates scale linearly with masses (Equations (5.4) – (5.7)), whereas the forced eccentricities depend only on the ratio M_d/m_p (Equation (5.12)). Thus, varying the disc and planet masses (while $M_d/m_p = \text{const}$) should only change the secular evolution timescale, but not the details of the secular dynamics. This simply is a restatement of the fact that scaling both M_d and m_p does not affect the relative strength of perturbations due to the disc and the planet. Consequently, if one increases both the disc and planet masses in any of my simulations, then the very same dynamical end-states – hence, disc morphology – will be achieved within shorter timescales, and vice versa. Note that, in principle, this scaling rule applies as long as $M_d, m_p \ll M_c$, since otherwise the Laplace–Lagrange description in Section 5.2 becomes unreliable (Murray & Dermott, 1999). However, looking at Figure 5.8, one can see that this limitation is not a concern in my case since the most massive "allowed" planet has $m_p \sim 10^{-2} M_c$.

5.6.2.4 Variations with the Mass distribution in the Disc

My calculations so far have assumed a disc with density profile $\Sigma_d \propto 1/a$, i.e., with a power-law index of $p = 1$ in Equation (5.1). I now discuss how my results would change for different values of p, when all else is kept the same. Since the slope of the surface density p effectively controls the precession rate of both the planetesimals and the planet (Equations (5.5) and (5.7)), it is natural to expect that the location of the secular resonance will shift as the mass distribution in the disc is varied; see also Equation (5.17). I found that this is indeed the case, and I further confirmed that it does not qualitatively affect the three evolutionary stages presented in Section 5.6.1.

I generally find that when $a_{\rm in} \ll a_{\rm res} \ll a_{\rm out}$, the resonance location shifts at most by only about 10 percent as p is varied between 0.5 and 1.5. However, the direction in which the resonance shifts in a given setup is rather subtle to characterise for the following reasons. First, larger values of p lead to larger $A_{d,p}$ (and *vice versa*) as now more mass will be concentrated in the inner disc parts than in the outer regions, causing the planet to precess at a faster rate. Second – and relatedly – the disc-induced precession rate of planetesimals A_d at $a \gg a_{\rm in}$ decreases in absolute magnitude, since it is proportional to the local surface density of the disc (Equation (5.5)).[9] To summarise, varying p has opposite effects on $A_{d,p}$ and $|A_d|$, and it is the detailed balance between these two effects that determines whether the resonance shifts outwards or inwards in a given setup; see Equation (5.16). For the parameters of HD 107146 in Figure 5.8, I find that the resonance shifts inwards from its nominal location, i.e., $a_{\rm res} = 70$ au, when a larger value for p is adopted (and vice versa). Thus if I were to generate a version of Figure 5.8 with, e.g., $p = 1.5$ rather than $p = 1$, the values of M_d required to reinstate the resonance at $a_{\rm res} = 70$ au would be a factor of ~ 1.1 lower.

5.7 Discussion

The results of previous sections show that the secular interaction between a low-eccentricity planet and an external, coplanar debris disc can lead to the formation of a gap in the disc. This occurs through the excitation of planetesimal eccentricities at around one of the two secular resonances arising due to the combined gravitational influence of the disc[10] and the planet. The novelty of this mechanism is that it requires the presence of only a single planet interior to a less-massive disc, and is also robust, in the sense that it operates over a wide range of parameters.

As an example, I applied my model to the HD 107146 disc and investigated the general features of the disc and gap morphology in the course of secular evolution. In the following, I first discuss (in a general context) how the results of my model compare with the observed features in HD 107146 (Section 5.7.1). I next discuss the application of my model to other systems which are known to harbour double-ringed debris discs (Section 5.7.2). I also discuss the potential applicability of my model to planetary systems composed of two (or more) planets (Section 5.7.3). Finally, I discuss

[9] I remind the reader that $A_d(a)$ depends also on p through the coefficient ψ_1; however, the latter changes by less than a factor of two within the range $0.5 \leq p \leq 1.5$ (e.g. Figure 4.5; Silsbee & Rafikov, 2015a).

[10] Recall that in this chapter I ignore the non-axisymmetric component of the disc gravity. See Section 5.8.2 for further discussion of this point.

the implications of my results for determining the masses of debris discs (Section 5.7.4), and for their dynamical modelling in general (Section 5.7.5).

5.7.1 Comparison with the Observed Structure in HD 107146

By applying my model to HD 107146, I have shown that a gap can be readily sculpted at the observed location, i.e., around 70 au (Marino et al., 2018a), for a wide range of planet–disc parameters; see, e.g., Figure 5.8 and Section 5.6. Additionally, my results show that the produced gaps invariably have a fractional depth of about 0.5 (Section 5.6), which is consistent with that observed in HD 107146 (Marino et al., 2018a). While these results are encouraging, there are some issues with my model that need to be highlighted when it comes to comparing with the observational data of HD 107146 (Marino et al., 2018a).

First, as already mentioned in Section 5.5.2, ALMA observations of HD 107146 indicate that its disc is axisymmetric and characterised by a circular gap (Marino et al., 2018a); see e.g. Figure 2.6. My model, however, produces gaps that are asymmetric in the azimuthal direction (Section 5.6), with the disc surface density being depleted to a greater extent and over a wider region in the direction of planet's pericentre. I further found that the gap asymmetry cannot be mitigated, as one might naively expect, by adopting lower values for the planetary eccentricity – see Section 5.6.2.2.

Second, as already stated in Section 5.5.3, the observed gap in HD 107146 is ~ 40 au wide. This is larger by about a factor of two compared to the gap in my fiducial configuration (Section 5.6.1). In principle, my model can yield such wide gaps with a combination of high-eccentricity and large semimajor axis for the planetary orbit; see Sections 5.6.2.1 and 5.6.2.2. However, this would also impose more notable non-axisymmetric structure on the disc, which, given the discussion above, is problematic for HD 107146. Thus the conclusion is that, within the model adopted in this chapter, it is difficult to sculpt a gap as wide and as axisymmetric as that in HD 107146 without invoking additional processes.

Third, observations of HD 107146 indicate that the surface brightness of the outer and inner rings are comparable (Marino et al., 2018a); see Figure 2.7. Since submillimetre dust emission at a distance r scales as $T(r) \propto r^{-1/2}$ (assuming blackbody emission in the Rayleigh-Jeans limit), this observation suggests an *increasing* surface density with radius, which may seem unnatural in the context of protoplanetary discs. As a result, this has been taken as evidence for collisional depletion of planetesimals in the inner disc regions (Ricci et al., 2015; Yelverton & Kennedy, 2018). Thus, if my collisionless model were applied to any physically realistic profile (i.e., with $p > 0$,

Equation (5.1)), it is unlikely that I would reproduce the observed brightness peaks. However, it is possible that a shallower density slope than $p = 1$ could generate comparable brightness peaks at times $t \sim \tau_{\rm sec}$, when my model produces an overdensity just exterior to the depletion (see Stage 2 in Section 5.6.1).

The above discussion suggests that although my mechanism acting alone can produce a structure qualitatively similar to that observed in HD 107146, it does not provide a quantitative interpretation of the observations. However, I re-emphasise that my aim in this chapter was not to provide a complete description of the HD 107146 disc, but rather to provide a proof-of-concept for my mechanism and its feasibility. I also stress that the limitations of my simple model need to be assessed before making any definitive conclusions (see Section 5.8 for a detailed discussion). Nevertheless, my results serve as a starting point to guide future, more comprehensive studies that aim to match the observations of the HD 107146 disc, or any other disc with an observed gap.

5.7.2 Application to Other Systems

Given the potential ubiquity of gaps in debris discs (e.g. Kennedy & Wyatt, 2014; Marino et al., 2020, see also Section 2.3.1), it is also possible that future surveys will reveal a sample of discs with asymmetric gaps. Two potential candidates for such systems are HD 92945 (Marino et al., 2019) and HD 206893 (Marino et al., 2020; Nederlander et al., 2021), which I discuss below.

5.7.2.1 HD 92945

I first consider the system HD 92945 (Golimowski et al., 2011), which is often viewed as a sibling to HD 107146 in many ways. Both systems not only have stars with similar masses and ages ($1 M_{\odot}$ and $100 - 300$ Myr, Plavchan et al. (2009)), but also their discs show some similarities in terms of their radial structure. Indeed, ALMA observations of Marino et al. (2019) show that the HD 92945 disc, extending from ~ 50 to 140 au, is double-peaked with a gap centred at about ~ 73 au, roughly coincident with that in HD 107146; see Figure 2.7. However, and in contrast to HD 107146, the gap in HD 92945 appears to be asymmetric and is relatively narrow with an estimated width of 20^{+10}_{-8} au (Marino et al., 2019).

These features speak in favour of my model, so I could use my results in Section 5.4 to determine the properties of the planet and disc such that the gap is sculpted by secular resonances. Figure 5.13 summarises the results of my analysis (following a

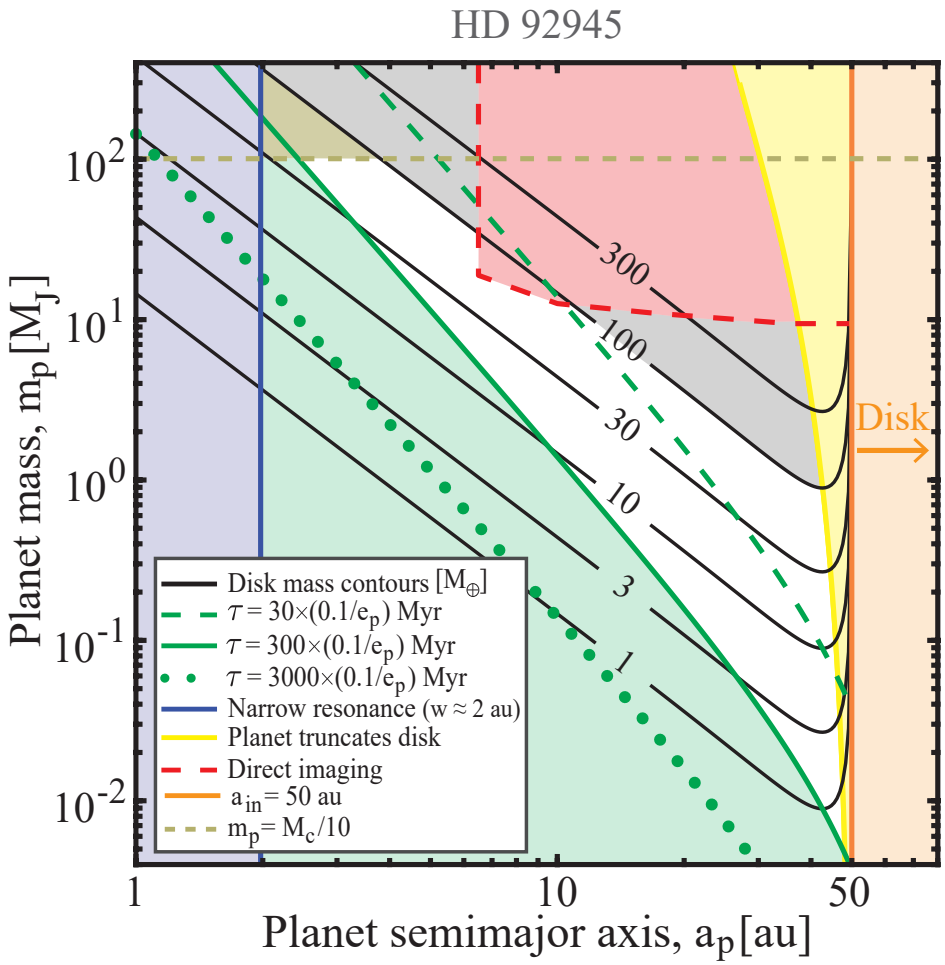

Figure 5.13. Similar to Figure 5.8, but for an HD 92945–like disc. The white region represents the disc–planet parameters that place a secular resonance at 73 au such that it acts on a time-scale less than the stellar age (i.e., ≤ 300 Myr) and is wide enough to have an observable effect. All other notations and exclusion criteria are similar to those in Figure 5.8, except that here I have also excluded planet masses that exceed one-tenth of the central star mass, i.e., $m_p \geq M_c/10$ (olive shaded region in the top part of the parameter space). See the text (Section 5.7.2.1) for details.

similar reasoning as for HD 107146 in Section 5.5). I find that a companion with a semimajor axis a_p in the range $\sim 3 - 50$ au and mass m_p between $\sim 10^{-2}$ and $10^2 M_J$ can produce a wide enough gap at the observed location within the stellar age, provided that $1 \lesssim M_d/M_\oplus \lesssim 100$ – see the white region in Figure 5.13. These limits are in agreement with (i) direct imaging constraints explored by Biller et al. (2013) – see red curve in Figure 5.13, and (ii) disc mass estimates of $\sim 100 - 200 M_\oplus$ derived from collisional models in Marino et al. (2019).

Finally, I note that since the inner disc edge in HD 92945 is located at ~ 50 au, i.e., farther out than in HD 107146, it is possible for the planet to be on a more distant orbit than in HD 107146 (Figure 5.13). However, I confirmed that this is only necessary if the true gap width is toward the upper end of its estimated range (recall that increasing $a_p/a_{\rm in}$ in my model leads to wider gaps). For instance, I find that invoking a planet similar to that in `Model A` (but with a disc of mass $M_d \approx 16.4 M_\oplus$) produces a ~ 16 au wide gap, which is comparable to that observed. Future observations of this system could help to put better constraints on the disc mass and planetary properties.

5.7.2.2 HD 206893

I next consider HD 206893, a $50 - 700$ Myr old F5V star, which hosts a debris disc (Marino et al., 2020; Nederlander et al., 2021) as well as one brown dwarf companion, HD 206893 B, detected using direct imaging (Milli et al., 2017a). ALMA observations of Marino et al. (2020) show that this disc, extending from ~ 30 to 180 au, features an asymmetric ~ 27 au wide gap centred at ~ 75 au (see also Figure 2.7). Given that HD 206893 B orbits interior to the disc with $a_p \sim 11$ au (Delorme et al., 2017), this system is ideally suited to test whether my model can reproduce the observed gap.

To assess this, I adopt the minimum possible mass of HD 206893 B ($\sim 12 M_J$, Delorme et al., 2017) and calculate, using Equation (5.16), the disc mass that would place a secular resonance at the observed gap location, i.e., $a_{\rm res} = 75$ au. Assuming a surface density profile with $p = 1$ (Equation (5.1)), I find that the required disc mass is $M_d \approx 170 M_\oplus$; see also Equation (5.18). This is roughly consistent with the disc mass estimates of Marino et al. (2020) based on collisional models. Moreover, I also confirmed that the gap width w_g obtained from my model agrees well with that observed: adopting the best-fitting eccentricity of HD 206893 B, $e_p \sim 0.15$ (Marino et al., 2020), I find that $w_g \approx 26$ au after ~ 20 Myr of evolution. If future observations with better resolution confirm that the gap in the HD 206893 disc is indeed wider toward the companion's pericentre position, this will then provide a strong support to my model.

Finally, I note that recent analyses of HD 206893 have indicated that it is likely that this system harbours a second inner companion at ~ 2 au (Grandjean et al., 2019; Marino et al., 2020). While in this and subsequent chapters I only consider single-planet systems, my model may easily be extended to two-planet systems (or more). In this case, depending on the strength of perturbations from the companion(s), my results both in general (e.g., Section 5.2) and for HD 206893 may or may not be

affected significantly. Although such an analysis is beyond the scope of this thesis, I briefly discuss this caveat next.

5.7.3 Potential Application to Multi-planet Systems

Throughout this chapter, I only considered what is arguably the least complex planetary system architecture: a single planet orbiting interior to a massive disc. This was partly motivated by the desire to demonstrate that, for all its simplicity, the model can produce debris disc morphologies (in particular, gaps) which would otherwise may not be possible with massless disc models. This said, however, the model presented in Section 5.2 may easily be extended to systems composed of two (or more) planets orbiting interior to the disc. The presence of additional planet(s) may or may not affect my results, depending on the perturbation strength of the additional planet(s).

In a two-planet system, for instance, it is straightforward to expect that my results would remain roughly the same if the perturbations due to the additional planet are negligible, e.g., if it is much less massive and closer to the central object than its counterpart. The extreme, of course, is a system where the additional companion overshadows the gravitational effects of the disc – even if the latter is relatively massive, say, with $M_d \sim 100 M_\oplus$. Such a case would be reminiscent of the setup in Yelverton & Kennedy (2018), where the authors show that two planets carve a crescent-shaped gap – similar to that I find in my study (Section 5.6) – centred around *one* of the two secular resonances they establish within an external, *massless* disc. The transition between these two extreme cases remains an interesting scenario to explore. In this case it may be possible to carve either two or a single but broader gap in the disc, depending on the properties of the secular resonances of the "two planets + massive disc" system, which, in principle, can feature up to four resonances (where two of them will be near $a_{\rm in}$ due to disc edge effects; see Section 5.3.2). A detailed investigation of the potential effects of an additional planet on my results is beyond the scope of this dissertation and is best deferred to a future study. Nevertheless, I acknowledge that it could be important for the location (if not number) of secular resonances and thus is crucial for constraining the disc–planet parameters based on imaged gap structures.

5.7.4 Implications for Disc Mass Estimates

My results may be used to infer the presence of a yet-undetected planet in any system harbouring a double-ringed debris disc. The inferences are, of course, degenerate with the assumed system parameters but, more importantly, they are subject to the

condition that there be sufficient mass in the disc (Sections 5.4 and 5.5). Thus, the detection of planets with the inferred properties will not only provide strong support to my model, but also – and more importantly – provide a unique way to indirectly measure the total mass of the debris disc M_d (see, e.g., Section 5.7.2.2). This is particularly appealing, considering the fact that M_d cannot be accessed using other techniques – not least without invoking theoretical collisional models to extrapolate observed dust masses to the unobservable larger planetesimals that carry most of the disc mass (for a detailed discussion, see Section 2.6 and Krivov & Wyatt, 2021). This represents a promising avenue to consider in the future, in particular with the advent of new generation instruments such as JWST, which could detect planets with $m_p \lesssim 10 M_J$ at $a_p \sim 10$ au separations. Conversely, the results of Section 5.4 may be used to investigate whether or not the debris disc of a known planet-hosting system should have a gap. Future observations of such systems, e.g., with ALMA looking for evidence – or lack thereof – of a gap could help in constraining the total disc mass.

5.7.5 The Importance of Disc Self-gravity in Dynamical Modelling of Debris Discs

The study presented in this chapter has further consequences beyond an explanation of gap formation in debris discs. Particularly, my findings strongly emphasise the need to account for the (self-)gravitational effects of discs in studies of planet–debris disc interactions. As I showed in this chapter, the end-state of secular interactions between a single planet and a disc having only a modest amount of mass can be radically different from the naive expectations based on a massless disc. Indeed, if it were not for the disc gravity in my model, secular resonances would have not been established and so no gap would have formed in the disc – at least not without invoking two or more planets (e.g., as done by Yelverton & Kennedy, 2018), or a single but precessing planet (Pearce & Wyatt, 2015).

This also highlights an important caveat related to the dynamical modelling of debris discs in general. While studies treating debris discs as a collection of massless particles seem to successfully reproduce a large variety of observed disc features by invoking unseen planets (e.g., see Chapter 2.4 and reviews by Krivov, 2010; Wyatt, 2018), their inferences about the underlying planetary system architecture may be compromised. The inclusion of disc gravity would – at least – impose modifications on the masses and orbital properties, if not numbers, of invoked planets. Thus, caution

must be exercised in the interpretation of observed disc structures when the disc mass is ignored.

Recently, Dong et al. (2020) raised a similar point when it comes to ascribing observed morphologies of discs (assumed to be massless) to single planets in situations where the potential presence of a second planet is ignored. I urge a similar analysis to be performed by considering a natural hypothesis of having nonzero disc mass in contrast to the potential presence of additional planets. Although this is beyond the scope of my dissertation, the formalism outlined in Section 5.2 could provide a useful starting point for such an analysis. To summarise, the inclusion of disc self-gravity in studies of planet–disc interactions should be considered in dynamical modelling of debris discs.

5.8 Limitations of the Model

I now discuss some of my model assumptions and limitations. This will also serve as a motivation for the study presented in the forthcoming chapter (i.e., Chapter 6).

5.8.1 Treating Planetesimals as Test Particles

Throughout this chapter, I treated planetesimals as massless test particles, and analysed their secular evolution under the influence of gravity from both the planet and the debris disc. To this end, I modelled the debris disc as being *passive*: that is, as a rigid slab that provides *fixed* axisymmetric gravitational potential (see Equation (5.3) and Section 5.2; disc non-axisymmetry is discussed next in Section 5.8.2). Thus, at first glance, it appears that instead of the planetesimals contributing to the collective potential of the disc, they are enslaved by the fixed disc potential given in Equation (5.3). In reality, though, these two approaches are subtly similar. This is because the orbit-averaged disturbing function for a planetesimal of mass m_j due to all other N massive planetesimals in a disc – in the continuum limit (i.e. $N \to \infty, m_j \sim N^{-1}$) – is equivalent to that in Equation (5.3). This can be verified by a somewhat tedious but straightforward calculation, as done in Chapter 4, which requires softening the gravitational interaction between massive planetesimals, integrating radially over all planetesimals, and taking the limit of zero softening – see e.g. Figure 4.10, Section 4.6.3, and Hahn (2003).

To further justify this equivalence, I simulated the secular dynamics of disc–planet systems by modeling the disc as a swarm of N massive planetesimals, each represented

as a ring[11], that interact via softened gravity (see e.g. Chapter 4, Hahn 2003; Touma et al. 2009; Batygin 2012). I found that simulations carried out with a negligible softening parameter accurately reproduce the analytical solutions presented in Section 5.2.3 (which is, of course, possible only when the non-axisymmetric perturbations due to simulated disc particles are neglected, i.e., as in Section 5.2). Quantitative results showing this equivalence are provided later in Section 6.2.3.2 of Chapter 6, where the softened "N-ring" model is introduced.

5.8.2 Non-axisymmetric Component of Disc Gravity

A major limitation of the work presented in this chapter is that I only accounted for the axisymmetric contribution of the disc gravity, ignoring its non-axisymmetric component (Section 5.2). That is to say, my model does not account for the non-axisymmetric perturbations that disc particles can exert both among themselves and onto the planet (see Section 5.8.1), even though I found that the disc naturally develops non-axisymmetry (Sections 5.6.1 and 5.6.2). This omission allowed me to elucidate the key effects of disc gravity (semi-)analytically. This comes, however, at the expense of reduced coupling within the system that inhibits the exchange of angular momentum between the disc and planet. Thus, the theory outlined in this chapter serves as a first step toward a comprehensive understanding of the role played by disc gravity and its observational implications.

Previous studies of gravitating disc–planet systems (which include the full gravitational effects of disc particles) have shown that an eccentric planet could launch a long, one-armed, spiral density wave at a secular resonance in the disc (Ward & Hahn, 1998a; Hahn, 2003, 2008). Such spiral waves propagate away from the resonance location as trailing waves with pattern speed equal to the planetary precession rate. These waves also transfer angular momentum from the disc to the planet in a way that damps the planet's eccentricity, without affecting its semimajor axis[12] (Goldreich & Tremaine, 1980; Ward & Hahn, 1998a; Tremaine, 1998; Ward & Hahn, 2000). My idealised model is not designed to capture the full richness of such dynamical phenomena. Thus, a more sophisticated analysis is crucial, and this will be the subject of the forthcoming chapter (i.e., Chapter 6).

[11] Recall that orbit-averaging is equivalent to smearing particles into massive rings along their orbits, where the line-density of each ring is inversely proportional to the orbital velocity of each particle (Murray & Dermott, 1999).

[12] This process is referred to in the literature as "resonant friction" (Tremaine, 1998) or "secular resonant damping" (Ward & Hahn, 2000).

5.8.3 Secular Approximation

I limited the expansion of the secular disturbing function to second order in eccentricities (e.g., see Sections 5.2 and 6.2). Hence, my results are only approximate at high eccentricities, e.g., in the vicinity of the secular resonances, where it is necessary to include higher-order terms in the disturbing function (e.g., see Section 4.6.4; Sefilian & Touma, 2019). Such an exercise would, primarily, limit the eccentricity amplitude at the resonance (Malhotra, 1998). Nevertheless, it seems unlikely that this would affect the gap formation. For instance, from Figures 5.9 and 5.10, one can see that the gap is already well-developed when eccentricities at the resonance are still rather modest, i.e., $e \sim 0.2$. Higher-order terms, however, could give rise to mild quantitative differences in terms of the dynamical timescales, e.g., period of eccentricity oscillations.

I also ignored MMRs between the planet and the planetesimals. Previously, Tabeshian & Wiegert (2016) found in simulations of synthetic debris discs that gaps can be carved at the 2:1 MMR with an internal low-e planet ($e_p \lesssim 0.1$; see also Regály et al., 2018). In my simulated systems, this can occur around $\simeq a_{\rm in}$. However, as the authors explain, MMR gaps will be blurred or even washed out by high-eccentricity planetesimal orbits farther out in the disc. In my case, this could be easily achieved by planetesimals in the vicinity of the secular resonance.

5.9 Summary

In this chapter, I explored the secular interaction between an eccentric planet and an external self-gravitating debris disc, using a simplified analytic model. The model is simplified in the sense that it only accounts for the axisymmetric component of the disc (self)-gravity, ignoring its non-axisymmetric contribution. Despite this limitation, however, this is the first time (to the best of my knowledge) that the effects of disc gravity have been considered analytically in such detail in the context of debris discs. I employed the analytic model to assess the possibility of forming gaps in debris discs through excitation of planetesimal eccentricities by the secular apsidal resonances of the system. I summarise my key results below.

(i) When the debris disc is less massive than the planet, $10^{-4} \lesssim M_d/m_p \lesssim 1$, the combined gravity of the disc and the planet can mediate the establishment of two secular apsidal resonances in the disc.

(ii) I mapped out the behaviour of the characteristics of the secular resonances – i.e. locations, time-scales, and widths – as a function of the disc and planet

parameters. In particular, I found that one of the secular resonances can lead to the formation of an observable gap over a broad region of parameter space.

(iii) As an example, I applied my results to HD 107146 and HD 92945, and showed how the properties of a yet-undetected planet, together with the mass of the debris disc, can be constrained to produce a gap at the observed location. In the case of HD 206893, I find that the directly imaged companion can sculpt the observed gap if the debris disc is $\approx 170 M_{\oplus}$ in mass.

(iv) By investigating the secular evolution in such systems, I identified three distinct evolutionary stages that occur on timescales measured relative to the planetary precession period. I find that the gap forms by the time the planet has completed approximately one precessional cycle, on a timescale of tens of megayears.

(v) Independent of the system parameters, the gap carved around the secular resonance is asymmetric: it is both wider and deeper in the direction of the planetary pericentre. Additionally, its fractional depth is always about 0.5. The gap width, however, increases with increasing planetary semimajor axis and/or eccentricity.

(vi) More generally, my results suggest that the gravitational potential of debris discs can have a notable effect on the secular evolution of debris particles. I advocate the inclusion of disc gravity in studies of planet–debris disc interactions.

The mechanism presented here represents what is arguably the simplest pathway to forming gaps in debris discs, akin to those observed in HD 107146, HD 92945, and HD 206893. It may indeed obviate the need for invoking more complicated scenarios, e.g. multiple planets interior to or within the disc.

Finally, I remark that the work presented in this chapter should be envisaged as a first step toward an in-depth exploration of the effects of disc gravity in planet–debris disc interactions. In the forthcoming chapter, I will extend my current calculations using numerical techniques to properly account for the full gravitational effects of the disc.

Chapter 6

Planet–Debris Disc Interactions II: Development of a Self-consistent Model

6.1 Introduction

The previous chapter investigated the secular interaction between an eccentric planet and a massive, external debris disc using a simple analytical model. The model accounted for both the gravitational coupling between the disc and the planet, as well as the disc self-gravity – with the limitation that it ignored the non-axisymmetric component of the disc (self)-gravity. This omission allowed me to elucidate the basic effects of disc gravity and present a novel pathway to sculpting gaps in debris discs. However, the assumption of an axisymmetric disc potential is too simplistic since, as we already saw, the disc can naturally develop some degree of non-axisymmetry. Such a disc would have a non-axisymmetric component of its gravitational potential, affecting the evolution of both the constituent planetesimals and the planet. Thus, a more complete treatment of the disc (self-)gravity is warranted to uncover its complete spectrum of effects in secular planet–debris disc interactions.

This chapter provides a natural but important extension of Chapter 5 by incorporating the full gravitational effects of the disc, both axi- and non-axisymmetric contributions. I accomplish this goal by working within the context of the continuum version of the classical Laplace-Lagrange theory (Murray & Dermott, 1999) using the softening prescription of Hahn (2003): i.e., by modelling the disc as a collection of N softened, massive rings interacting both with each other and with the planet (Chapter

4). Owing to the more complex nature of this problem, I approach it in two steps. First, in addition to the disc's axisymmetric potential, I account for its non-axisymmetric gravitational component acting on the planet, i.e., without considering its effects on the disc particles – a case I refer to as 'nominal' simulations. I then move on to account for the full gravitational field of the disc, including its non-axisymmetric perturbations both on the planet and the disc itself – a case I refer to as 'full' simulations. The results of such full simulations are preliminary, in the sense that I plan on expanding on them in the near future, and thus should be regarded as a work in progress. For ease of comparison with the results of Chapter 5, I perform such N-ring simulations using the same planet–disc parameters identified there as capable of reproducing the gap observed in the HD 107146 disc (Figure 5.8). As we shall see, this approach will prove useful in building up a more systematic understanding of the effects of disc gravity in planet–debris disc interactions, both generally and specifically for gapped debris discs.

The structure of this chapter is as follows. First in Section 6.2, I introduce the framework describing the continuum version of the classical Laplace–Lagrange theory – hereafter, the N-ring model – that will be used to simulate the secular evolution of planet–debris disc systems. After performing some tests of the N-ring model, I present the main results obtained within the set of 'nominal' simulations in Section 6.3. Then in Section 6.4, I analyse the results of a suite of nominal N-ring simulations and provide quantitative explanations for the differences identified when compared to the expectations based on the simplified model of Chapter 5. Section 6.5 then presents and discusses some of my key preliminary results obtained within the set of full N-ring simulations, describing the two qualitatively different outcomes observed. I discuss the overall results of this study along with their theoretical and observational implications in Section 6.6, where I also critically assess the limitations of my model, discuss planned future work, and summarise the key results presented in this chapter.

6.2 Laplace-Lagrange secular theory: continuum version

Laplace–Lagrange theory represents one of the best-known results of celestial mechanics for studying the secular evolution of planetary systems. However, as already discussed in Section 4.1, this framework is ill-posed from a mathematical point of view when applied to self-gravitating, massive discs. Nevertheless, the results presented in Chapter 4 suggest that the continuum version of the classical Laplace–Lagrange theory, when modified by the introduction of a small but non-zero softening parameter, can well

reproduce the expected results arising from unsoftened calculations of the orbit-averaged disc potential, e.g. as done by Heppenheimer (1980). In this section, I present the equations describing the continuum version of the classical Laplace–Lagrange theory, or the N-ring model, based on the softening prescription of Hahn (2003) (Chapter 4). However, prior to this, it is necessary to define a number of basic parameters of the model.

6.2.1 Parameters of the Model

The general setup of the planet–disc system that I consider in this chapter is the same as in Chapter 5: a central star of mass M_c that is orbited by a planet of mass m_p and an external, coplanar debris disc of mass M_d, such that $m_p, M_d \ll M_c$ – see Figure 5.1. Similar to Chapter 5, I also assume the planetary orbit – which is characterised by its semimajor axis a_p, eccentricity e_p, and longitude of pericentre ϖ_p – to be initially slightly eccentric (i.e., $e_p \lesssim 0.1$) so that it does not cross the disc along its orbit. Additionally and once again, I consider the debris disc to be initially axisymmetric with a surface density profile that is described by the truncated power-law given by Equation (5.1).

Unlike Chapter 5, however, in which the debris disc was modelled as a continuous entity providing a fixed gravitational potential (Section 5.8.1), I now adopt a discretised description in which the orbits of disc particles slowly flex over time due to perturbations arising from the gravity of both the disc and the planet. In other words, I model the debris disc as a series of $N \gg 1$ nested massive rings (see e.g. Figure 4.1), each of which interacts with the other disc rings as well as the planet, and in the process affect the gravitational potential generated by the entire disc. This allows for a self-consistent representation of the disc's (self-)gravity. Note that qualitatively speaking, replacing particle orbits by massive rings is equivalent to orbit-averaging the gravitational potential generated by the particles (Murray & Dermott, 1999, see also Chapter 3), and thus each ring will be characterised by a non-uniform linear density that is inversely proportional to the orbital velocity of a particle at a particular phase of its orbit; see Chapters 3 and 4. Note that throughout this chapter, I shall use the words "planetesimals", "disc rings", and "debris particles" interchangeably.

Conceptually each of the disc rings may be thought of as a swarm of planetesimals, all sharing the same semimajor axis a_j, eccentricity e_j, and longitude of pericentre ϖ_j (with $j = 1, ..., N$). Thus, I also assume that the orbital eccentricity of each ring is small throughout the disc, so that the velocity dispersion of neighbouring rings is small when compared to the Keplerian velocity at a given semimajor axis. Accordingly,

I define the aspect ratio H of the disc as a small parameter and assume that it is constant throughout the entire range of the disc:

$$H = \frac{\mathfrak{h}_j}{a_j} = \text{const} \ll 1, \tag{6.1}$$

with $\mathfrak{h}_j \propto a_j$ being a measure of the vertical and radial half-thickness of the j-th disc ring (Hahn, 2003). Here, it is important to note that the parameter H is one of the fundamental parameters entering in the N-ring model, as it represents the magnitude of the gravitational softening parameter; see Chapter 4 and Hahn (2003).

To complete the discretised description of the debris disc, one further needs to assign the semimajor axes and masses of the disc rings. To this end, the semimajor axes of the disc rings are taken to be constant in time and distributed logarithmically between the inner and outer disc edges, a_{in} and a_{out}, respectively, such that the ratio of spacing between any two adjacent rings is constant, i.e., $a_{j+1}/a_j = (a_{\text{out}}/a_{\text{in}})^{1/N}$. The first of these assumptions is justified by the fact that I am primarily interested in the secular evolution of the system, and thus the semimajor axes are conserved (Murray & Dermott, 1999). As to the masses of the disc rings m_j, they are assigned such that the debris disc's surface density follows the profile given by Equation (5.1), and thus the total disc mass $M_d = \sum_{j=1}^{N} m_j$ is given by Equation (5.2). For reference, a geometrical representation of the debris disc is shown in Figure 4.1. Note that this setup is very similar to that explored in Hahn (2003) in the context of secular interactions between Neptune and the primordial, massive Kuiper belt.

In this chapter, unless otherwise stated, I adopt a fiducial disc model with $a_{\text{in}} = 30$ au and $a_{\text{out}} = 150$ au (i.e., $\delta \equiv a_{\text{out}}/a_{\text{in}} = 5$), and a surface density power-law index of $p = 1$ (Equation 5.1). The reasoning for these choices is to make contact with the analytic work presented in Chapter 5. However, it is important to note that the N-ring model described in this chapter is applicable to arbitrary surface density profiles, whether power-law or otherwise. Additionally, I model the disc as composed of $N = 5000$ rings with an aspect ratio of $H = 0.1$, while the planet is modelled as an unsoftened thin ring, i.e., having $H = 0$. These choices are motivated by the results of Chapter 4, namely Sections 4.5 and 4.6.3, and will be further justified and discussed later in Section 6.2.3.3.

6.2.2 The N-ring Model: Governing Equations

With the preliminary specifications of the setup now in place, I move on to describe the equations governing the secular evolution of the planet–disc system.

6.2.2.1 The Disturbing Function

Consider the eccentricity dynamics of a ring labeled by index $j = 1, ..., N$ embedded within a broad disc. According to Chapter 4, the secular disturbing function R_j that governs the evolution of the ring j is determined by the contribution of all other perturbing rings $k \neq j$ such that, to second order in eccentricities, one has:

$$R_j = n_j a_j^2 \left[\frac{1}{2} A_{jj} e_j^2 + \sum_{k=0, k\neq j}^{N} A_{jk} e_j e_k \cos(\varpi_j - \varpi_k) \right], \tag{6.2}$$

where $n_j = \sqrt{G(M_c + m_j)/a_j^3}$ is the mean motion of the perturbed j-th ring, and the meanings of different constants is explained below. To arrive at Equation (6.2), I have used the definition of the mass density per unit semimajor axis appearing in Equations (4.4) – (4.6) of Chapter 4 – namely, $\mu_d(a) = 2\pi a \Sigma_d(a)$ – and replaced the integrals appearing in those equations by a summation over discrete semimajor axes. Note that in Equation (6.2), the disc is broken up into N rings indexed as $j = 1, ..., N$, with j running from the inner to the outer disc edge, and I have introduced one additional ring indexed as $j = 0$ to represent the planet. Thus, the entire planet-disc system is modelled as composed of $N + 1$ rings. The introduction of the planet in Equation (6.2) is justified by the fact that the mathematical structure of the disturbing function due to a disc (see e.g. Section 4.2 and Equation 4.4) is the same as the disturbing function due to a planet when the softening is set to zero (see e.g. Chapter 4 and Equation 5.3).

The coefficients A_{jj} and A_{jk} appearing in Equation (6.2) are the discretised analogues of A_d and $\mathbf{B}_d$ defined in Chapter 4 (Equations (4.5) and (4.6)), and thus their expressions take the following forms:

$$A_{jj} = \frac{1}{4} n_j \sum_{k=0, k\neq j}^{N} \frac{m_k}{M_c + m_j} f(\alpha_{jk}, H), \tag{6.3}$$

$$A_{jk} = \frac{1}{4} n_j \frac{m_k}{M_c + m_j} g(\alpha_{jk}, H), \qquad j \neq k, \tag{6.4}$$

for the softening prescription of Hahn (2003). In Equations (6.3) and (6.4), α_{jk} is defined such that $\alpha_{jk} \equiv a_k/a_j$, and the functions $f(\alpha, H)$ and $g(\alpha, H)$ which fully characterise the ring-ring interactions are given by:

$$f(\alpha, H) = 8\phi_{11} = 8\phi_{22} = \alpha \mathcal{B}_{3/2}^{(1),H} - 3\alpha^2 H^2 (2 + H^2) \mathcal{B}_{5/2}^{(0),H}, \tag{6.5}$$

$$g(\alpha, H) = 4\phi_{12} = -\alpha \mathcal{B}_{3/2}^{(2),H} + 3\alpha^2 H^2 (2 + H^2) \mathcal{B}_{5/2}^{(1),H}, \tag{6.6}$$

where $\mathcal{B}_s^{(m),H}$ are the softened Laplace coefficients corresponding to the softening formalism of Hahn (2003) – see Equation (4.14) – and the expressions of the functions $\phi_{ij}(\alpha, H)$ (with $i, j = 1,\ 2$) are borrowed from Table 4.1.

Note that unlike the expressions of the disturbing function developed in standard textbooks (e.g. Brouwer & Clemence, 1961; Murray & Dermott, 1999), the disturbing function given by Equation (6.2) is valid for cases where the perturbed ring is interior to or exterior to the perturbing ring, i.e., $\alpha_{jk} > 1$ or $\alpha_{jk} < 1$, respectively, and there is no need for two distinct expressions. This follows from the fact that the coefficients ϕ_{ij} resulting from the softening prescription of Hahn (2003) are symmetric when α is replaced with α^{-1}; see Chapter 4 and Equation (6.11). Nevertheless, it is trivial to show that the disturbing function (6.2) reduces to the classical expressions in Murray & Dermott (1999) (e.g. their equations 7.6 and 7.7) upon setting the softening parameter equal to zero for all considered rings, $H = 0$.

Finally, I remind the reader of the physical meanings of the coefficients A_{jj} and A_{jk}, highlighting their relation to the definitions used in Chapter 5. The coefficient A_{jj} given by Equation (6.3) represents the precession rate of the free eccentricity vector of the j-th ring due to all other rings in the system. Thus, A_{jj} with $j = 0$ is the free precession of the planet due to the disc, i.e., $A_{d,p}$ given by Equation (5.7) in Chapter 5, whereas A_{jj} with $1 \leq j \leq N$ is the free precession rate of the j-th disc particle due to the axisymmetric components of both the disc and planet gravity, i.e., the sum of A_d and A_p given by Equations (5.5) and (5.4) in Chapter 5, respectively. Similarly, the coefficient A_{jk} given by Equation (6.4) represents the non-axisymmetric perturbations of the j-th ring due to the perturbing ring indexed k. Thus, A_{jk} with $k = 0$ represents the non-axisymmetric perturbations exerted by the planet on the j-th disc ring, i.e. B_p given by Equation (5.6) in Chapter 5. On the other hand, the coefficient A_{jk} evaluated at $j = 0$ represents the non-axisymmetric perturbations that the k-th disc ring exerts on the planet, and when evaluated at $j \neq 0$, it represents the non-axisymmetric perturbations that the disc rings exert amongst themselves. These two contributions to the secular dynamics of the system were neglected in Chapter 5, and thus the model presented here allows for a more self-consistent treatment of the problem.

6.2.2.2 Evolution Equations and Their Solution

With the expression of the disturbing function in place (Equation 6.2), the time variation of the rings' orbital elements can be determined by the aid of Lagrange's planetary

equations which, to leading order in eccentricities, read as (Murray & Dermott, 1999):

$$\begin{aligned} \frac{dk_j}{dt} &\approx -\frac{1}{n_j a_j^2}\frac{\partial R_j}{\partial h_j} = -\sum_{k=0}^{N} A_{jk} h_k, \\ \frac{dh_j}{dt} &\approx \frac{1}{n_j a_j^2}\frac{\partial R_j}{\partial k_j} = \sum_{k=0}^{N} A_{jk} k_k. \end{aligned} \tag{6.7}$$

Here, and as is customary in planetary dynamics, I have expressed the eccentricity vector in terms of polar coordinates so that $\mathbf{e} \equiv (k, h)$ with

$$k = e\cos\varpi, \quad \text{and} \quad h = e\sin\varpi. \tag{6.8}$$

The system of equations (6.7) can be written in a more compact form when expressed in terms of the complex Poincaré variable, $\zeta \equiv e\,\exp(i\varpi) = k + ih$, so that:

$$\frac{d\zeta_j}{dt} = i\sum_{k=0}^{N} A_{jk}\zeta_k. \tag{6.9}$$

Equation (6.9) represents the key equation needed for the work presented in this chapter, allowing me to probe the secular evolution of the N-rings system.

Here, I remark that the coefficients A_{jj} and A_{jk} (Equations 6.3 and 6.4) can be regarded as the time-independent entries of an $(N+1)\times(N+1)$ square matrix $\mathbf{A}$ (Murray & Dermott, 1999), which fully encapsulates the mutual gravitational interactions among the $N+1$ rings in the system. Thus, the equations of motion given by Equation (6.9) constitute an eigen-system which can be solved using standard methods, i.e., akin to the problem of N coupled harmonic oscillators (see Chapter 7 of Murray & Dermott, 1999). Indeed, the time evolution of ζ_j can be written in closed form as follows (Murray & Dermott, 1999):

$$\zeta_j(t) = \sum_{k=0}^{N} T_k E_{jk}\,\exp[i(g_k t + \delta_k)], \tag{6.10}$$

where g_k and E_{jk} represent the eigenvalues and eigenvectors of the matrix $\mathbf{A}$, respectively, while δ_k and T_k are constants of integration determining the phases and relative amplitudes of the eigenvectors, respectively. A handy recipe for determining these constants is given in Murray & Dermott (1999). Despite the analytic nature of the solution, however, this method can become cumbersome and inefficient for the large number of rings, $N \sim \mathcal{O}(10^3)$, that must be considered for accurate representation of the secular dynamics (see Section 4.6.3). Thus, one can alternatively integrate the equa-

tions of motion in Equation (6.9) numerically e.g. with a conventional Runge–Kutta ODE solver (Press et al., 2002). This is the approach I opted for in this chapter.

6.2.2.3 Numerical Implementation

Based on the description of the N-ring model above, I developed a code to solve the resultant equations of motion. First, the code calculates the matrix $\mathbf{A}$ for a given set of rings. Note that the entries of this matrix depend only on the masses of the rings m_j, as well as their semimajor axes a_j and aspect ratio $H = \mathfrak{h}_j/a_j$ (with $j = 0, ..., N$) – see Equations (6.1), (6.3), and (6.4). The latter two parameters appear in the definition of the Laplace coefficients $B_s^{(m),H}$, which can be rapidly evaluated by making use of their relationship to elliptic integrals outlined in Appendix C.3, that is, without relying on numerical integration of Equation (4.14). Additionally, note that $\mathcal{B}_s^{(m),H}$ obeys the relationship given in Equation (C.4) which relates the Laplace coefficients of arguments α and α^{-1}, and thus one also has:

$$\begin{aligned} f(\alpha^{-1}, H) &= \alpha f(\alpha, H), \\ g(\alpha^{-1}, H) &= \alpha g(\alpha, H), \end{aligned} \tag{6.11}$$

further reducing the number of evaluations required to populate the matrix $\mathbf{A}$. Once the matrix $\mathbf{A}$ is computed, the code then takes as input the initial eccentricities and apsidal angles of the rings, $e_j(0)$ and $\varpi_j(0)$, and integrates the equations of motion (6.9) using a six-stage, fifth-order, Runge-Kutta method (e.g. Press et al., 2002) with a variable time step for efficient computation. The times steps are chosen in such a way that they ensure a maximum relative error of 10^{-8} per time step. Finally, to check that my code behaves as expected, I devised and carried out several tests: these are presented next in Section 6.2.3.

In the coming sections, I will report on simulations carried out using this N-ring code at three levels of approximation: (1) a 'simplified' case, whereby the disc rings' non-axisymmetric perturbations both amongst themselves and onto the planet are turned off, i.e., $A_{jk} = 0$ for all but $k = 0$ (where by definition $j \neq k$, Equation 6.4); (2) a 'nominal' case, whereby the disc's non-axisymmetric potential does operate on the planet but not on the disc particles themselves, i.e., $A_{jk} = 0$ for all but $j = 0$ and $k = 0$; and finally (3) a 'full' case, whereby all perturbations are accounted for.

6.2.3 Tests of the N-ring Model

In this section, I describe the tests that I carried out in order to ensure the proper operation of the N-ring model outlined above. More specifically, I show that the N-ring model (i) conserves the total angular momentum of the system to second order in eccentricities, i.e. to the same level of precision as the disturbing function (Equation 6.2); (ii) reproduces the analytical results of Chapter 5, if and when the non-axisymmetric gravitational potential of the disc is neglected; (iii) is not significantly sensitive to the assumed number of rings N and softening parameter H, as long as the conditions specified in Section 4.6.3 are satisfied; and (iv) reproduces the evolution of planetary systems not comprising any discs, provided the softening parameter is set to zero. The reader not interested in the details of these tests may skip to the next section without loss of continuity.

6.2.3.1 Conservation of Angular Momentum

I first demonstrate analytically that the N-ring model conserves the total angular momentum of the system, L, which can be written as (Murray & Dermott, 1999),

$$L = \sum_{j=0}^{N} L_j = \sum_{j=0}^{N} m_j n_j a_j^2 \sqrt{1-e_j^2}, \tag{6.12}$$

where the sum runs over all rings in the system, i.e., those representing the disc and the planet. Retaining terms up to second order in eccentricities in Equation (6.12) – that is, to the same degree of precision to which the secular disturbing function is expanded (Equation 6.2) – L can be written as $L \approx L_0 - L_e$, where

$$L_0 = \sum_{j=0}^{N} m_j n_j a_j^2, \quad \text{and} \quad L_e = \sum_{j=0}^{N} \frac{1}{2} m_j n_j a_j^2 e_j^2. \tag{6.13}$$

Here, L_0 is the system's total circular angular momentum which, by virtue of orbital averaging, is a constant of motion. This is simply because secular perturbations do not alter the individual orbital energies, or semimajor axes, of interacting bodies (e.g. Chapter 3, see also Murray & Dermott, 1999). The term L_e, on the other hand, is the system's total angular momentum deficit (AMD) which, physically, quantifies the amount of angular momentum that needs to be injected into the system to circularize the orbits of all interacting bodies.

While the individual AMD of each ring is not a conserved quantity and can evolve significantly over secular timescales, the system's total AMD, L_e, is conserved. This can

be demonstrated as follows. First, by making use of the relationship $\frac{1}{2}\frac{de_j^2}{dt} = k_j\frac{dk_j}{dt} + h_j\frac{dh_j}{dt}$ and Equation (6.9), I can express the time derivative of L_e as follows:

$$\begin{aligned}\frac{dL_e}{dt} &= \sum_{j=0}^{N}\sum_{k=0,k\neq j}^{N} m_j n_j a_j^2 A_{jk}(h_j k_k - h_k k_j), \\ &= \sum_{j=0}^{N}\sum_{k=0,k\neq j}^{N} \frac{1}{4}\frac{m_j m_k}{M_c + m_j} n_j^2 a_j^2 g(\alpha_{jk}, H)(h_j k_k - h_k k_j).\end{aligned} \tag{6.14}$$

I next write $dL_e/dt \equiv S_1 - S_2$, with S_1 and S_2 being the terms that involve summing over the $h_j k_k$ and $h_k k_j$ terms in the second line of Equation (6.14), respectively. Since j and k are dummy indices, I interchange them in the expression of S_1, and upon using the relationships $(n_k/n_j)^2 = \alpha_{jk}^{-3}(M_c + m_k)/(M_c + m_j)$ and $g(\alpha_{kj}, H) = \alpha_{jk} g(\alpha_{jk}, H)$, it can be shown after some straightforward algebra that $S_1 = S_2$. This provides an analytic proof that the system's angular momentum deficit, and thus the total angular momentum, is conserved within the N-ring model to second order in eccentricities, i.e., $dL/dt = dL_e/dt = 0$.

Given this result, one can thus use the conservation of the total angular momentum as a reliable diagnostic for the quality of the numerical scheme employed to evolve the N-rings system (Equation 6.9). To this end, I analysed all of the nominal N-ring simulations presented in this Chapter (see Table H.1), finding that the total AMD is conserved in all runs to within a fractional error of $|\Delta L_e/L_e| \sim 10^{-9} - 10^{-8}$. Note that this is the case at all times as long as no ring attains an eccentricity comparable to or larger than 1 at and around the secular resonance in the course of the evolution[1]. Finally out of curiosity, I also repeated some of the simulations by modelling the planet as a thick softened ring, rather than razor-thin unsoftened one, finding that it makes no difference in terms of preserving the system's total angular momentum.

6.2.3.2 Tests Against Known Analytic Solutions

I now demonstrate that the N-ring model can successfully reproduce the analytical results derived in Chapter 5 for the time evolution of the planetesimal eccentricities and apsidal angles – namely, Equations (5.10) and (5.11).

At the outset, I remind that Equations (5.10) and (5.11) were derived under the assumption that the disc contributes to the secular evolution of the system – both the planet and planetesimals – through its axisymmetric component of (self)-gravity

[1] I remind that eccentricities larger than unity are not physical and simply result from adopting a second-order Laplace-Lagrange theory; see e.g. Malhotra (1998); Murray & Dermott (1999).

only, i.e., ignoring the non-axisymmetric component of the disc gravity. Thus, for a meaningful comparison between the numerical results of the N-ring model and the analytical results in Chapter 5, I switched off the terms in the matrix $\mathbf{A}$ which represent the non-axisymmetric perturbations due to the disc rings, i.e., by setting $A_{j,k\neq 0} = 0$ (Equation 6.4), and followed the evolution of the N-rings system. To this end, I adapted the parameters of the fiducial planet–disc system (`Model A`, Tables 5.1 and H.1), modelling the disc as a collection of $N = 5000$ softened rings with $H = 0.1$, each initiated with $e(0) = 10^{-5}$ and $w(0) = -\pi/2$. The planet's potential was left unsoftened. This allows me to directly compare the numerical results with Equations (5.10) and (5.11) stemming from the analytical theory in Chapter 5, and thus directly verifying the accuracy of the N-ring model.

The results of this exercise are illustrated in Figure 6.1, where I plot the time evolution of the eccentricities e (left panels) and apsidal angles $\Delta\varpi$ (right panels) of planetesimals orbiting at semimajor axes of 50, 70, 90, 110, and 130 au (from top to bottom). Note that the apsidal angles are measured relative to that of the precessing planet, i.e., $\Delta\varpi = \varpi - \varpi_p$. Looking at Figure 6.1, one can see that the agreement between the evolution computed by the N-ring model (shown in blue) and the analytical results of Chapter 5 (shown in red) is very good at all semimajor axes. Indeed, the curves of $e(t)$ and $\Delta\varpi(t)$ resulting from the N-ring model follow closely the theoretical predictions of Equations (5.10) and (5.11), with agreement to within several percent in both the amplitudes of eccentricity variations as well as the periods associated with the oscillations of $e(t)$ and $\Delta\varpi(t)$. Generally, the deviations are negligible throughout almost the entire disc, $a_{in} \lesssim a \lesssim a_{out}$, with differences being about $2-3\%$ when compared to the analytical results, see e.g. Figures 6.1(a)–(d). However, the deviations become more pronounced and are about $\approx 10\%$ for planetesimals near the outer edge of the disc, $a \approx a_{out}$, see e.g. Figure 6.1(e). Nevertheless, these deviations are of no practical importance for the purposes of the study presented here, as they are unlikely to have significant effect e.g. on the spatial appearance of the disc due to the already low eccentricities in the outer parts.

Having said that, note that the origin of such deviations can be well explained within the context of the study presented in Chapter 4. There it was shown that the number of disc rings required to accurately capture the secular effects of a continuous disc scales as $N \sim CH^{-2}$ when considering particles far from the disc edges, with the constant of proportionality C being roughly inversely proportional to the desired level of accuracy; see Section 4.6.3 and Figure 4.10. The deviations observed in Figures 6.1(a)–(d) are consistent with those findings. Indeed, according to Figure 4.10, $N \approx 500$

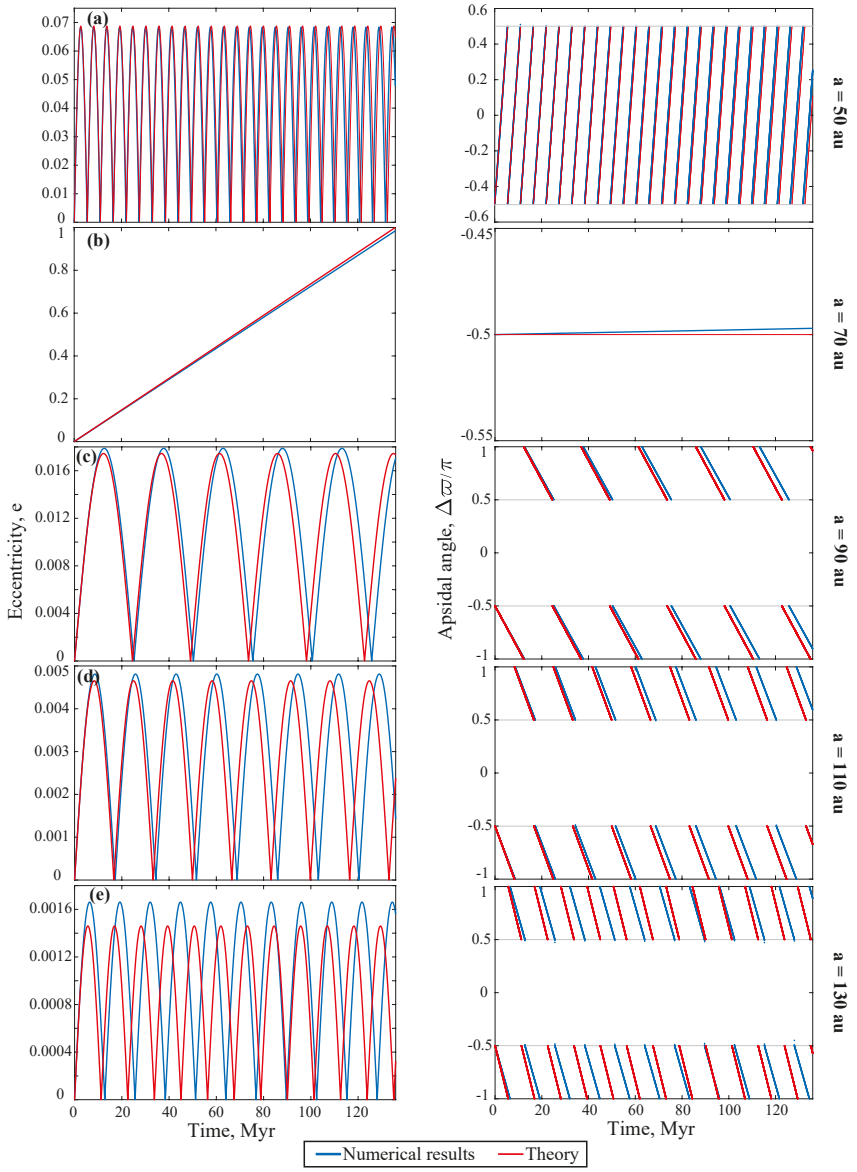

Figure 6.1. Verification of the numerical results obtained for the planetesimal orbital evolution within the softened N-ring model (Section 6.2) using the unsoftened analytical results developed in Chapter 5 (Equations 5.10 and 5.11). The planetesimals were initiated on circular orbits ($e(0) = 10^{-5}$) in the fiducial planet–disc model (`Model A`; Table 5.1). The N-ring calculations assumed 5000 disc rings, each with a softening parameter of $H = 0.1$, while leaving the planet's potential unsoftened. The time evolution of the planetesimal eccentricity e (left panels) and apsidal angle $\Delta\varpi$ (right panels, measured relative to that of the precessing planet) is shown for different values of planetesimal semimajor axes, as indicated to the right of each panel. One can see a very good agreement between the numerical results of the N-ring model and the theoretical results of Chapter 5 at all semimajor axes. See the text (Section 6.2.3.2) for more details.

rings are required to ensure accuracy to within 10% for a softening of $H = 0.1$ (see dark blue line in Figure 4.10), and thus a discrepancy of $\sim 1\%$ is expected for $N = 5000$ rings as in Figure 6.1. Similarly, the relatively larger discrepancy observed in Figure 6.1(e) is consistent with the findings in Section 4.5, where I showed that very small values of softening (i.e., $H \lesssim 0.001$) are required to accurately capture the secular effects near the edges of a disc. Thus, in principle, large deviations should also have been seen near the inner edge of the disc, e.g. at $a = 50$ au in Figure 6.1: however, this is not the case here, since the planet dominates the dynamics in the inner parts of the disc (see e.g. Section 5.3), rendering departures to within several percent in terms of the disc induced precession of planetesimal orbits irrelevant. Nevertheless, as can be seen in Figure 6.4, this causes the system to not establish an additional secular resonance near the inner edge of the disc, i.e, at $\approx a_{\rm in}$ (apart from the one at 70 au, or more generally $\gtrsim a_{\rm in}$), as already anticipated from Chapter 5 – see e.g. Section 5.3.2.

To summarise, the general conclusion one can draw from the comparisons displayed in Figure 6.1 is that the N-ring model described in Section 6.2 can reproduce the analytical results developed in Chapter 5 to within an acceptable degree of accuracy. This conclusion will be further supported in Section 6.4, where I analyse the outcomes of N-ring simulations of 67 planet–disc models (Table H.1) and compare them with the expectations stemming from Chapter 5, e.g. in terms of the expected planetary precession rate, the location of the secular resonance, and the associated timescale for exciting planetesimal eccentricities.

6.2.3.3 Sensitivity to input parameters N and H

Apart from physical parameters describing the initial state of a system, such as the masses and orbital elements of the interacting bodies, the N-ring model employs two other key numerical parameters as inputs: namely, the number of disc rings N and their aspect ratio H. I thus checked how the results of the simulations performed in this Chapter (see Table H.1) would be affected if they were ran with values of N and H differing from their fiducial values (i.e., $N = 5000$ and $H = 0.1$, Section 6.2.1), with all else being kept the same. To this end, I followed the evolution of the fiducial planet–disc model (`Model A`, see Tables 5.1 and H.1) by running two sets of simulations which I describe below.

In the first set of simulations, I varied the number of rings representing the disc while holding their aspect ratio constant at $H = 0.1$. I generally found that varying N within the range $500 \leq N \leq 10000$ does not qualitatively affect the evolution of the planet–disc system. Nevertheless, with a closer look, I found that there are

some quantitative differences towards the smallest end of N with, for instance, the resonance location shifting inwards at most by only a few percent ($\lesssim 2\%$) compared to the nominal location. No significant differences were observed, on the other hand, for $N \gtrsim 2000$, indicating that the solutions obtained with the fiducial value of N can be considered as converged for the adopted value of H. I also checked that this is true regardless of whether the disc's non-axisymmetric gravity is included or not. In retrospect, this behaviour is consistent with the analysis presented in Chapter 4 about the e.g. convergence of the disc-induced secular effects in discretised softened discs to the expected results in continuous discs – see Section 4.6.3 and Figure 4.10.

In the second set of simulations, I varied the aspect ratio of the disc rings within the range $0.05 \leq H \leq 0.2$, while holding the number of rings constant at $N = 5000$. In this case, I found no significant qualitative changes in the evolution of the system. However, some quantitative differences were observed in the evolution of planetesimal orbits in the outer parts of the disc, i.e., away from the secular resonance, in terms of their eccentricity oscillations and precession periods. Running several test simulations with and without accounting for the non-axisymmetric component of the disc gravity, I learned that the differences largely arise due to the non-axisymmetric component of the disc self-gravity. In retrospect, this can be expected from the results of Chapter 4. Indeed, looking at Figure 4.2(A), one can see that the axisymmetric component of the softened gravitational potential (with the prescription of Hahn (2003)) matches the expected un-softened results for all values of $H \lesssim 0.2$. However, as can be seen from Figure 4.2(B), this is not the case for the non-axisymmetric component, which fully converges only for $H \lesssim 0.01$.

In summary, while the results of the N-ring model could be sensitive to the adopted number of disc rings and their aspect ratio, they do not vary strongly provided that $H \lesssim 0.1$ and $N \gtrsim 2000$, justifying my specific fiducial choice of their values (i.e., $H = 0.1$ and $N = 5000$; Section 6.2.1). Note that, in principle, I could have adopted a softening value of $H \lesssim 0.01$; however, this would have necessitated a much larger value of N than that adopted, making the N-ring model computationally expensive and slow to operate. Finally, it is important to note that the requirements discussed here strictly apply to the softening prescription of Hahn (2003) and discs with initial surface density slope of $p = 1$, and thus might differ for other values of p and softening prescriptions – see e.g. Figures 4.2 and 4.10.

6.2.3.4 Solar System Dynamics: Jupiter and Saturn

Finally, as already mentioned in Section 6.2, the softened N-ring model represents the continuum version of the classical Laplace–Lagrange theory (Murray & Dermott, 1999). Thus, as a natural check, I also tested the ability of my code to accurately reproduce the secular evolution of systems harbouring planets, but not discs. To this end, I adopted the example of Murray & Dermott (1999) on the evolution of Jupiter and Saturn in the Solar System (see their Section 7.3), and found that my code reproduces the system's evolution to the same level of accuracy reported in Murray & Dermott (1999) (e.g. their figure 7.1), provided the softening parameter is set to zero, $H = 0$, and the mean motion is defined as $n_j = \sqrt{GM_c/a_j^3}$. I also checked that the N-ring model can reproduce the evolution of test-particles perturbed by Jupiter and Saturn (Section 7.5 in Murray & Dermott, 1999) by modelling the test-particles as unsoftened, massless rings, finding that my code accurately reproduces e.g. the forced eccentricities of the test-particles as a function of semimajor axis (e.g. figure 7.5 in Murray & Dermott, 1999). However, I found that stringent conservation of the system's angular momentum requires defining the mean-motion as $n_j = \sqrt{G(M_c + m_j)/a_j^3}$, rather than $n_j = \sqrt{GM_c/a_j^3}$ as done in Murray & Dermott (1999) – in agreement with previous results on the same subject (Hahn, 2003).

6.3 Results: Nominal N-ring Simulations

Having outlined the details of the softened N-ring model, I now begin my investigation of the secular evolution of planet–debris disc systems. I start with the case I refer to as 'nominal' simulations: that is, with the non-axisymmetric component of the disc rings operating only on the planet, but not onto themselves (Section 6.2.2.3). My specific aim here is to analyse the dynamical effects of introducing this component on the evolution of the planet–disc system with an eye on its potential consequences on the gap-forming mechanism presented in Chapter 5.

To this end, I ran a set of 67 nominal N-ring simulations using the planet and disc parameters identified in Chapter 5 as capable of reproducing the observed depletion in the HD 107146 disc (Section 5.5). To allow for some generality, however, I chose the combinations of the planet and disc masses, as well as the planetary semimajor axes, from both within and outside the allowed (i.e., unshaded) portion of the parameter space portrayed in Figure 5.8. The parameters of the chosen models together with the simulation times are listed in Table H.1. In all simulations, the ring representing the

planet was initiated with $\varpi_p(0) = 0$, while the $N = 5000$ disc rings were initiated with $e(0) = 10^{-4}$ and $\varpi(0) = -\pi/2$; see Section 6.2.1 and Appendix H for more details.

Despite the broad range of adopted planet–disc parameters, I found that the evolution of all systems followed the same qualitative behaviour. Thus, to facilitate the interpretation of the simulations results and how they compare with those of Chapter 5, here I present the results obtained for `Model A` – which, I remind, was the fiducial configuration considered in Chapter 5 – focusing first on the orbital evolution of the planet and the planetesimals (Section 6.3.1), and then on the evolution of the disc morphology (Section 6.3.2).

6.3.1 Evolution of Planetesimal and Planetary Orbits

The nominal N-ring simulations differ from the analytical calculations of Chapter 5 by the introduction of the non-axisymmetric torque that the debris disc exerts on the planetary orbit. For this reason, I start by presenting results showing the orbital evolution of the planet. This will also aid in interpreting many of the dynamical features of the planetesimal evolution described later in Section 6.3.1.2.

6.3.1.1 Planetary Evolution

Figure 6.2 summarises the evolution of the planetary orbit throughout the simulation of `Model A`: the behaviour of its eccentricity e_p and apsidal angle ϖ_p as a function of time are shown in the left and right panels, respectively; see the blue curves therein. For ease of comparison with the results of Chapter 5, I have also plotted the corresponding results when the disc's non-axisymmetric gravity is switched off (i.e., by performing a 'simplified' N-ring simulation); see the red curves in Figure 6.2.

A striking feature in Figure 6.2(A) is the behaviour of the planetary eccentricity which, instead of remaining constant in time as in Chapter 5, undergoes a long-term decline. Indeed, one can see that in the course of the evolution, the planetary orbit circularises significantly, with its eccentricity decreasing from the initial value of 0.05 to ≈ 0.013 by the time that the simulation is stopped at $t = \tau \approx 332$ Myr, i.e., by a factor of ≈ 4. Second, it is also notable that this long-term decline is accompanied with additional small amplitude oscillatory behaviour; see the inset in Figure 6.2(A). Third, and more importantly, the decay of the planetary eccentricity seems to follow an exponential behaviour rather closely at all times (note that Figure 6.2(a) is a semi-log plot). Indeed, I find that the eccentricity decline can be perfectly fitted by the exponential model $e_p(t) \propto \exp(-Dt/2)$, with $D \approx 7.7 \times 10^{-3}\text{Myr}^{-1}$. Here, it is

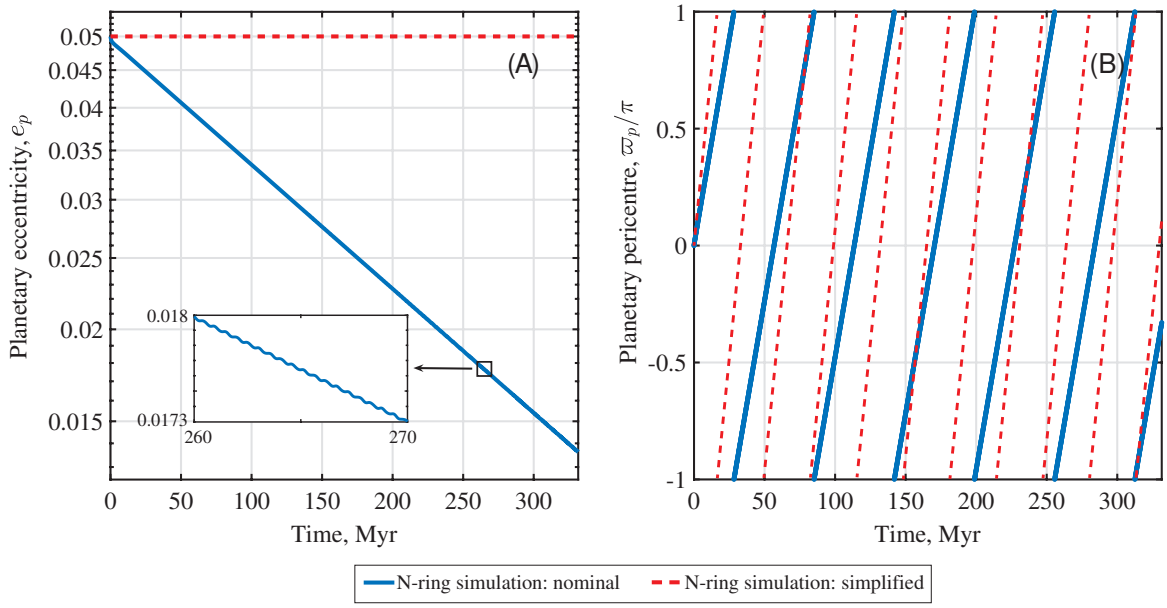

Figure 6.2. The evolution of the planetary eccentricity e_p (panel A) and longitude of pericentre ϖ_p (panel B) throughout the nominal N-ring simulation of `Model A` (Table H.1). For reference, the results obtained within the simplified simulation of the same planet–disc system are shown in red dashed lines in each panel. One can see that in the nominal simulations, the planetary eccentricity does not remain constant, but rather decays over time exponentially. The decay is also accompanied by small-amplitude oscillations (see the inset in panel A). It is also evident that the planet precesses at a slower rate in the nominal simulation compared to the simplified one. See the text (Section 6.3.1.1) for more details.

worthwhile to note that for this particular simulation, the maximum fractional change in the system's total angular momentum deficit is on the order of $|\Delta L_e/L_e| \sim 10^{-8}$; see Section 6.2.3.1. Thus, the decline of planetary eccentricity reported in Figure 6.2(A) is a real effect, and not due to e.g. diffusion of numerical errors within the simulation. As a matter of fact, and as we shall see later in Section 6.4.2, the circularisation of the planetary orbit is a generic phenomenon that occurs as a result of a process known as "resonant friction" or "secular resonant damping" in the literature (Tremaine, 1998; Ward & Hahn, 1998a, 2000). This will be studied in details in Section 6.4.2.

Looking now at the right panel of Figure 6.2, one can see that the planetary orbit precesses at a constant rate in a prograde manner, as already expected from Chapter 5. What is interesting to note, however, is that the rate at which the planet completes its precession cycle is different from the predictions of Chapter 5. Indeed, instead of precessing at the rate of $\dot{\varpi}_p = A_{d,p}$ as given by Equation (5.7) (see the dashed red lines), it now seems to precess at a slower rate such that $\dot{\varpi}_p \approx 0.58 A_{d,p}$. Thus, when compared with Chapter 5, the precession period is extended from $T_{\rm sec} \equiv 2\pi/\dot{\varpi}_p \approx 33$ Myr to ≈ 57

Myr. This indicates that the axisymmetric and non-axisymmetric components of the disc gravity drive planetary precession in opposite senses, with the former inducing a prograde precession and the latter inducing a retrograde precession. I will characterise this behaviour in more detail later in Section 6.4.1.

In summary, Figure 6.2 shows that the planetary orbit not only precesses under the action of the disc gravity, but also circularises over time. As we shall see, this will have important consequences for the evolution of planetesimal orbits and the subsequent development of a gap within the debris disc.

6.3.1.2 Planetesimal Evolution

Having described the evolution of the planetary orbit, I now present results showing the evolution of the debris particles. Figure 6.3 shows snapshots of the disc rings' eccentricities e and longitudes of pericentre $\Delta\varpi$ (relative to that of the planet) as a function of their semimajor axes at different times, as indicated in each panel. The times t were chosen such that they correspond roughly to the same ratios of t/τ as in Figure 5.3 of Chapter 5, which, I remind, measures the time into the evolution relative to the time it takes for the eccentricity at the resonance to grow to unity. For reference, an animated version of Figure 6.3 is made available online. There are several notable features in this figure, which I discuss below.

To begin with, and as expected from Chapter 5, the evolution of the disc rings in the inner and outer parts of the disc proceeds differently. Indeed, looking closely at Figure 6.3, one can see that at semimajor axes of $a \lesssim 70$ au, the eccentricities of the disc rings are maximised when they are aligned with the planetary orbit, i.e., $\Delta\varpi = 0$, while at semimajor axes of $a \gtrsim 70$ au, their eccentricities are maximised when the rings are anti-aligned, i.e., $\Delta\varpi = \pi$. This is to be expected since in the reported simulation, the disc's non-axisymmetric gravity acts only on the planet and thus does not modify the free precession rates of the disc particles when compared to Chapter 5 (e.g., see compare Figures 6.4 and 5.2). Consequently, the dynamics of disc particles remains planet- and disc-dominated at small and large distances from the planet, respectively (Section 5.3). This can also be seen by looking at Figure 6.4, where I show the radial profile of the free precession rate of the disc particles, together with that of the planet, as extracted from the diagonal of the matrix $\mathbf{A}$ (i.e., A_{jj} as given by Equation 6.3) corresponding to the simulated system (i.e., `Model A`). Looking at Figure 6.4, it is evident that $A \approx A_p$ at $a \lesssim 70$ au, and $A \approx -|A_d|$ at $a \gtrsim 70$ au. Additionally, the results of Figure 6.3 clearly show that the transition between the two regimes occurs

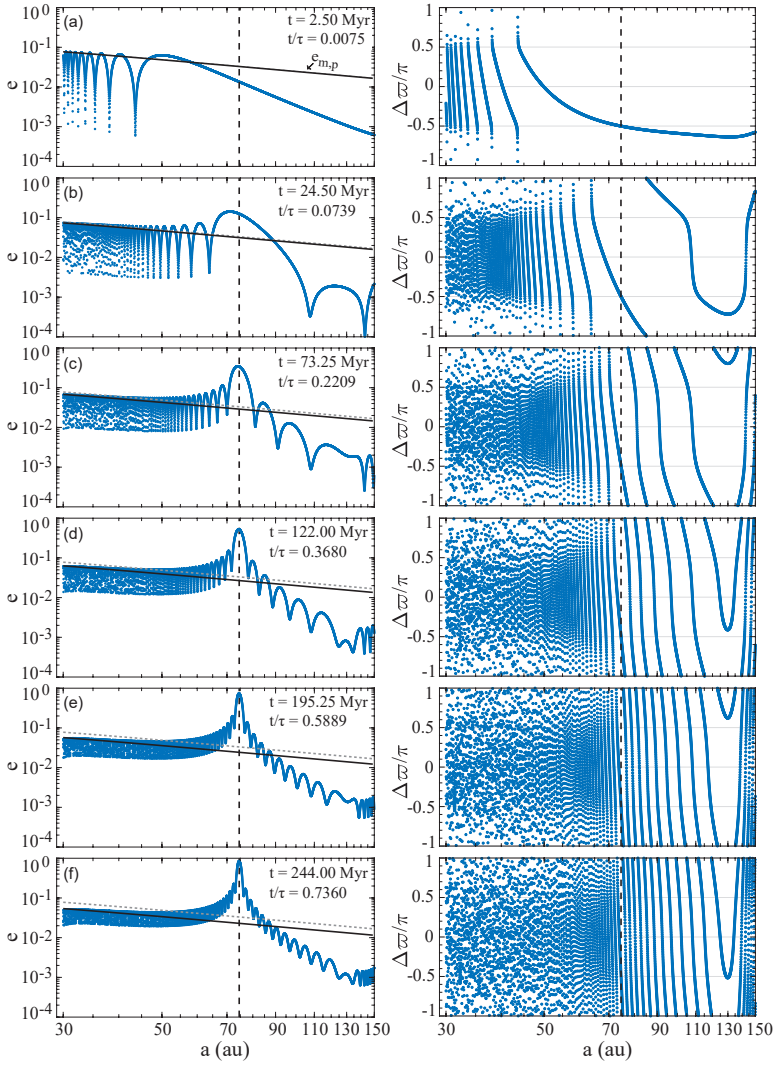

Figure 6.3. Snapshots of the planetesimal eccentricities e (left panels) and apsidal angles $\Delta\varpi$ (right panels, measured relative to that of the precessing planet) as a function of semimajor axis a. The snapshots are taken at $t = 2.5$, 24.5, 73.25, 122, 195.25, and 244 Myr (top to bottom) into the evolution of `Model A` (Table H.1) in its nominal N-ring simulation. The time is also indicated relative to $\tau \approx 332$ Myr, which represents the time at which $e \to 1$ at the site of the secular resonance. For reference, the secular resonance location ($a_{\rm res} \approx 75$ au) is shown by the dashed vertical lines. The solid black lines in the left panels show the maximum planetesimal eccentricities driven by the planet in the absence of the disc $e_{m,p}$ (see e.g. Equation 6.15). Note that the curve of $e_{m,p}(a)$ decreases over time and for reference, its initial value is shown by small-dashed grey lines (Equation 5.13). An animation of this figure is available online. See the text (Section 6.3.1.2) for more details.

via a secular resonance, where the eccentricity of the disc ring grows in time until it reaches unity, $e(a_{\rm res}) \to 1$; as already expected from Chapter 5.

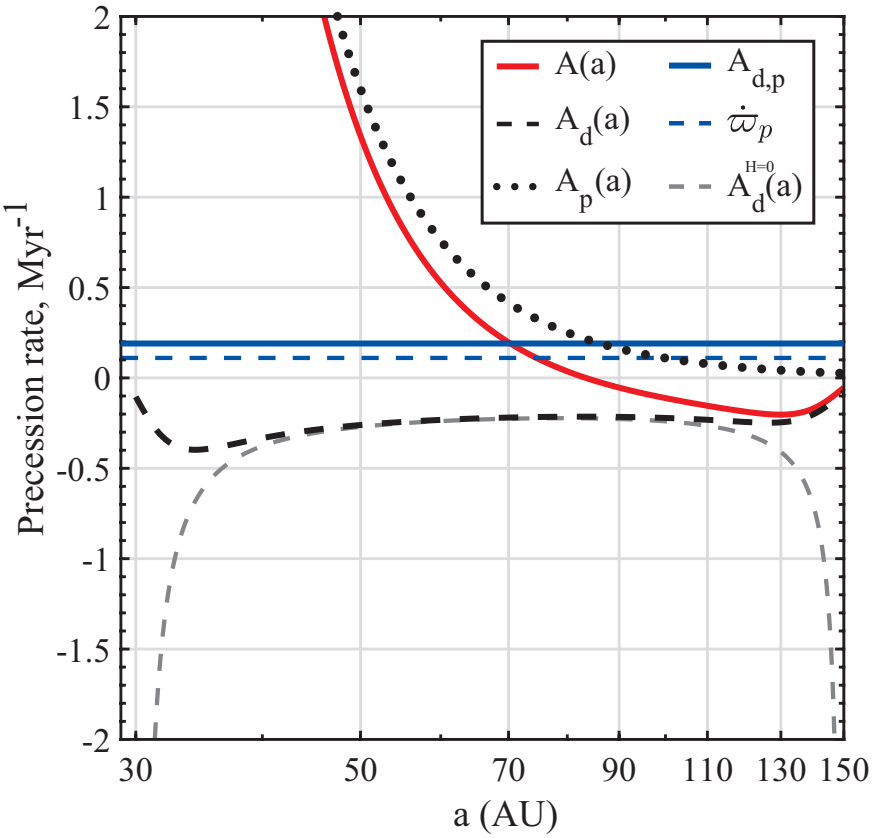

Figure 6.4. Planetesimal free precession rate $A = A_d + A_p$ due to both the planet and the disc as a function of semimajor axis (red curve). Calculations are done using the softened N-ring model described in Section 6.2 for the fiducial planet–disc model with $N = 5000$ and $H = 0.1$ (`Model A`, Table H.1). The dotted and dashed black curves represent $A_p(a)$ and $A_d(a)$, respectively. The solid and dashed blue lines represent the free and total precession rate of the planet due to the disc, $A_{d,p}$ and $\dot{\varpi}_p$, respectively. This figure must be understood as the softened analogue of Figure 5.2 of Chapter 5: for ease of comparison, the curve representing the unsoftened version of $A_d(a)$ is shown in dashed grey line and labelled as $A_d^{\mathrm{H}=0}(a)$. Note that $A(a) = A_{d,p}$ at 70 au, and $A(a) = \dot{\varpi}_p$ at ≈ 75 au. See the text (Section 6.3.1) for details.

This said, however, there are notable differences in terms of the orbital evolution of the disc particles between the results shown here and those of Chapter 5, both away from and at the resonance location; e.g. see and compare Figures 6.3 and 5.3. First, looking at the right column of Figure 6.3, one can see that soon after the evolution starts, the apsidal angles $\Delta\varpi$ of the disc rings, both in the inner and outer disc parts, span the entire range $[-\pi, \pi]$ over time. This is in contrast with the results of Chapter 5, where I found that $\Delta\varpi$ remain confined at all times within the ranges $[-\pi/2, \pi/2]$ and $\pm[\pi/2, \pi]$ in the inner and outer parts, respectively. Nevertheless, with a closer look at Figure 6.3, one can see that this expectation does actually hold true for the

majority of disc rings as long as not much time has elapsed from the beginning of the simulation, i.e., $t/\tau \lesssim 0.1$ or so, after which more and more disc rings populate the entire $[-\pi, \pi]$ range at all semimajor axes; see e.g. Figures 6.3(a),(b) and Figure 6.5.

Second, looking at the left column of Figure 6.3, it is evident that the maximum amplitude of the eccentricity oscillations of all disc rings does not remain constant in time, but rather undergoes a slow decline – see also the animated version of this figure. Indeed, one can see that the maximum eccentricities decline by about a factor of 2 at all semimajor axes, although it is a bit difficult to discern this effect in the outer parts of the disc due to the small eccentricities in that region. Additionally, I find that this decline occurs roughly over the same timescale at all semimajor axes, and that it is not accompanied with a change in the periods of the associated oscillations. This can be seen more clearly in Figure 6.5, where I plot the time evolution of the eccentricities and apsidal angles of disc rings at five different semimajor axes. This figure can also be compared with Figure 6.1, which depicts the orbital evolution of the disc rings in the negligence of the disc's non-axisymmetric gravity perturbations.

Apart from the features discussed above, the results depicted in Figure 6.3 indicate that while a secular resonance is established within the disc as expected from Chapter 5, its location is slightly different than anticipated. Indeed, one can see that the resonance occurs at ≈ 75 au rather than at 70 au – see the dashed vertical lines in Figure 6.3 – where the eccentricity grows to unity over time. Note also that at the resonance, the apsidal angle remains fixed at $-\pi/2$ throughout the simulation, in line with the results of Chapter 5. Additionally and interestingly, Figure 6.3 shows that the growth of the eccentricity at the resonance does not occur linearly in time, as was expected in Chapter 5. Instead, it displays a linear growth phase at early times, i.e. $t/\tau \lesssim 0.1$, and then smoothly becomes slower, resembling more of a quadratic curve: this can be seen more clearly in Figure 6.5(b). The change in the growth rate is significant, in the sense that it extends the time needed for the eccentricity at the resonance to grow to unity by a factor of ≈ 2.5 relative to the expectations from Chapter 5: namely, from $\tau \approx 135$ Myr to $\tau \approx 332$ Myr (e.g., see Figure 6.5(b) and the animated version of Figure 6.3)

Finally, note that the simulation results show no evidence of a secular resonance at $\approx a_{\rm in}$ (apart from the one at ≈ 75 au), which was found in Chapter 5. This is because the free precession rate $A_d(a)$ of the disc rings resulting from the disc gravity converges to a finite value as the edges of the disc are approached, i.e., $a \to a_{\rm in},\ a_{\rm out}$, rather than diverging as in Chapter 5; see Figure 6.4. As already explained in Chapters 4 and 5, this is a direct consequence of modelling the disc with a small but non-zero thickness (or softening parameter), as done in this chapter (Section 6.2). This also

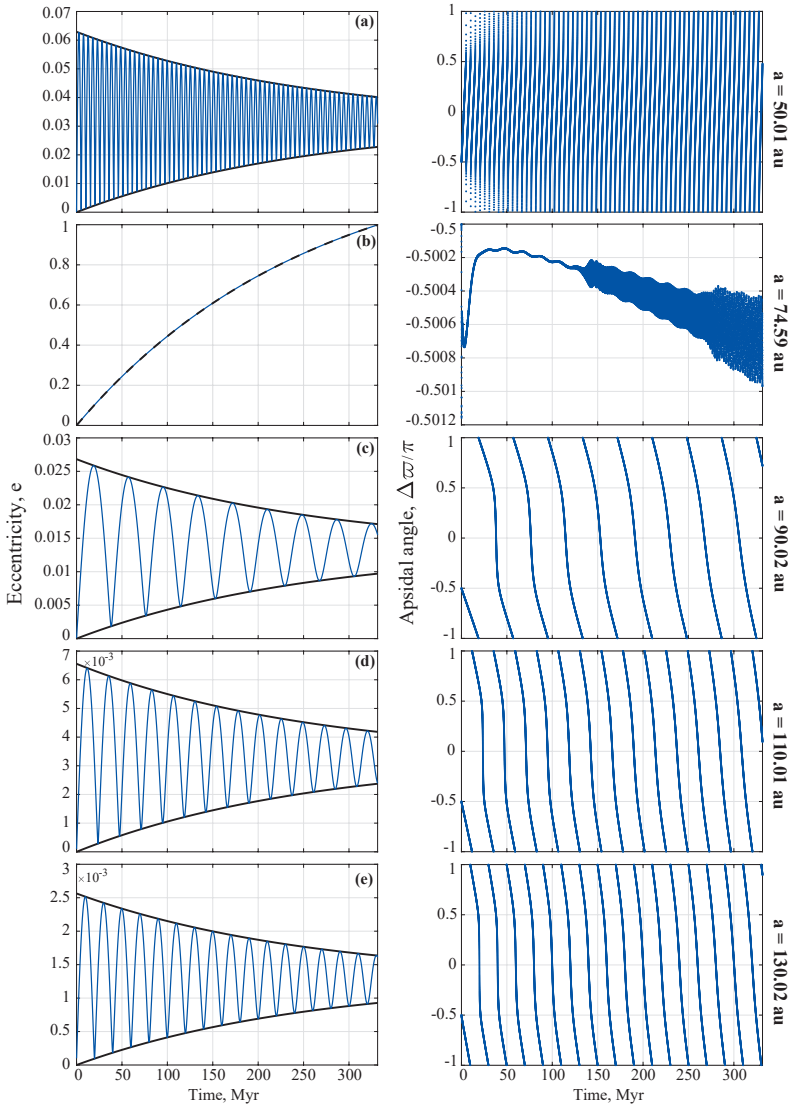

Figure 6.5. The time evolution of the planetesimal eccentricities e (left panels) and apsidal angles $\Delta\varpi$ (right panels, relative to the planet) at five different semimajor axis, as extracted from the nominal N-ring simulation shown in Figure 6.3 (blue curves). The values of the probed semimajor axes are indicated on the right side of the figure. The envelope of the eccentricity oscillations shown in black curves are obtained using Equation (I.3) of Appendix I (see also Section 6.3.1.3). Note that at the resonance $\Delta\varpi \approx -\pi/2$ at all times (panel b). It is also evident that the eccentricities at the resonance grow following a quadratic curve, which is well reproduced by Equation (6.16) (dashed black curve in panel b). This figure can be compared to Figure 6.1, which portrays the results of the same planet–disc model (`Model A`, Table H.1) in the simplified simulations. See the text (Sections 6.3.1.2 and 6.3.1.3) for more details.

explains as to why the apsidal angles $\Delta\varpi(a)$ in the right-hand panels of Figure 6.3 have a positive slope within the range of 130 au $\lesssim a \lesssim a_{\rm out}$: the convergence of $A_d(a)$ to a finite value as $a \to a_{\rm out}$ renders the total free precession rate $A(a) \approx A_d(a)$ in that region an increasing function of semimajor axis (rather than decreasing as in Chapter 5); see e.g. Figure 6.4.

6.3.1.3 The Coupling between the Planet and Planetesimals

Before moving on to present results showing the evolution of the disc morphology, let us pause here and try to understand the physics behind the behaviour of planetesimal evolution (Figure 6.3). Obviously, planetesimal dynamics is affected only indirectly by the introduction of the disc's non-axisymmetric torque on the planet, and thus the behaviour of planetesimal dynamics in Figure 6.3 should, in principle, be understood as a result of the coupling between the planet and the planetesimals.

Let us for a moment ignore the decay of planet's eccentricity, and focus on the consequences of its precessional behaviour which, according to Figure 6.2, is slowed down to a rate below $A_{d,p}$, i.e., $\dot{\varpi}_p \lesssim A_{d,p}$. For reference, the value of $\dot{\varpi}_p$ as extracted from the simulation of `Model A` is overplotted in Figure 6.4, see the blue dashed line therein. Looking at this figure, one can see that while $A(a) = A_{d,p}$ at $a = 70$ au as expected from Chapter 5, one has $A(a) = \dot{\varpi}_p$ at around $a \approx 75$ au, which coincides with the exact location of the secular resonance; see Figure 6.3. This suggests that the shift in the resonance location could be the consequence of the slower precession rate of the planetary orbit. I will further test and confirm this hypothesis in Section 6.4.3.

Now, let us consider the effects of the decaying planetary eccentricity on the planetesimal dynamics. As can be expected, this effect renders the secular gravitational potential of the planet time-dependent, even when evaluated in a frame co-rotating with the planet: this is because in the notation of Chapter 5 (see Section 5.2.2.3), the term $B_p \propto e_p$ will now be a function of time, complicating the dynamics. To gain additional insights, in Appendix I I derive a full time-dependent solution for the planetesimal eccentricities $e(t)$ and apsidal angles $\Delta\varpi(t)$ in the presence of a circularising planet in a frame co-precessing with the planet; see Equations (I.3)–(I.6). The main takeaway from this calculation is that the decay of e_p renders the forced component of the planetesimal eccentricity a time-dependent function:

$$e_{\rm forced,decay}(a,t) = e_{\rm forced}(a) \times \exp(-Dt/2), \tag{6.15}$$

where $e_{\rm forced}(a)$ is given by Equation (5.12) and must now be understood as evaluated at $e_p(0)$, while the free eccentricity – which, I remind, is set by the initial conditions (e.g. Section 3.2.4.1) – remains the same as in Chapter 5. As a result, and unlike in Chapter 5, the eccentricity does not oscillate between its initial value of 0 and $e_m = 2|e_{\rm forced}(a)|$. Instead, the eccentricity now oscillates with a decreasing amplitude while at the same the minimum attained in the course of oscillations increases over time; as can be seen in Figures 6.3 6.5. Eventually, as $t \rightarrow \infty$ and $e_{\rm forced,decay}(a) \rightarrow 0$ (Equation 6.15), eccentricities converge to the free eccentricity. This behaviour is well captured by Equation (I.3) of Appendix I which shows that the minimum and maximum eccentricity during the oscillations change in time following $\propto 1 - \exp(-Dt/2)$ and $\propto 1 + \exp(-Dt/2)$, respectively – see the black curves in the left panels of Figure 6.5. For reference, the black lines in Figure 6.3 show the maximum eccentricity $e_{m,p}$ predicted by Equation (I.3) upon neglecting A_d and $\dot{\varpi}_p$, which is valid for the inner disc parts where the dynamics is planet-dominated. One can see a very good agreement between the simulations results and the curve of $e_{m,p}$ (see also the animated version of Figure 6.3). Equation (I.3) also explains the behaviour of eccentricity growth at the resonance: taking the limit of $A = A_d + A_p \rightarrow \dot{\varpi}_p$ in Equation (I.3), one finds that

$$e(t) = \frac{2|B_p(0)|}{D}\Big[1 - \exp(-Dt/2)\Big] \approx |B_p(0)|t\Big(1 - \frac{1}{4}Dt\Big), \tag{6.16}$$

where the approximation assumes that $Dt \approx 0$, and all quantities are evaluated at the resonance location $a_{\rm res}$. For reference, Equation (6.16) as evaluated at the values of D and $a_{\rm res}$ extracted from the simulation of `Model A` is plotted in Figure 6.5(b): one can see the perfect agreement between the numerical results and Equation (6.16).

Finally, note that this analysis also clarifies as to why the disc particles evolve to populate the entire range of $\Delta\varpi = [-\pi, \pi]$ (e.g. see the left panels of Figures 6.3 and 6.3). Without loss of generality, let us consider the disc particles in the inner disc parts: with their orbits being initially circular, their eccentricities start at the origin of the $e(\cos\Delta\varpi, \sin\Delta\varpi)$ plane and precess counter-clockwise in a circle (since $A(a) \approx A_p > 0$) around the forced eccentricity vector (see e.g. Figure 3.3 and Section 3.2.4.1). Given that $e_{\rm forced} > 0$ in this case, the circle would initially be restricted to the plane containing the positive x-axis; meaning that $\Delta\varpi$ can only acquire values between $-\pi/2$ and $+\pi/2$. As time progresses and the forced eccentricity decays, however, the centre of the circle slowly shifts towards the origin of the $e(\cos\Delta\varpi, \sin\Delta\varpi)$ plane, causing the circle to cross through all quadrants of the plane and so allowing $\Delta\varpi$ to be within the entire range of $[-\pi, \pi]$. This 'thought experiment' also explains as to why

the minimum of the eccentricity oscillations grows to values larger than 0. A similar argument can be applied to the outer disc parts, explaining the results shown in the left-hand panels of Figure 6.3.

6.3.2 Evolution of the Disc Morphology

In the previous section, I described the orbital evolution of the planetesimal and planetary orbits in the fiducial configuration (i.e., `Model A`, Table H.1). I now move on to present results showing the evolution of the disc surface density. To this end, I convert the orbital element distributions of planetesimals shown in Figure 6.3 into surface density distributions following the same technique adopted in Chapter 5; see Appendix G. The resulting two-dimensional maps of the (normalised) disc surface density, Σ, as well the radial profiles of the azimuthally averaged surface density, $\langle\Sigma(r)\rangle$, are shown in Figures 6.6 and 6.7, respectively. For reference, in Figure 6.6 I also show the planetary orbit and its pericentre position which, for the considered model, precesses with a period of $\tau_{\rm sec} \approx 57$ Myr (Figure 6.2). The six snapshots shown in both Figures 6.6 and 6.7 correspond to the same times as in Figure 6.3, and their animated versions covering the entire simulation – i.e., from $t = 0$ to $t = \tau \approx 332$ Myr – are made available online. These figures can be compared to Figures 5.9 and 5.10 of Section 5.6.1 (Chapter 5).

A quick look at Figures 6.6 and 6.7 reveals that by and large, the disc morphology evolves following the same three stages identified in Chapter 5 (see Section 5.6.1), although with subtle differences mainly in terms of the gap's morphology at late times, i.e., as $t \to \tau$. For completeness, however, I briefly describe these three stages below, pointing out the differences in the disc morphology when compared to the description provided in Chapter 5.

Stage 1 ($0 \leq t \lesssim \tau_{\rm sec}$): At early times, and similar to Chapter 5, the planet launches a trailing spiral wave at the inner edge of the disc $a_{\rm in}$ which propagates outwards with time while wrapping 360° around the star; see panels (a) and (b) in Figures 6.6 and 6.7. Once again, the speed at which the spiral moves through the disc gets slower as it extends to larger radii due to the fact that $A(a)$ is a decreasing function of semimajor axis (Figure 6.4). Compared to Section 5.6.1, however, the main difference is that the spiral extends out to about a radius of ≈ 80 au rather than ≈ 70 au; e.g., see and compare Figures 6.6(b) and 5.9(b), as well as Figures 6.7(b) and 5.10(b). At the same time, the planet's eccentricity decreases from its initial value of $e_p(0) = 0.05$ to ≈ 0.04, i.e., by about $\approx 20\%$ (Figure 6.2). This, however, does not affect the spatial appearance of the disc: for instance, and similar to Chapter 5, the

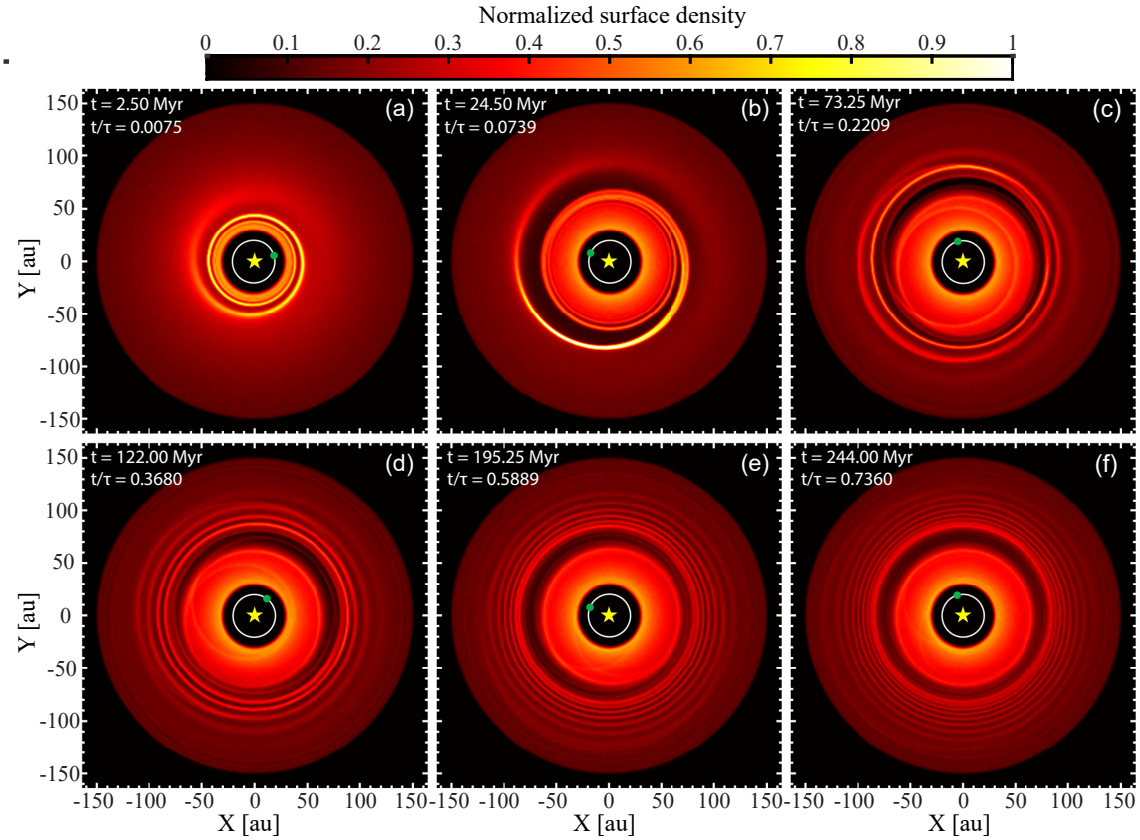

Figure 6.6. Series of two-dimensional snapshots showing the evolution of the (normalised) disc surface density Σ in `Model A` (Table H.1). The results shown are obtained by making use of the dynamical state of planetesimals in the nominal N-ring simulations of Figure 6.3 (see Appendix G for the details of the procedure). The snapshots correspond to the same moments of time t as in Figure 6.3, and are indicated in each panel for reference. The time is also indicated relative to $\tau \approx 332$ Myr. All panels have 400×400 pixels and share the same surface density scale (and normalisation constant) as shown in the colour bar. The stellar position is marked by the yellow star in each panel. The planet's orbit and its pericentre position are shown by the white solid line and green circle, respectively. Note the launching of a spiral arm at the inner disc edge at early times (panel a), its outward propagation in time (panel b), and the eventual sculpting of a gap centred around the secular resonance, $a_{\rm res} \approx 75$ au (panels c–f). An animation of this figure is available online. See the text (Section 6.3.2) for more details.

surface density distribution in the inner parts of the disc looks roughly axisymmetric. This is because the spirals in the this region become so tightly wound together as planetesimal orbits become phase-mixed over time, i.e., $\Delta\varpi(a)$ spans the range $[-\pi, \pi]$ – although with a higher concentration within $[-\pi/2, \pi/2]$, the same range as in Chapter 5 (see panels a and b in Figure 6.3).

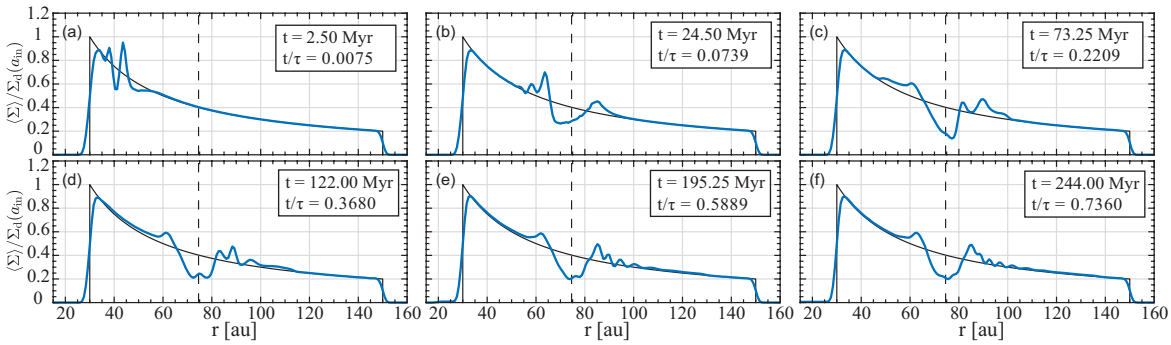

Figure 6.7. The azimuthally averaged surface density of the disc $\langle\Sigma\rangle$ as a function of radial distance r from the central star (solid blue lines). Each panel corresponds to each of the snapshots of `Model A` (Table H.1) shown in Figure 6.6. The results are obtained by splitting the disc into 200 annular bins (Appendix G), and are normalised with respect to the initial disc surface density $\Sigma_d(a)$ (Equation 5.1, with $p = 1$) evaluated at the inner disc edge, $a = a_{\rm in}$. The solid black lines show the normalised profile of the initial $\Sigma_d(a)$, for reference. Note the appearance of a clear depletion in the surface density around the location of the secular resonance ($a_{\rm res} \approx 75$ au, vertical dashed lines). See the text (Section 6.3.2) for more details. An animation of this figure is available online.

Stage 2 ($t \sim \tau_{\rm sec}$): By the time the planet has completed approximately one precession cycle, a gap forms around the location of the secular resonance – similar to Chapter 5; see Figures 6.6(c), 6.7(c) and their animated versions. Now, however, the gap is centred around ≈ 75 au (rather than ≈ 70 au), which represents the location of the secular resonance (Figure 6.3). At this stage, the features of the gap are similar to that in Chapter 5, in that it is crescent-shaped pointing in the direction of the planet's pericentre. Within the gap, in an azimuthally averaged sense, about a half of the initial surface density is depleted. The width of the gap, however, is slightly narrower than that in Section 5.6.1, with a value of ~ 15 au (compared to ~ 20 au); e.g. compare Figures 6.7(c) and 5.10(c). The shape of the gap can be understood using the same reasoning in Chapter 5: by this time, the planetesimals interior to the gap have formed a coherent eccentric structure that is apsidally aligned with the planet and offset relative to the star to roughly the same degree as in Section 5.6.1. This is because despite the decay of the planetary eccentricity, the maximum amplitudes of eccentricity oscillations in the inner disc parts are still comparable to those in Section 5.6.1; see Figure 6.3(c).

Stage 3 ($\tau_{\rm sec} \lesssim t \lesssim \tau$): At times where at least one secular period has elapsed for particles at the resonance (i.e., $t \gtrsim \tau_{\rm sec}$), the gap continues to precess while maintaining its alignment and coherence with the planetary longitude of pericentre; see panels (d)–(f) in Figures 6.6 and 6.7. As this happens, and similar to Chapter 5, the gap's

depth and width practically remain invariant: indeed, in a time- and azimuthally averaged sense, I find that $w_g \approx 15.45 \pm 0.21$ au and $d_g \approx 51 \pm 2\%$ (relative to the initial density at $a_{\rm res}$). Additionally, and as in Chapter 5, the disc part exterior to the gap develops a spiral pattern, first near the gap and then at larger radii as the system evolves further: this can be explained by the same arguments provided in Section 5.6.1 – see Stage 3 therein – despite the fact that now the apsidal angles $\Delta\varpi(a)$ span the entire range of $[-\pi, \pi]$ (Figure 6.3). What is interesting to note, however, is that further into the evolution, i.e., as $t \to \tau$, the gap starts to slowly evolve away from its crescent 'arc' shape and develops more of an axisymmetric or circular structure, in the sense that the depletion becomes visible 360° around the star – see e.g. Figure 6.6(f) and its animated version. This is in stark contrast with the results of Chapter 5; e.g. see and compare Figures 6.6(d)–(f) and Figures 5.9(d)–(f). The transition from asymmetry to symmetry is not perfect though, in the sense that one can still discern that the gap is both wider and deeper toward the planetary pericentre, but only to a relatively small degree. This behaviour can be understood by noting that by the time $t \to \tau$, the planetary eccentricity undergoes a significant decay (relative to its initial value, see Figure 6.2(A)) and in turn forces the planetesimal eccentricities throughout the entire disc to decrease (e.g., Figures 6.3 and 6.5). This also causes the planetesimal orbits to be phase-mixed as described in Section 6.3.1.3, with $\Delta\varpi(a)$ spanning the entire range $[-\pi, \pi]$ (see Figure 6.3). As a result, the disc parts both interior and exterior to the gap – which have already settled into a lopsided, precessing coherent structure – become less offset relative to the star, individually, and thus in combination, decrease the asymmetry of the gap in between. This is easier to see in the region interior to the gap, where the eccentricities are naturally larger than in the outer parts. Here, it is also noteworthy to mention that the effect of the circularising planet on the gap asymmetry cannot be reproduced by a planet of constant but smaller eccentricity, e.g. as in `Model A-Loep` of Chapter 5; see Section 5.6.2.2 and Figure 5.12.

Before concluding this section, I remark that the results on the secular evolution of the planet and disc particles described in this section are general, in the sense that the same qualitative behaviour – both in terms of orbital evolution and disc morphology – is reproduced in all simulated planet–disc systems (Table H.1). Quantitatively speaking however, the results differ from one system to another depending on the parameters adopted, but the same scaling rules identified in Chapter 5 do generally apply (see Section 5.6.2). For instance, varying both the planet and disc masses simultaneously, i.e., so that $M_d/m_p = \text{const}$ (to ensure the resonance location does not change), will

only affect the secular evolution timescale, but not the details of the ensuing dynamics – see also Section 5.6.2.3. Increasing the planet's initial eccentricity will (i) shorten the evolution timescale; (ii) render the spirals more open and prominent; (iii) increase the width of the resulting gap; and vice versa – see also Section 5.6.2.2. Last but not least, increasing (decreasing) the planet's semimajor axis causes (i) the gap to be wider (narrower), see Section 5.6.2.1; (ii) its orbit to circularise more rapidly (slowly); and (iii) the shift in the resonance location – relative to that expected from Chapter 5 – to become larger (smaller). Finally, and as in Chapter 5, varying the planet–disc parameters does not significantly affect the gap depths. The scalings concerning the new effects uncovered by the simulations carried out for the purposes of this section is explored in details in the next section.

6.4 Analysis and Predictions

As pointed out in the previous section, there are several qualitative and quantitative differences between the results of this chapter and those described in Chapter 5 in terms of the evolution of the planet and the disc. In this section, I aim to characterise these differences in greater detail in order to better understand the underlying physics governing the secular planet–disc interaction. Wherever possible, I also provide quantitative explanations for the observed differences using dynamical theory.

To this end, I ran a second set of simulations using the N-ring model described in Section 6.2, but this time by setting the terms representing the non-axisymmetric perturbations due to the disc rings equal to zero. Such 'simplified' N-ring simulations (Section 6.2.2.3) are expected to accurately reproduce the analytical results of Chapter 5; see Section 6.2.3.2. I performed such simulations for all the 67 planet–disc models considered in this work (Table H.1), and compared the results with those obtained within the nominal set of simulations. As we shall see, this approach allows me not only to gain insights into how the various components of the disc potential affect the system's evolution, but also to make some analytical progress towards characterising the resultant effects. Below is a brief account of what I learned, supplemented by an appropriate figure, when called for.

6.4.1 Precession Rate of the Planetary Orbit

Numerical results of the previous section show that in the same planet–disc model (i.e., `Model A`, Table H.1), the planetary orbit precesses at a slower rate than that expected

based on the analytical results of Chapter 5; see Figure 6.2(B). Here, I investigate this behaviour in more detail, and provide an explanation for the observed differences. To do so, for each of the two sets of simulations, i.e., nominal and simplified, I extract the time evolution of the planet's longitude of pericentre $\varpi_p(t)$ after each simulation ends, and measure the rate $\dot{\varpi}_p$ at which it advances from the slope of the numerical data[2]. Additionally, since it is expected that the precession rates will scale linearly with masses (e.g. Section 5.6.2.3), I normalise the extracted slopes by the theoretical values of $A_{d,p}$ (Equation 5.7) corresponding to each system's model parameters (Table H.1). This essentially should render the results dependent only on the planetary semimajor axis, except if there is a dependency on a parameter not accounted for (e.g. disc and planet eccentricities).

Figure 6.8 shows the results obtained by applying the above procedure to each of the two sets of 67 simulations. There are several notable features in Figure 6.8. First, it shows that in the absence of the disc's non-axisymmetric effects on the planet (shown in red filled circles), the planetary orbit precesses at the expected rate of $A_{d,p}$ regardless of the system parameters (Section 5.2.2.2): indeed, one can see that $\dot{\varpi}_p/A_{d,p} \approx 1$ in all considered planet–disc models. This also further confirms the validity of the N-ring model, in addition to the tests presented in Section 6.2.3. Second, looking at Figure 6.8, one can see that when the disc's non-axisymmetric effects on the planet are included (shown in blue filled circles), a planet orbiting at a given semimajor axis precesses at a rate that is generally smaller than that anticipated in Chapter 5, i.e., $\dot{\varpi}_p/A_{d,p} \lesssim 1$, as already pointed out in Section 6.3.1.1. It is also clear that the differences between the two sets of simulations become more pronounced for planets orbiting closer to the disc inner edge than to the star. Indeed, increasing the planetary semimajor axis from $a_p/a_{\rm in} \approx 0.1$ to $a_p/a_{\rm in} \approx 1$, the differences grow from a factor of roughly 1 to a factor of ≈ 5, respectively. Another important feature in Figure 6.8 is that similar to the simplified simulations, the nominal simulations reveal little or no evidence of scatter in the results at any given value of a_p, despite the different initial conditions – see Table H.1. This indicates that while the disc's non-axisymmetric effects do affect the planetary precession rates, they do so in a way that is independent of the disc eccentricity (which naturally is imposed by the planet's eccentricity, Section 6.3).

The behaviour of the planetary precession rates in the nominal N-ring simulations can be explained as follows. According to Chapter 5, the planetary orbit precesses at a rate given by $\dot{\varpi}_p = A_{d,p}$ (Equation 5.7): however, strictly speaking, this is the *free*

[2]For all simulations, the curves of $\varpi_p(t)$ were well fitted with a straight line obtained by the least-squares method, with a correlation coefficient of ≈ 1.

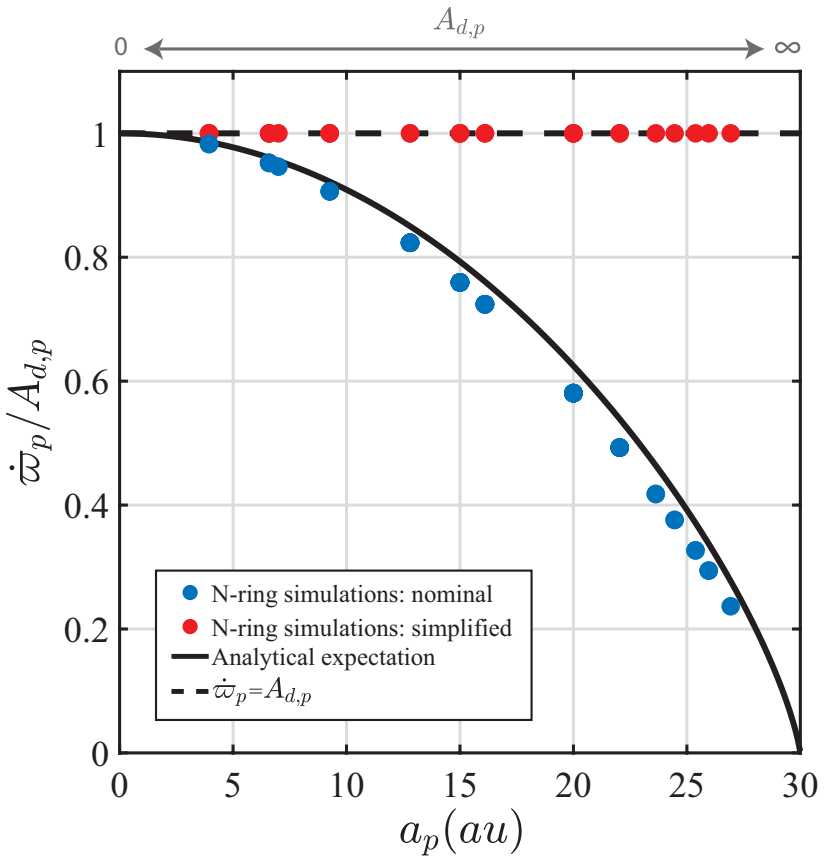

Figure 6.8. The planetary apsidal precession rates $\dot{\varpi}_p$ relative to the theoretical free precession rate $A_{d,p}$ (Equation 5.7) as a function of planetary semimajor axis a_p, as measured from N-ring simulations. Calculations are based on 67 different planet–disc models (Table H.1), each evolved numerically using the N-ring model within two sets of simulations: simplified (red circles) and nominal (blue circles). One can see that the inclusion of the disc's non-axisymmetric torque on the planet reduces the rate at which the planetary orbit precesses compared to the expectation when that torque is ignored. The reduction is clearly dependent on the planetary semimajor axis, and is well represented by Equation (6.17) which, for reference, is shown in black solid line. Note that the N-ring simulations accurately reproduce the theoretical expectations of Chapter 5, i.e., $\dot{\varpi}_p/A_{d,p} = 1$, when the disc's non-axisymmetric torque on the planet is ignored. See the text (Section 6.4.1) for more details.

precession rate, i.e., the rate at which the planet precesses if the disc potential were axisymmetric – which is what was assumed in Chapter 5. In reality however, there is also a contribution to the planet's precession rate due to the disc eccentricity, which manifests itself as a non-axisymmetric contribution to the disc potential. This is the *forced* precession rate, corresponding to the term $i\sum_{k\neq 0}^{N} A_{0k}\zeta_k$ in Equation (6.9) with $j = 0$. Thus, it is the combination of the free and forced contributions that dictates the planetary precession rate.

To quantify the effects of the above-mentioned physical mechanism and confirm its validity, in Appendix J I derive an analytical expression for $\dot{\varpi}_p/A_{d,p}$, accounting for both the free and forced precession rates induced by the disc. I find that within a set of reasonable assumptions, the planetary precession rate can be written as follows:

$$\begin{aligned}\frac{\dot{\varpi}_P}{A_{d,p}} &= 1 + \frac{1}{2}\frac{\phi_2}{\phi_1}\frac{b^{(2)}_{3/2}(a_p/a_{\rm in})}{b^{(1)}_{3/2}(a_p/a_{\rm in})}, \\ &\approx 1 - \frac{13}{16}\left(\frac{a_p}{a_{\rm in}}\right)^2.\end{aligned} \tag{6.17}$$

Here, the terms ϕ_1 and ϕ_2 govern the strengths of the axisymmetric and non-axisymmetric effects of the disc on the planet, respectively (Equations (E.5) – (E.8)), and the approximation is obtained for the fiducial disc model ($p = 1$, $\delta = 5$) in the limit of $a_p/a_{\rm in} \to 0$, assuming a disc eccentricity profile scaling as that forced by the planet (Equation 5.13), i.e., $e_d(a) \propto 1/a$ – see Appendix J for details and justification. First, Equation (6.17) shows that the precession rate is directly proportional to the coefficient ϕ_2, which is a proxy to the strength of the disc's non-axisymmetric torque. This explains why $\dot{\varpi}_p/A_{d,p} = 1$ in the simplified simulations (and in Chapter 5) which have $\phi_2 = 0$ by construction. Second, given that $\phi_2 < 0$ (Equation E.6), one can see from Equation (6.17) that the disc non-axisymmetry causes the planet to precess at a slower rate than when ignored. Note that the contribution depends only on the semimajor axis of the planet relative to the disc inner edge, $a_p/a_{\rm in}$, explaining why there is no scatter in terms of $\dot{\varpi}_p/A_{d,p}$ at a given $a_p/a_{\rm in}$ in Figure 6.8. Third, and more importantly, Equation (6.17) approximates the numerical solutions of the nominal N-ring simulations very well, even for relatively large values of a_p – see the black line in Figure 6.8.

6.4.2 Decay of Planetary Eccentricity and Resonant Friction

As noted in Section 6.3, the planetary eccentricity evolves significantly in the nominal N-ring simulations, undergoing a long-term exponential decline with additional small amplitude oscillatory behaviour in the meantime (Figure 6.2). This behaviour is in contrast to the case studied in Chapter 5, where the planet's eccentricity remains constant throughout the evolution. I now present an in-depth analysis of this phenomenon, providing evidence that it ensues from the so-called process of 'resonant friction' or 'secular resonant damping' (Tremaine, 1998; Ward & Hahn, 1998a, 2000).

Resonant friction is a secular process that results from the gravitational coupling of a planet and disc particles at and around the site of a secular apsidal resonance (see

Tremaine, 1998; Ward & Hahn, 1998a, 2000, for a detailed discussion). The excitation of planetesimal eccentricities at $a_{\rm res}$ gives rise to a strong torque that acts on the planet in a way that the system's angular momentum is rapidly redistributed between the planet and the disc, without affecting the system's angular momentum budget. Indeed, the torque exerted by the disc on the planet acts to transport angular momentum from the disc particles (mainly around $a_{\rm res}$) to the planet, in a way that damps the planet's eccentricity[3]. The rate at which resonant friction damps the planetary eccentricity is given by (e.g., see equation 20 in Tremaine, 1998):

$$\begin{aligned} \frac{1}{e_p^2}\frac{de_p^2}{dt} &= -\frac{\pi^2}{4} n_p \frac{m_p}{M_c}\frac{\Sigma_d(a)a^2}{M_c}\frac{n(a)}{|dA/d\log a|}\left(\frac{a_p}{a}\right)^3\left[b^{(2)}_{3/2}\left(\frac{a_p}{a}\right)\right]^2\Bigg|_{a=a_{\rm res}}, \\ &\equiv -D < 0, \end{aligned} \tag{6.18}$$

where as before $A(a) = A_d(a) + A_p(a)$. For a detailed derivation of Equation (6.18), the reader is referred to Tremaine (1998) and Ward & Hahn (1998a, 2000). Note that one has $D \to 0$ and thus $de_p/dt = 0$ when (i) there is no secular resonance within the disc, e.g. if the disc is considered to be massless; (ii) either the planet and/or the disc do not exert non-axisymmetric torques on each other, e.g. if either's potential is taken to be axisymmetric. It is also important to note that Equation (6.18) is derived by neglecting the terms representing the non-axisymmetric perturbations that planetesimals exert amongst themselves (see Tremaine, 1998; Ward & Hahn, 1998a, 2000) – similar to the setup of my nominal N-ring simulations. The solution of Equation (6.18) is a simple exponential function of time, so that $e_p(t)$ decays

$$e_p(t) = e_p(0)\,\exp\left(-Dt/2\right), \tag{6.19}$$

with a characteristic half-life of $\tau_D \equiv 2\log 2/D$.

In order to ascertain that the decay of planetary eccentricities observed in my simulations is indeed due to resonant friction, I performed the following exercise. At the end of each simulation, I extracted the evolution of $e_p(t)$ from each of the 67 nominal N-ring simulations (Table H.1) up until the time when planetesimals at the secular resonance attained an eccentricity of unity, or lower if they did not have sufficient time to achieve that throughout the simulation time, i.e., $e(t) \lesssim 1$; see Table H.1. Then, I fitted the numerical results with the functional form of Equation (6.19), while keeping

[3]Formally, this process is somewhat akin to the interactions of planets and discs at co-orbital Lindblad resonances, which also act to damp the planet's eccentricity (e.g. Goldreich & Sari, 2003). This is because apsidal resonances are a special type of Lindblad resonances, but with pattern speed equal to the apsidal precession rate of the planet – rather than its mean motion.

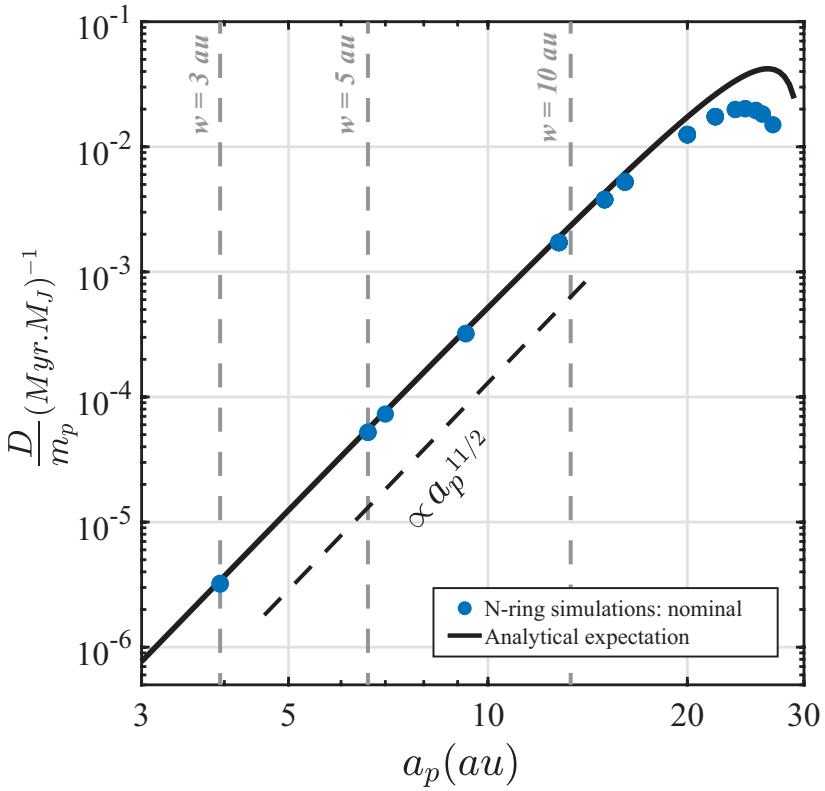

Figure 6.9. The rates of exponential decay of planetary eccentricities per unit mass of the planet D/m_p as a function of its semimajor axis a_p. The blue filled circles represent the numerical data extracted from N-ring simulations of 67 different planet–disc models (Table H.1). The black line represents the theoretical expectation (Equation 6.18), evaluated at the system parameters of interest. The dashed black line shows the scaling of D/m_p with a_p (Equation 6.20). The grey dashed lines mark the planetary semimajor axes for which the expected resonance widths are $w = 3$, 5, and 10 au (based on Chapter 5, see Equation 5.21). See the text (Section 6.4.2) for details.

both $e_p(0)$ and D as free parameters. In each case, the fitting procedure yielded an output $e_p(0)$ agreeing with the input value for the simulations to within a difference of less than 1%. The results obtained for the decay rate are shown in Figure 6.9, where the fitted values of D are plotted per unit mass of the planet as a function of its semimajor axis. For ease of comparison, Figure 6.9 also shows in black solid curve the results of evaluating Equation (6.18) at the system parameters of interest (Table H.1), i.e., the combinations of M_d/m_p and $a_p/a_{\rm in}$ which according to Chapter 5 place a secular resonance at $a_{\rm res} = 70$ au.

Looking at Figure 6.9, one can see that the simulation results agree very well with the theoretical expectations for all considered values of a_p. The agreement is perfect at all values of $a_p/a_{\rm in} \lesssim 0.5$, beyond which the numerical results tend to be smaller

than the theoretical expectations by at most a factor of ≈ 2. The reason for this slight discrepancy as $a_p \to a_{\rm in}$ is likely due to the fact that the derivation of Equation (6.18) assumes that as $t \to \infty$ and $e(a_{\rm res}) \to 1$, all of the resonant friction arises from the coupling of the planet with the planetesimals located *exactly* at the secular resonance; see e.g. Tremaine (1998). In other words, it neglects the contribution of planetesimals around $a_{\rm res}$ which could also attain relatively large eccentricities (Figure 6.3). This explanation is further supported by the fact that the resonance widths – as defined in Section 5.4.3 – increase proportionally with a_p (Equation 5.20), and so the effects of planetesimals around the resonance become more pronounced as $a_p/a_{\rm in} \to 1$. For reference, the values of a_p for which the expected resonance widths are $w = 3$, 5, and 10 au are highlighted in Figure 6.9. In summary, the results presented in Figure 6.9 provide strong evidence that the decay of $e_p(t)$ in the nominal N-ring simulations results from resonant friction.

Before moving on, it is also worth discussing the implications of the results portrayed in Figure 6.9. First, Figure 6.9 shows that for a fixed a_p, the decay rate scales linearly with m_p so that $D/m_p = \text{const}$. To understand this behaviour, it is important to recall that simulations for a fixed value of a_p have different masses m_p and M_d but the same ratio M_d/m_p (to ensure the location of the resonance remains the same, see Chapter 5). Thus, similar to the case for precessional timescales (see Section 5.6.2.3 for detailed discussion), the larger (smaller) the masses are the shorter (longer) the secular timescales are, including the decay rate. Second, it is clear that D/m_p grows with increasing a_p such that $D/m_p \propto a_p^{11/2}$ (see the dashed black curve), meaning that the decay is faster for planets situated closer to the disc than to the central star. This makes intuitive sense since the torque exerted by the disc on the planet will be stronger when the planet and the resonance location are closer together, i.e., when $a_p/a_{\rm res} \to 1$.

All of the dependencies discussed above can be understood by considering the approximate form of the decay rate given by Equation (6.18). Indeed, inserting the condition for secular resonance, i.e. Equation (5.16) or (6.23), with $p = 1$ into the expression of D in Equation (6.18), and taking the limits of $a_p/a_{\rm in} \to 0$ and $a_{\rm in} \ll a_{\rm res} \ll a_{\rm out}$ so that one can use the asymptotic behaviours of $b_{3/2}^{(m)}$ and dA/da (Equations F.4), it is straightforward to arrive at the following scaling relationship:

$$D \approx 2.5 \times 10^{-2}\ {\rm Myr}^{-1} \left(\frac{m_p}{1 M_J}\right) \left(\frac{M_c}{1.09 M_\odot}\right)^{-1/2} \left(\frac{a_p}{20\ {\rm au}}\right)^{11/2} \left(\frac{a_{\rm res}}{70\ {\rm au}}\right)^{-15/2}, \quad (6.20)$$

where the numerical coefficient is obtained for the fiducial disc model (i.e., $p = 1, \delta \equiv a_{\rm out}/a_{\rm in} = 5$). Using Equation (6.20), one can also estimate the half life τ_D of the

eccentricity decay so that,

$$\tau_D = \frac{2\log 2}{D} \approx 55 \text{ Myr} \left(\frac{m_p}{1M_J}\right)^{-1} \left(\frac{M_c}{1.09M_\odot}\right)^{1/2} \left(\frac{a_p}{20 \text{ au}}\right)^{-11/2} \left(\frac{a_{\rm res}}{70 \text{ au}}\right)^{15/2}. \quad (6.21)$$

Equations (6.20) and (6.21) – which can alternatively be expressed in term of M_d using Equation (5.18) – provide useful formulae for approximating a priori the decay rate of e_p in a given planet–disc system, without running any simulation.

Finally, I note that in principle, the very same torques that cause the planetary eccentricity to decay could also lead to the migration of the planet, which was not allowed for within my model. Indeed, according to Ward & Hahn (1998a), the rate at which the planetary semimajor axis varies is given by (see their equation 38):

$$\frac{1}{a_p}\frac{da_p}{dt} = 2e_p \frac{de_p}{dt}\frac{\dot{\varpi}_p}{n_p}. \quad (6.22)$$

It is clear, however, that the planetary semimajor axis would essentially be unaffected during the decay of the eccentricity, since $\dot{\varpi}_p \ll n_p$. This is not surprising since, by definition, resonant friction is a secular process.

6.4.3 Location of Secular Resonance

Results of Section 6.3 show that for the same planet–disc model (i.e. `Model A`), the secular resonance occurs at a larger semimajor axis in the nominal N-ring simulations than that expected from Chapter 5 – namely, ≈ 75 au compared to 70 au. This is a generic behaviour across all of my simulations, and thus requires an explanation. In this section, I interpret this difference as a result of the slower planetary precession rate in the simulations due to the non-axisymmetric component of the disc gravity (Section 6.4.1).

As already mentioned in Chapter 5, secular apsidal resonances occur at semimajor axes $a = a_{\rm res}$ where the precession rate of planetesimal free eccentricity due to both the disc and planet gravity is commensurate with the apsidal precession rate of the planet due to the disc:

$$A_d(a_{\rm res}) + A_p(a_{\rm res}) = \dot{\varpi}_p, \quad (6.23)$$

see Equation (5.16). In the previous chapter, given that the disc's potential was taken to be axisymmetric, $\dot{\varpi}_p$ represented the free precession rate of the planet, i.e., $\dot{\varpi}_p = A_{d,p}$ – see Section 5.2.2.2. In the nominal N-ring simulations presented here, however, this is no longer true since the non-axisymmetric component of disc's gravity causes the

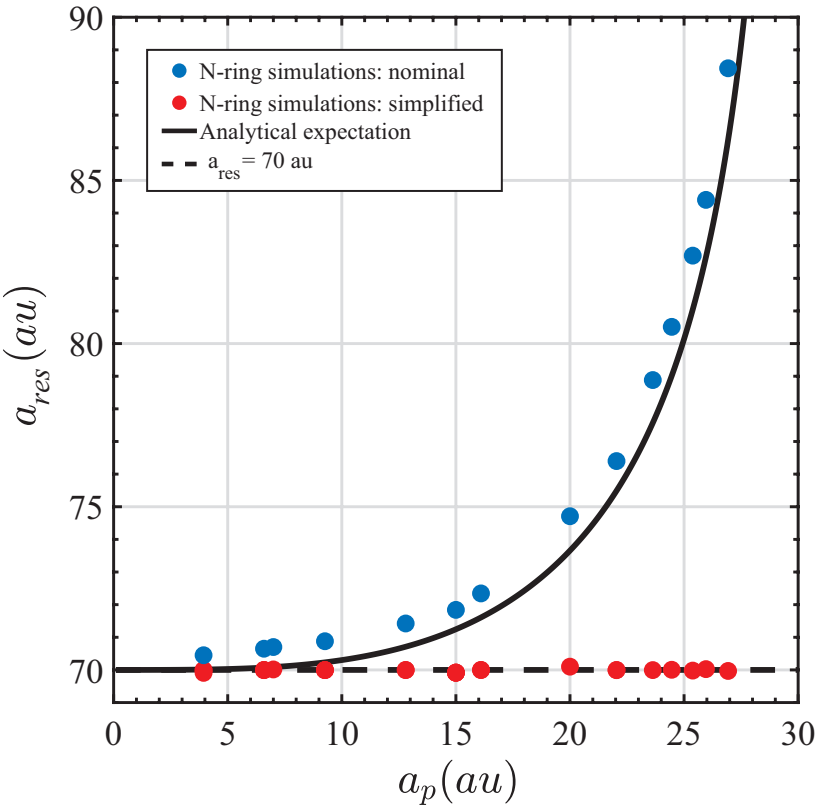

Figure 6.10. The location of the secular resonance $a_{\rm res}$ as a function of planetary semimajor axis a_p, as measured from N-ring simulations. Simulations are done for 67 different planet–disc models (Table H.1) within two sets: simplified (red circles) and nominal (blue circles). One can see that when the disc's non-axisymmetric torque on the planet is accounted for, $a_{\rm res}$ shifts to larger values than those expected when ignored – namely, $a_{\rm res} = 70$ au (dashed black line). The numerical data extracted from nominal N-ring simulations follow closely the black solid curve, which represents the expected resonance location upon accounting for the reduced planetary precession rate due to the disc's non-axisymmetric torque on the planet. See the text (Section 6.4.3) for more details.

planet to precess at a slower rate so that $\dot{\varpi}_p \lesssim A_{d,p}$; see Section 6.4.1 and Equation (6.17). In terms of the resonance condition, this means that for a given planet–disc system, the right hand side of Equation (6.23) will be systematically smaller than in Chapter 5, while the left hand side will remain the same as it is determined by the axisymmetric components of both the planet and disc gravity. It then follows that for a given planet–disc system, the resonance location should be expected to be pushed outwards to a larger semimajor axis in the nominal N-ring simulations than that in Chapter 5 (or the simplified N-ring simulations). This could also be understood, for instance, by looking at the curves for $A(a)$, $A_{d,p}$, and $\dot{\varpi}_p$ in Figure 6.4.

To test this explanation, I computed the theoretical locations of the secular resonance that would be expected upon setting $\dot{\varpi}_p$ in Equation (6.23) equal to the analytical expression given by Equation (6.17) – rather than equal to $A_{d,p}$ – and compared them with the numerical results of $a_{\rm res}$ extracted from the suite of 67 nominal N-ring simulations. For a meaningful comparison, the theoretical calculations assumed the combinations of M_d/m_p and $a_p/a_{\rm in}$ that would guarantee a secular resonance at $a_{\rm res} = 70$ au if $\dot{\varpi}_p$ were equal to $A_{d,p}$ (Section 5.4.1), which are the same as those adopted in the numerical simulations (Table H.1). The results obtained are shown in Figure 6.8, which illustrates how the resonance location varies with planetary semimajor axis both stemming from the theoretical calculations (black curve) and the nominal N-ring simulations (blue filled circles). One can clearly see that accounting for the disc's non-axisymmetric gravity in the condition of secular resonance provides a very good description of the behaviour observed in the nominal N-ring simulations. Note that the fact that the theoretical expectation underestimates the simulation results is not surprising given that the theoretical values of $\dot{\varpi}_p/A_{d,p}$ overestimate their numerical counterparts (Figure 6.8). As another check, I also analysed the set of simplified N-ring simulations, in which case $\dot{\varpi}_p = A_{d,p}$ (see Figure 6.8), and found that the expected value of $a_{\rm res} = 70$ au is recovered in every simulation; see the red filled circles in Figure 6.10. In summary, the results shown in Figure 6.10 confirm the explanation for the shift in resonance locations being due to the non-axisymmetric component of the disc gravity.

Before moving on, however, there are other features in Figure 6.10 worth noting. First, looking at Figures 6.10, one can see that the resonance location is not shifted significantly for $a_p \lesssim 20$ au, or at least not as strongly as $\dot{\varpi}_p/A_{d,p}$ does in the same region; see Figure 6.8. This follows from the fact that for $a_p/a_{\rm in} \ll 1$, one has $\dot{\varpi}_p \to A_{d,p}$ and $A_{d,p} \propto a_p^{3/2} \to 0$ (see Figure 6.8 and Equation 5.7), so that the planet's precession rate does not contribute much to the resonance condition (Equation 6.23) – with or without the corrections to the planetary precession due to the disc's non-axisymmetric torque on the planet. Second, and relatedly, Figure 6.10 shows that the effects of the reduced planetary precession rate become prominent for $a_p \to a_{\rm in}$, and the resonance locations shift considerably relative to the predictions based on Chapter 5: for instance, by as much as $\approx 30\%$ for $a_p/a_{\rm in} \approx 1$. Obviously, this shift results from the fact that for $a_p \to a_{\rm in}$ one has $\dot{\varpi}_p \to 0$ (Figure 6.8), and thus affects the resonance condition considerably when compared to Chapter 5 which, in the same limit, had $\dot{\varpi}_p = A_{d,p} \to \infty$ (e.g., see Equation 5.7 and Appendix E). In other words, the numerical results suggest that when the disc's non-axisymmetric gravity is included,

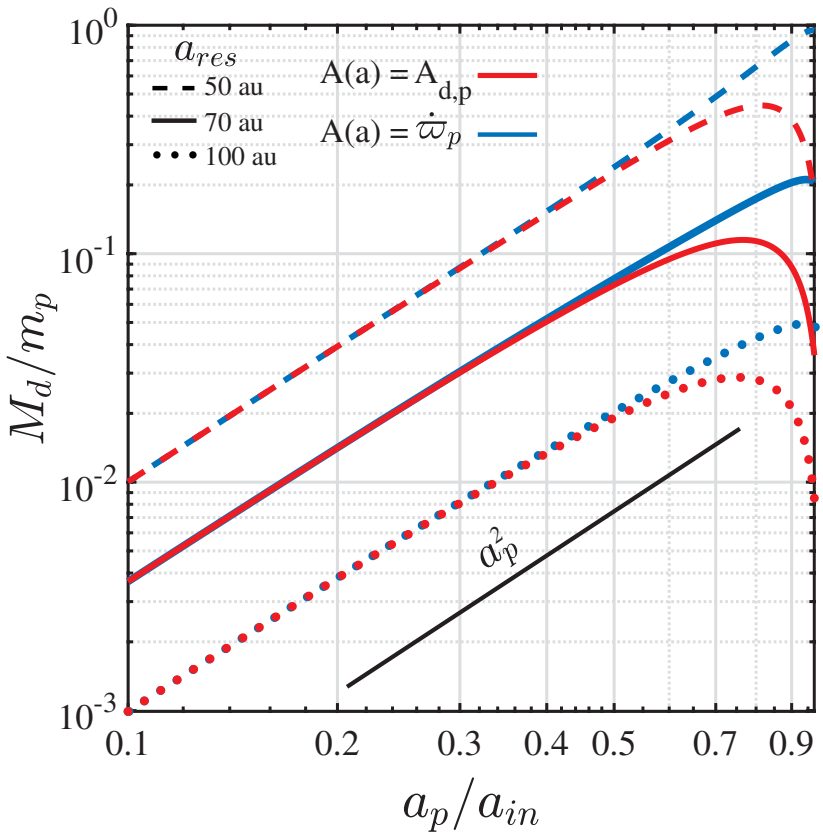

Figure 6.11. The disc-to-planet mass ratio M_d/m_p required, as a function of planetary semimajor axis $a_p/a_{\rm in}$, to place a secular resonance within the disc at $a_{\rm res}$. Calculations are done for the fiducial disc model ($p = 1, \delta = a_{\rm out}/a_{\rm in} = 5, a_{\rm in} = 30$ au) by solving the resonance condition given by Equation (6.23) under two assumptions: once within the context of simplified simulations (i.e., $A(a) = A_{d,p}$, red curves), and once within the context of nominal simulations (i.e., $A(a) = \dot{\varpi}_p$, blue curves). The results obtained for three different values of $a_{\rm res}$ are shown by different line types, as indicated in the legend. One can see that for planets orbiting close to the disc, M_d/m_p is larger in the setting of nominal simulations than that of simplified ones. Note also that the scaling of M_d/m_p with $a_p/a_{\rm in}$ in the nominal case is well captured by the black curve representing Equation (5.18), even in the limit of $a_p \to a_{\rm in}$. See the text (Section 6.4.3) for more details.

one can safely neglect the planetary precession rate in the resonance conditions for all values of $a_p/a_{\rm in}$, i.e., $A_d(a_{\rm res}) + A_p(a_{\rm res}) \approx 0$ – an assumption valid only for $a_p/a_{\rm in} \to 0$ in Chapter 5.

The above observation thus implies that Equation (5.18) as derived in Chapter 5 – i.e., by neglecting the contribution of $\dot{\varpi}_p = A_{d,p}$ – provides a better explanation for the behaviour of the resonance locations in the nominal N-ring simulations than in the simplified ones. This is so even at large values of $a_p/a_{\rm in}$, where previously Equation (5.18) overestimated the resonance location for a given M_d/m_p and $a_p/a_{\rm in}$ due to $A_{d,p}$

diverging (Figure 5.5 and Section 5.4.1). For instance, evaluating Equation (5.18) at the parameters of `Model A`, one obtains $a_{\rm res} \approx 77$ au (not so dissimilar from the result in Figure 6.10), rather than $a_{\rm res} = 70$ au. Conversely, this discussion indicates that for the N-ring simulations to establish a secular resonance at some $a_{\rm res}$, the ratio M_d/m_p for a given $a_p/a_{\rm in}$ must be larger than that expected from Chapter 5 (e.g. Figure 5.5). This is confirmed in Figure 6.11, where I compare the curves of M_d/m_p as a function of $a_p/a_{\rm in}$ for three different values of $a_{\rm res}/a_{\rm in}$ resulting from solving the resonance condition, i.e., Equation (6.23), for the fiducial disc model ($p = 1, \delta = 5$) by first equating $\dot{\varpi}_p$ to $A_{d,p}$ given by Equation (5.7) and then to the expression given by Equation (6.17). One can see that the ratio M_d/m_p must be corrected by increasing its value by at most a factor of $\approx 2 - 3$ for reasonable large values of $a_p/a_{\rm in}$.

6.4.4 Timescale for Excitation of Planetesimal Eccentricities

Results of Section 6.3 show that for the same planet–disc model (i.e., `Model A`), the excitation of eccentricity at the location of secular resonance $a_{\rm res}$ takes a longer time in the nominal N-ring simulation than that expected from Chapter 5 – namely, $\tau \approx 332$ Myr instead of ≈ 135 Myr. In this section, I conduct a more quantitative analysis of this behaviour by analysing the complete suite of 67 planet-disc models within the nominal set of N-ring simulations (Table H.1). For ease of discussion, in this section I use $\tau_{\rm n.axi}$ and $\tau_{\rm axi}$ to refer to the time it takes for $e(a_{\rm res}) \to 1$ in the presence and absence of the disc's non-axisymmetric torque on the planet, respectively.

As already mentioned in Section 6.3.1.3, the delay in the time at which $e(a_{\rm res})$ attains a value of unity is caused by the fact that the planetary eccentricity decays in the course of the evolution, and in turn affects the rate at which $e(a_{\rm res})$ grows; see e.g. Figure 6.5(b) and Equation (6.16). Given this, one can then invert Equation (6.16) to solve for the time $\tau_{\rm n.axi}$ at which $e(\tau_{\rm n.axi}) = 1$, finding that

$$\begin{aligned} \tau_{\rm n.axi} &= \log\left[\left(1 - \frac{D}{2|B_p(0)|}\right)^{-\frac{2}{D}}\right], \\ &\approx \frac{1}{|B_p(0)|}\left(1 + \frac{D}{4|B_p(0)|}\right), \end{aligned} \tag{6.24}$$

where all quantities are evaluated at the secular resonance, $a = a_{\rm res}$, and the approximation in the second line assumes $D \ll |B_p(0)|$. Looking at Equation (6.24), one can see that for a given planet–disc system, and depending on the strength of resonant

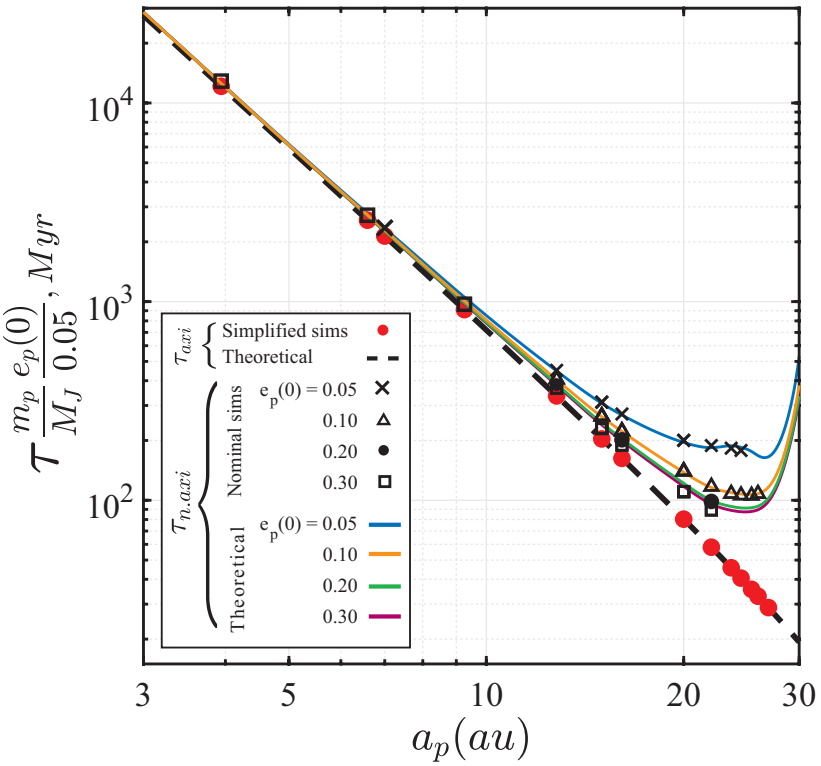

Figure 6.12. The timescale τ for planetesimal eccentricity excitation at the secular resonance as a function of planetary semimajor axis a_p, as measured from N-ring simulations. Simulations are done for 67 different planet–disc models (Table H.1) within two sets: (i) simplified, which are shown in red circles; and (ii) nominal, which are shown using various black symbols explained in the legend. Note that the timescales τ are scaled by the planet's mass m_p and initial eccentricity $e_p(0)$. One can see that when the disc's non-axisymmetric torque on the planet is accounted for, $\tau \equiv \tau_{\rm n.axi}$ is larger than those expected when ignored, $\tau = \tau_{\rm axi}$ (dashed black line, Equation 5.15). The numerical data extracted from nominal N-ring simulations follow closely the solid curves of different colours (see the legend), which represent the expected timescale upon accounting for the fact that the planet's eccentricity decays due to the disc's non-axisymmetric torque on the planet (Equation 6.24). See the text (Section 6.4.4) for more details.

friction (i.e., D), the value of $\tau_{\rm n.axi}$ should be expected to be equal to or larger than $\tau_{\rm axi}$ of Chapter 5 which, I remind, is given by $\tau_{\rm axi} = 1/|B_p(0)|$ – see Equation (5.15).

To test and confirm this explanation, I computed the theoretical values of $\tau_{\rm n.axi}$ using Equation (6.24) for various values of $e_p(0)$ and compared them with the numerical values of $\tau_{\rm n.axi}$ that I extracted from the suite of 67 nominal N-ring simulations (Table H.1). For a meaningful comparison, in the theoretical calculations I adopted the numerical values of D and $a_{\rm res}$ as extracted from the simulations and shown in Figures 6.9 and 6.10, respectively, rather than using their corresponding theoretical predictions (i.e. Equations 6.18 and 6.17). This is because while the theoretical predictions do

explain the behaviour of D and $a_{\rm res}$ as a function of a_p, they do not provide a perfect quantitative description; see Sections 6.4.2 and 6.4.3.

The results of this exercise are summarised in Figure 6.12, which shows both the theoretical and numerical values of $\tau_{\rm n.axi}$ as a function of planetary semimajor axis a_p for four different values of $e_p(0)$, as indicated in the legend. Note that the results are normalised by the planet's mass and initial eccentricity motivated by the fact that according to Chapter 5, $\tau_{\rm axi} \propto [m_p e_p(0)]^{-1}$ (Equation 5.15). For reference, Figure 6.12 also shows the normalised values of $\tau_{\rm axi}$ as extracted from the simplified simulations (see the red circles): as expected, they all lie perfectly along the black dashed curve representing Equation (5.15), irrespective of m_p and $e_p(0)$. More importantly, looking at Figure 6.12, one can clearly see that Equation (6.24) (shown in various solid lines; see the legend) provides a very good description of the numerical values of $\tau_{\rm n.axi}$ (shown in various black symbols; see the legend) for all considered values of $e_p(0)$ and a_p, confirming that resonant friction on the planet is indeed responsible for the delay in exciting the planetesimal eccentricities at the resonance.

Indeed, looking at Figure 6.12 one can see that the values of $\tau_{\rm n.axi}$ – regardless of $e_p(0)$ – converge to the curve of $\tau_{\rm axi}$ at small values of a_p. This is because resonant friction on the planet in that region is so weak that $e_p(t)$ remains roughly constant, and $\tau_{\rm n.axi} \to \tau_{\rm axi}$; see also Equation (6.24). In the opposite limit, however, i.e., as $a_p \to a_{\rm in}$ and the effects of resonant friction on the planet become significant, the values of $\tau_{\rm n.axi}$ diverge away from the curve of $\tau_{\rm axi}$, attaining values larger than $\tau_{\rm axi}$ by as much as a factor of ~ 5. This factor depends on the specific value of $e_p(0)$, with initially more eccentric planets leading to relatively smaller shifts in $\tau_{\rm n.axi}$ than nearly circular planets. This makes intuitive sense since the decay rate of e_p does not depend on $e_p(0)$ (Section 6.4.2) and by and large, $B_p \propto e_p$ controls the amplitude of planetesimal eccentricities (see e.g. Equations 6.16 and I.3).

This concludes my analysis of the simplified N-ring simulations.

6.5 Results: Full N-ring Simulations

The results presented thus far in this chapter have been obtained by running 'nominal' N-ring simulations, i.e., accounting for all gravitational effects associated with the planet and the debris disc, with the exception of the non-axisymmetric component of the disc self-gravity (see Section 6.3). This simplification allowed me to explore and develop a comprehensive understanding of the key dynamical effects due to the

non-axisymmetric torque exerted by the disc upon the planet (e.g., Section 6.4), in addition to those already explored in Chapter 5. However, it is important to emphasise that the 'nominal' N-ring simulations do not capture the full richness of the dynamical phenomena that may ensue from the secular interactions between planets and massive debris discs. By and large, this is because the negligence of the non-axisymmetric gravitational perturbations amongst the disc rings inhibits their exchange of angular momentum within the disc.

In this section, I report on simulation results obtained with the N-ring model, but this time by accounting for the full gravitational effects of the disc, both on the planet and the disc itself – simulations referred to as 'full' N-ring simulations (Section 6.2). At the outset, however, I would like to emphasise that the results presented here are preliminary, in the sense that I plan on extending them in the near future, and thus this section should be viewed as a work in progress.

As a basis for the investigation, I simulated the evolution of the same 67 planet–disc models considered thus far in this chapter; see Table H.1. Before diving into the details in the subsequent sections, here I state the main results briefly. Examining the complete set of the full N-ring simulations, I found that there are two qualitatively distinct simulation outcomes. The fundamental difference between these outcomes lies in the way both the planetary and planetesimals orbits evolve in time, leading to different evolutionary stages for the disc surface density and thus observed morphology. The key parameter dictating which outcome occurred was set by the semimajor axis of the planetary orbit a_p relative to the disc's inner edge $a_{\rm in}$: indeed, one outcome – hereafter, 'Outcome I' – occurred for relatively large planetary semimajor axes, that is, for $a_p/a_{\rm in} \to 1$, and a second outcome – hereafter, 'Outcome II' – occurred for smaller planetary semimajor axes, that is, for $a_p/a_{\rm in} \to 0$. Outcome I resulted in the circularisation of the planetary orbit and a disc structure characterised by a gap at the secular resonance (somewhat similar to the case in the nominal N-ring simulations). In addition to this, however, the disc region beyond the gap developed a long, one-armed spiral density wave. Outcome II, on the other hand, resulted in a constant planetary eccentricity and the disc maintained a roughly coherent structure. In addition to these two outcomes, I also identify an intermediary stage, whereby the system behaves akin to Outcome I in certain aspects, but akin to Outcome II in others. I describe these two outcomes together with the intermediary stage in detail below.

6.5.1 Outcome I: 'Gap + Spiral Wave'

I begin by describing the evolution of planet–disc systems whereby the planet orbits relatively close to the disc, i.e., $a_p \lesssim a_{\rm in}$ (see also Table H.1). For clarity, however, I remind that the planetary orbit does not cross the disc. As already mentioned above, I refer to the results of simulations with such configurations as Outcome I.

6.5.1.1 The Planet

In Outcome I, the planetary orbit evolves by and large in a similar manner as in the nominal N-ring simulations of Section 6.3. Indeed, the planet's eccentricity decays in time due to resonant friction, while at the same its longitude of pericentre precesses in a prograde fashion at a constant rate. The precession rate $\dot{\varpi}_p$ remains by and large unaffected when compared to the nominal simulations. This makes intuitive sense, since $\dot{\varpi}_p$ is dictated by the non-axisymmetric component of the disc gravity on the planet (Section 6.4.1). The decay of the planetary eccentricity, however, proceeds somewhat differently. Indeed, instead of $e_p(t)$ decaying according to the exponential behaviour discussed in Section 6.4.2 at all times, it undergoes a milder decline at early times and only thereafter follows the exponential profile given by Equation (6.19). This milder decline lasts for $t \sim 2\ \tau_{\rm sec}$, that is, for the first ~ 2 precession cycles of the planetary orbit. After this time, i.e., at $t \gtrsim 2\ \tau_{\rm sec}$, $e_p(t)$ behaves according to Equation (6.19) but with a decay rate D which is smaller by about a factor of $\sim 2-3$ than that in the nominal simulations (e.g. Figure 6.9). Last but not least, the exponential decay is accompanied with small-amplitude oscillations in time, similar to the case in the nominal simulations.

The behaviour described above can be seen by looking at Figure 6.13, where I plot the time evolution of the planetary orbit in `Model A` ($a_p = 20$ au, Table H.1) – which, I remind, is the fiducial configuration adopted both in Chapter 5 and in Section 6.3. Note that the results in Figure 6.13 are shown up until $t = \tau \approx 424$ Myr, where I remind that τ is the time at which the planetesimal eccentricities at the resonance location become equal to unity. For reference, panels (A) and (B) of Figure 6.13 also show in red lines the behaviour of $e_p(t)$ and $\varpi_p(t)$ as extracted from the nominal simulation of `Model A`, respectively.

6.5.1.2 The Planetesimals

In Outcome I, debris particles interior to the secular resonance location, i.e., those with $a \lesssim a_{\rm res}$, evolve following the same stages as in the nominal simulations. Thus,

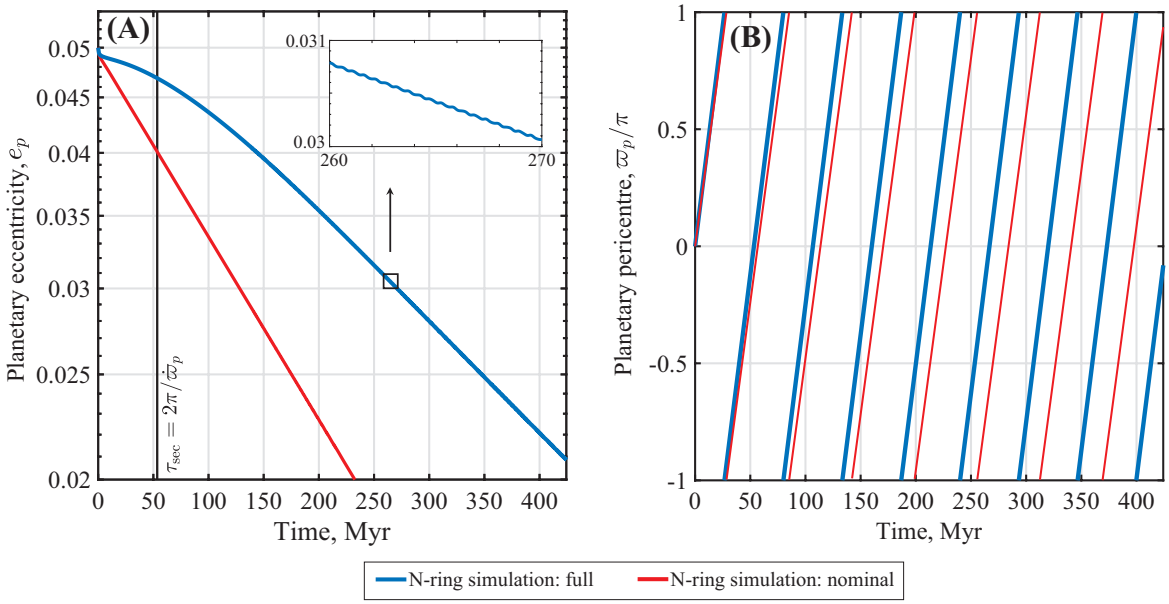

Figure 6.13. The evolution of the planetary eccentricity e_p (panel A) and longitude of pericentre ϖ_p (panel B) as a function of time in the full N-ring simulation of `Model A` ($a_p = 20$ au, see Table H.1). The vertical black line in panel (A) marks the period of the planetary precession, $\tau_{\rm sec} \approx 54$ Myr. For reference, the results obtained from the nominal simulation of `Model A` are shown by red lines in each panel. Note that e_p decays at a slower rate in the full N-ring simulation compared to the nominal simulation. Additionally, the decay does not follow an exponential behaviour until at least $\sim 2\tau_{\rm sec}$ have elapsed. It is also evident that the planetary precession rate is roughly the same in both types of simulations. The results shown here serve as an example of the general behaviour observed in all simulations resulting in Outcome I. See the text (Section 6.5.1) for more details.

for the sake of conciseness, I do not reiterate the description here and refer the reader to Section 6.3.1.2. This said, however, the evolution of planetesimals at and beyond $a_{\rm res}$ display various qualitative and quantitative differences when compared to the nominal simulations. This can be seen by looking at rows (a) and (b) in Figure 6.14, where I show the evolution of planetesimal eccentricities and apsidal angles in `Model A`, respectively[4]. The snapshot times were chosen such that they roughly correspond to $1/2, 1, 2$, and 5 times the planetary precession period, $\tau_{\rm sec} \approx 54$ Myr. Note the vertical dashed lines therein at $a \approx 74.11$ au, which denote the exact location where the resonance condition (Equation 6.23) is satisfied for this particular model.

Looking at rows (a) and (b) of Figure 6.14, one can see that by the time that the planet completes one precession cycle, $t \sim \tau_{\rm sec}$, the eccentricities at the resonance do not grow as much as in the nominal simulations (see e.g. the first two columns of

[4]Animated versions of rows (a) and (b) of Figure 6.14 which runs from $t = 0$ to $t = \tau \approx 424$ Myr are made available online.

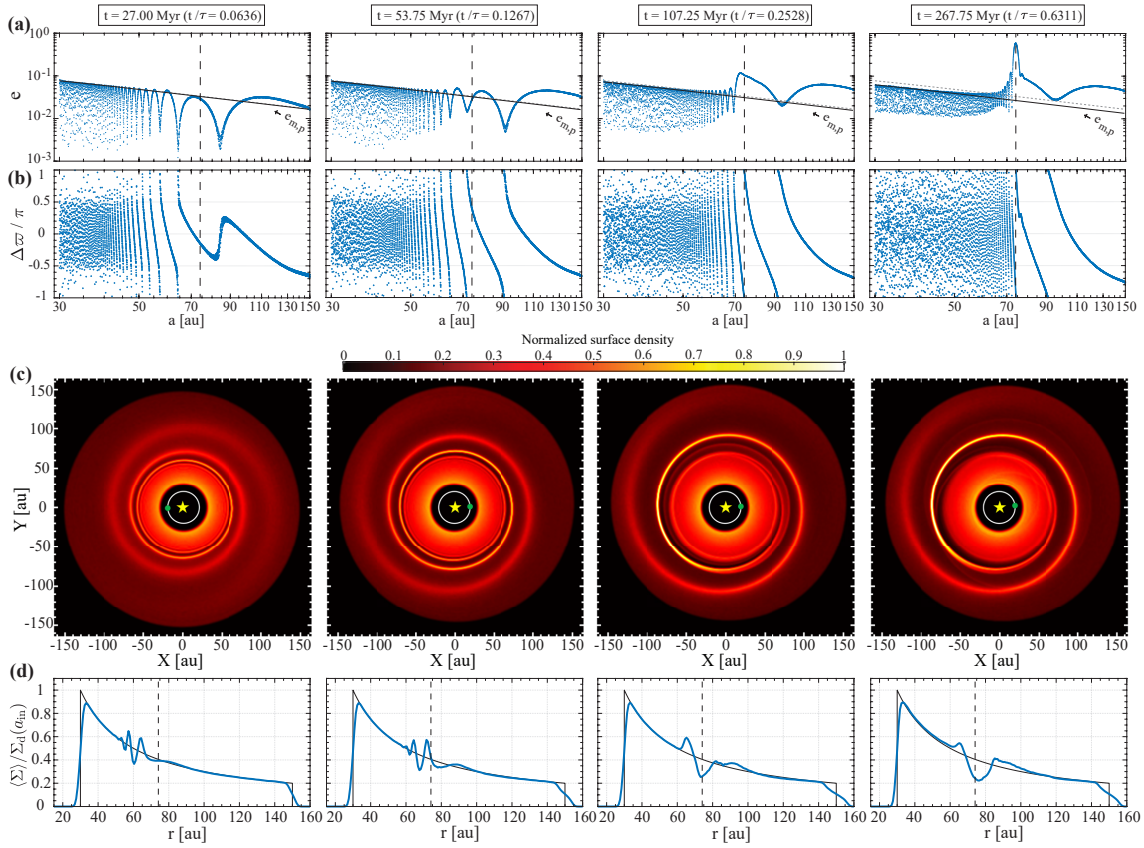

Figure 6.14. Summary of the results obtained by running a full N-ring simulation for `Model A` ($a_p = 20$ au, Table H.1), as an example of the general behaviour observed in all simulations resulting in Outcome I (Section 6.5.1). Four snapshots of the simulation are shown. The time t is given at the top of each column, which for reference is also indicated relative to $\tau \approx 424$ Myr (Equations 5.15, 6.21). Note that the chosen times correspond to 1/2, 1, 2 and 5 periods of planetary precession, $\tau_{\rm sec} = 2\pi/\dot{\varpi}_p \approx 54$ Myr. Rows (a) and (b) show the planetesimal eccentricities and apsidal angles (relative to that of the planet) as a function of semimajor axis, respectively. Rows (c) and (d) show the corresponding snapshots of the disc surface density and radial profiles of the azimuthally averaged surface density – see Appendix G for details. The dashed vertical lines in rows (a), (b), and (d) show the secular resonance location, $a_{\rm res} \approx 74.11$ au. The black solid line in row (a) represents the maximum planetesimal eccentricities $e_{m,p}$ driven by the circularising planet in the absence of the disc: it is calculated using Equation (6.15) but with the numerical results of $e_p(t)$ in Figure 6.13, and its initial value (i.e., with $e_p = e_p(0)$, Equation 5.13) is shown by small-dashed grey lines for reference. Animated versions of each of the four rows in this figure are made available online. See the text (Section 6.5.1) for details.

Figure 6.14). Instead, what happens is that the eccentricities first grow in the outer parts of the disc (and not at $a_{\rm res}$ as in the nominal simulations), where they attain amplitudes larger than $e_{m,p}$, i.e., the maximum eccentricity driven by the planet alone; see e.g. Equation (5.12). As this happens, the apsidal angles $\Delta\varpi(a)$ in this region

wrap twice all around the star, from $[-\pi, \pi]$. This happens such that the resulting inner winding is narrower than the outer one, in the sense that the former spans the semimajor axis range of $\approx 75 - 90$ au while the latter spans $\approx 90 - 150$ au.

By the time that the planet completes its second precession cycle, $t \sim 2\ \tau_{\rm sec}$, $\Delta\varpi(a_{\rm res})$ becomes fixed at $-\pi$ (and not $-\pi/2$ as in the nominal simulations), and $e(a_{\rm res})$ starts to grow significantly from thereon (see e.g. the third column of Figure 6.14). Here, it is interesting to note that this happens by roughly the same time that the planet's eccentricity starts to follow the exponential behaviour of Equation (6.19); see Figure 6.13. It is also evident that during the same interval of time, i.e., $\tau_{\rm sec} \lesssim t \lesssim 2\tau_{\rm sec}$ the behaviour of $\Delta\varpi(a)$ remains unchanged in the outer disc parts, showing little or no evidence of precession relative to the planetary orbit. This is also the case at later times, i.e., $2\ \tau_{\rm sec} \lesssim t \lesssim \tau$, when the only noticeable evolution is in terms of the eccentricities: $e(a_{\rm res})$ continues its growth to unity as $t \to \tau$ and, at the same time, the eccentricities in the outer disc parts shift down roughly uniformly at all a, but remain larger than $e_{m,p}$ – see also the animated versions of rows (a) and (b) of Figure 6.14.

Before moving on, it is worthwhile to mention that the planetesimal evolution illustrated by Figure 6.14 is general, in the sense that it applies to all other simulations with $a_p \lesssim a_{\rm in}$ that result in Outcome I. The main difference that arises by varying the value of a_p is that for $a_p \to a_{\rm in}$, rather than having two windings of $\Delta\varpi(a)$ in the range $a_{\rm res} \lesssim a \lesssim a_{\rm out}$, there can happen only one. This is mainly due to the fact as $a_p \to a_{\rm in}$, the location of the resonance moves out significantly (see e.g. Figure 6.10 and Section 6.4.3), and thus only the innermost winding evident in Figure 6.14 would be established.

6.5.1.3 The Disc Morphology

Given the behaviour of the planetesimal orbital elements, the physical debris structure in Outcome I features a gap centred around the resonance location in addition to a long, one-armed spiral pattern beyond it. This can be seen by looking at rows (c) and (d) of Figure 6.14, where I show snapshots of the disc surface density and its azimuthally averaged radial profile corresponding to `Model A`, respectively[5]. Looking at this figure, it is also evident that the gap develops fully by $t \sim 2\ \tau_{\rm sec}$ – see e.g. the third column in Figure 6.14. This essentially means that the timeframe of 'Stage 1' as introduced in Section 6.3 is now extended from $t \sim \tau_{\rm sec}$ to $t \sim 2\ \tau_{\rm sec}$, which makes

[5]Animated versions of rows (c) and (d) of Figure 6.14 showing the evolution over a period of $t = \tau \approx 424$ Myr are made available online.

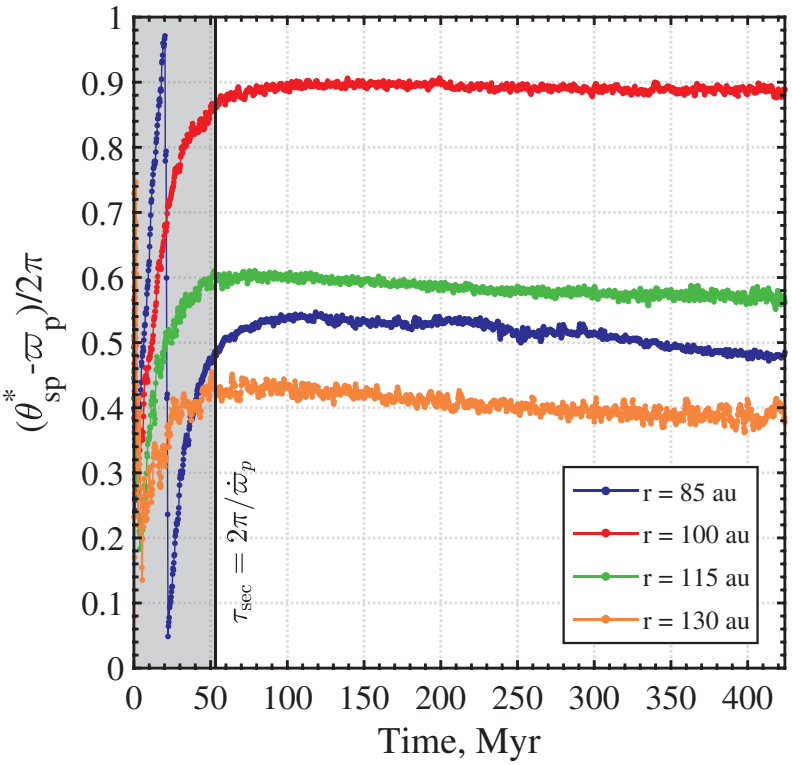

Figure 6.15. The time evolution of the relative difference between θ^*_{sp}, which measures the angular position where the disc surface density peaks at a given radial distance, and the planetary longitude of pericentre ϖ_p in the full N-ring simulation of `Model A` (Figure 6.14). Calculations are done at four different radii located beyond the resonance, $a_{\rm res} \approx 74.11$ au – see the legend. One can see that $\theta^*_{sp} - \dot{\varpi}_p \approx$ const in time at all radii, suggesting that the spiral's pattern speed is equal to the planetary precession rate $\dot{\varpi}_p$. See the text (Section 6.5.1.3) for more details.

sense given the behaviour of $e(a_{\rm res})$ discussed above in Section 6.5.1.2. Note also that during Stage 1, i.e., $0 \leq t \lesssim 2\ \tau_{\rm sec}$, the spiral arm launched by the planet at $a_{\rm in}$ extends out to about the outer edge of the disc as it wraps around the star, and not only to $a_{\rm res}$ as in the nominal simulations – see the first two columns of Figure 6.14. Once the gap is developed by $t \sim 2\ \tau_{\rm sec}$, however, the spiral arm in the outer disc parts would more or less stop wrapping onto itself, while the region interior to the gap settles into a coherent, almost axisymmetric structure (as in the nominal simulations) – see e.g. the third and fourth columns of Figure 6.14.

Here, it is worthwhile to note that the gap in Outcome I no longer appears axisymmetric as in the nominal simulations, despite e.g. the randomisation of $\Delta\varpi$ at $a \lesssim a_{\rm res}$. Instead, looking at Figure 6.14(c), (d), one can see that the depletion around the secular resonance location is of non-axisymmetric shape, and seemingly wider and

deeper towards the planetary apocentre (and not the pericentre). However, it remains difficult to say how the combination of the secular resonance and the underdensities caused by the spirals beyond that affect the appearance of the overall depletion. From an azimuthally averaged sense, however, Figure 6.14(d) shows that the gap width is ≈ 15 au at late times, which is comparable to the value found in the nominal simulation of the same planet–disc system (Section 6.3.2). Additionally, the gap depth does not seem to be affected by the underdensities induced by the spirals: indeed, Figure 6.14(d) shows that the surface density at the secular resonance is depleted by about $\approx 50\%$ relative to the initial density – similar to the nominal simulations.

Finally, I point out that the spiral pattern beyond the gap remains (by and large) stationary relative to the planet as it co-rotates with the latter's apsidal line – see e.g. Figure 6.14(c) and its animated version. To further illustrate this point, however, in Figure 6.15 I plot the time evolution of the angle θ^*_{sp} measuring the angular position of where the disc density peaks at a given radial distance in `Model A` (Figure 6.14). Looking at this figure, one can see that at each of the different radii considered – namely, $r = 85$, 100, 115, and 130 au – the relative difference between θ^*_{sp} and $\dot{\varpi}_p$ converges to a roughly constant value over time. This convergence happens after at least $\sim \tau_{\rm sec}$ has passed (see the shaded region in Figure 6.15), which, I remind, corresponds to one secular period at the resonance location. Note that Figure 6.15 also shows that the relative angle $\theta^*_{sp} - \dot{\varpi}_p$ actually drifts ever so slightly at times $t \gtrsim \tau_{\rm sec}$; however, this does not lead to substantial differential precession between the different radii. In conclusion, Figure 6.15 shows that to a good approximation, the pattern speed of the spiral wave evident in Figure 6.14 – and more generally, in Outcome I – is given by the planetary precession rate, $\dot{\varpi}_p$.

6.5.2 Outcome II: 'Coherent Disc'

The previous outcome occurred if the planet's orbit was relatively close to the disc, i.e., $0.4 \lesssim a_p/a_{\rm in} \lesssim 1$. If instead the planet's semimajor axis is below some critical value relative to the inner disc edge, the evolution of both the planet and the debris particles show significant qualitative differences from the previous regime. The value of this critical semimajor axis was around $a_p^c \approx 12$ au in my simulations (Table H.1) – although note that this is defined loosely (more on this later in Section 6.5.3). I refer to the results of simulations below this critical planetary semimajor axis as Outcome II.

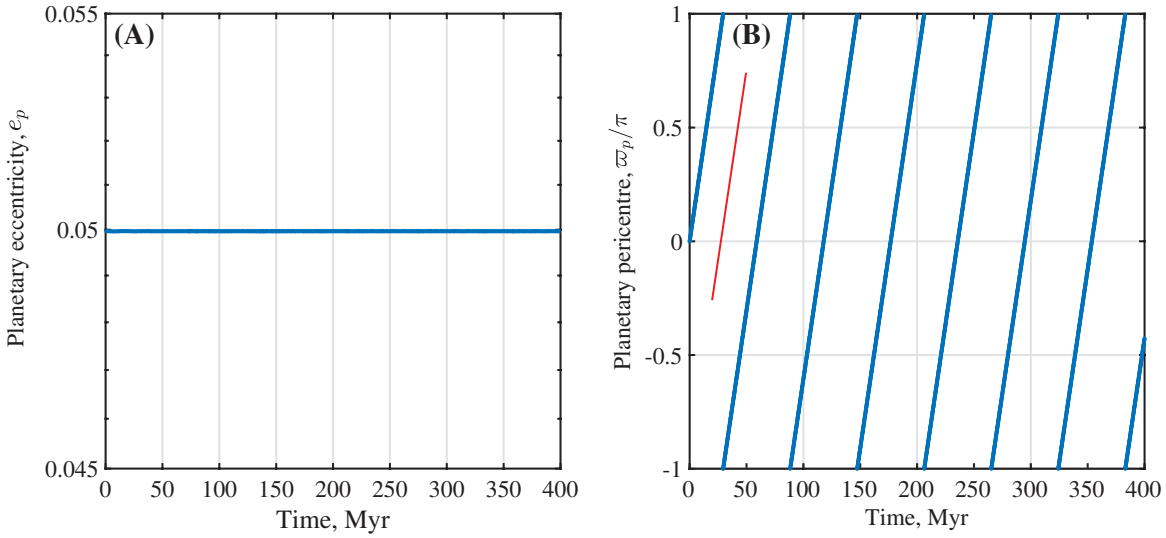

Figure 6.16. The evolution of the planetary orbit in the 'full' N-ring simulation of `Model B` ($a_p = 7$ au, Table H.1), as an example of the general behaviour observed in all simulations resulting in Outcome II (Section 6.5.2). Panels (A) and (B) show the evolution of the planet's eccentricity e_p and longitude of pericentre ϖ_p, respectively. Note that the planet's eccentricity remains constant in time. The red line in panel (B) represents the slope of $\varpi_p(t)$ extracted from the nominal simulation of `Model B` (Section 6.3). See the text (Section 6.5.2) for details.

6.5.2.1 The Planet

In Outcome II, the planetary orbit remains of constant eccentricity as it precesses under the action of the disc gravity, showing no signs of circularisation associated with resonant friction. This is illustrated in Figure 6.16, where I plot the evolution of the planetary orbit over a period of 400 Myr in `Model B` ($a_p = 7$ au; see Table H.1) – which, I remind, is one of the models also considered in Chapter 5. Note that once again the precession rate of the planetary orbit is similar to that predicted in Section 6.4.1: indeed, Equation (6.17) well approximates the period of $\tau_{\rm sec} = 2\pi/\dot{\varpi}_p \approx 59$ Myr measured in the simulation. For reference, the red line in Figure 6.16(B) shows the slope of $\varpi_p(t)$ as measured from the 'nominal' simulation of `Model B` (Section 6.3).

6.5.2.2 The Planetesimals

Similarly, the planetesimal orbits show a new and somewhat peculiar evolutionary behaviour. The main striking feature is that at and around the location where the resonance condition is satisfied, planetesimal eccentricities do not show any sign of growth to values of unity over time, contradicting expectations. Instead, planetesimal eccentricities are suppressed, remaining slightly lower than what would be expected

even in the limit of a *massless* disc, i.e., $e \lesssim e_{m,p}$ (Equation 5.13). This can be seen by looking at rows (a) and (b) in Figure 6.17, where I show the evolution of planetesimal eccentricities and apsidal angles in `Model B`, respectively, at three different times[6]. Note the vertical dashed lines therein at $a \approx 70.8$ au, which denote the exact location where the resonance condition (Equation 6.23) is satisfied for this particular model.

Moreover, while debris particles in the inner disc parts evolve by and large in line with the expectations from Chapter 5 – which, I remind, treated e_p as constant by construction – particles in the outer disc parts reveal a distinct behaviour; see e.g. rows (a) and (b) of Figure 6.17. Indeed, by the time that the planet completes one precession cycle, $t \sim \tau_{\rm sec}$, particles in the outer disc parts become nearly apse-aligned with the planet, co-precessing with the planet's apsidal line to within a small dispersion around $\Delta\varpi(a) = 0$ at all times $t \gtrsim \tau_{\rm sec}$. As this happens, planetesimal eccentricities become 'trapped' to small-amplitude oscillations in time, with an average magnitude which is lower than $e_{m,p}$ by a factor of a few and roughly independent of a. Note that interestingly, the oscillations of both e and $\Delta\varpi$ are negligible at the exact location where the resonance condition is satisfied, i.e., at $a \approx 70.8$ au – see the animated versions of rows (a) and (b) of Figure 6.17. Additionally, planetesimal eccentricities show a peculiar 'spiky' behaviour at about ≈ 65 au, with their values being pumped up to slightly larger values than the neighbouring planetesimals – see row (a) of Figure 6.17 as well as its animated version. The maximum eccentricity at this location does not grow any further than that shown in Figure 6.17, even if one integrates the system up to 2 Gyr.

6.5.2.3 The Disc Morphology

Given this behaviour, the debris disc no longer manifests a gap at the expected resonance location, but rather evolves into a coherent structure with some subtle substructure. This is illustrated in the third and fourth rows of Figure 6.17, where I show three different snapshots of the disc surface density and its azimuthally averaged radial profile corresponding to `Model B`, respectively[7]. Looking at this figure, one can see that by the time that the spiral arm launched by the planet at $a_{\rm in}$ propagates outwards to ≈ 70 au (i.e., $t \sim \tau_{\rm sec}$, akin to 'Stage 1' in Sections 5.6.1 and 6.3.2), the surface density settles into and maintains a roughly coherent axisymmetric distribution. Indeed, the only discernible substructure at late times, i.e., $t \gg \tau_{\rm sec}$, is the presence of a

[6] Animated versions of rows (a) and (b) of Figure 6.17 showing the evolution over a period of 400 Myr are made available online.

[7] Animated versions of rows (c) and (d) of Figure 6.17 showing the evolution over a period of 400 Myr are made available online.

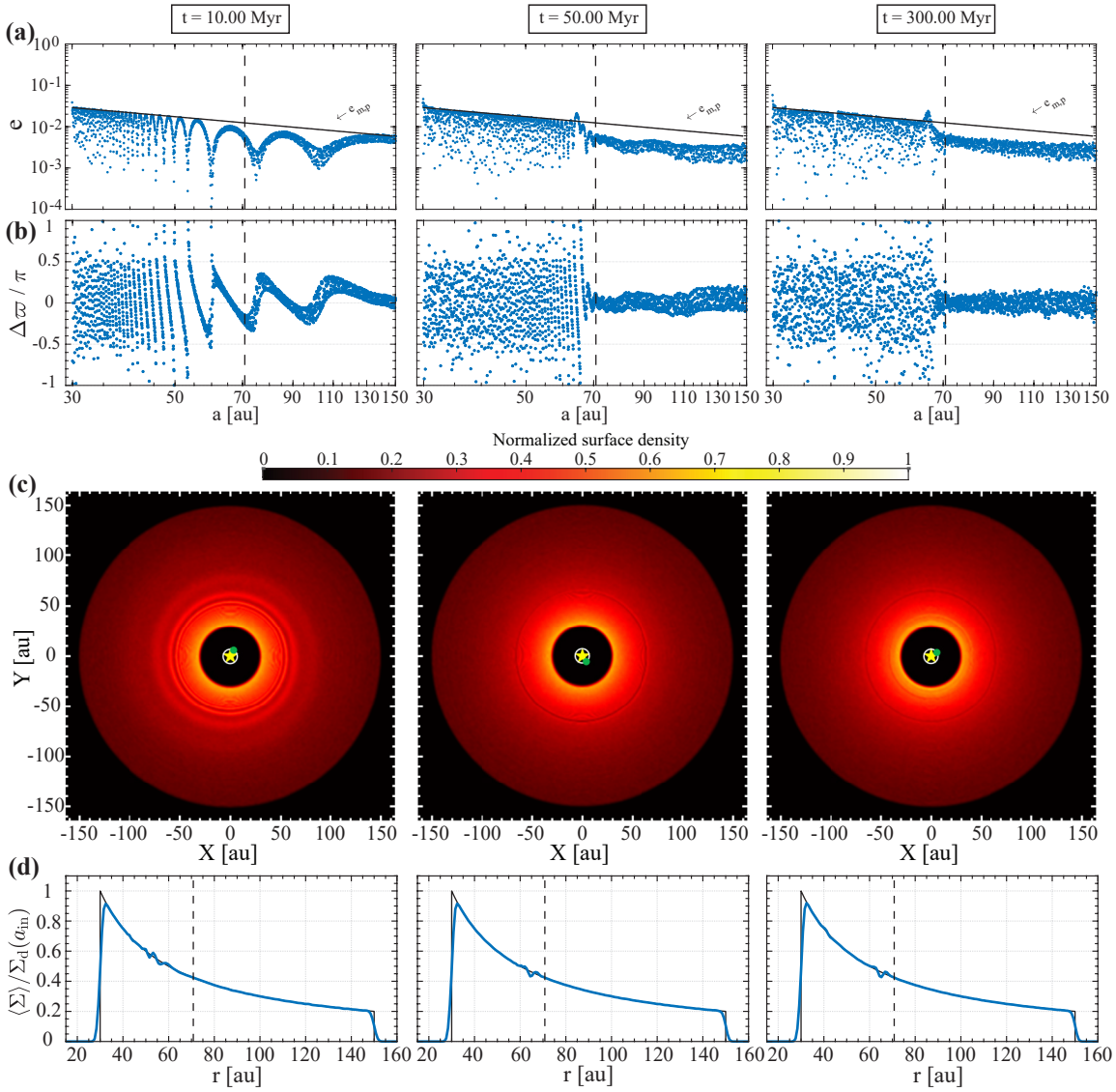

Figure 6.17. Similar to Figure 6.14, but now for `Model B` ($a_p = 7$ au, Table H.1) to illustrate the general behaviour observed in full N-ring simulations resulting in Outcome II (Section 6.5.2). One can see that planetesimal eccentricities do not show any sign of significant growth at the location where the secular resonance condition (Equation 6.23) is satisfied ($a = 70.80$ au; dashed vertical lines in rows (a), (b), and (d)). Note also the 'spiky' behaviour in the eccentricities at around 65 au. Around that region, the surface density manifests a very narrow (~ 1 au) region within which it is depleted by less than a few percent relative to the initial density. This is easier to see in the azimuthally averaged density profile at late times, e.g., at $t = 300$ Myr. Animated versions of each of the four rows in this figure are made available online. See the text (Section 6.5.2) for more details.

narrow, ~ 1 au, axisymmetric band at about 65 au (interior to the expected resonance location) where the surface density is only marginally lower than the initial density,

by less than about 5%; see row (d) of Figure 6.17. I also tested that this behaviour is maintained up to 2 Gyr e.g. for `Model B`, though for practical purposes, the animated versions of Figure 6.17 show the evolution only to 400 Myr.

Here, it is important to stress that the origins of the underdense band in Figure 6.17 should not be confounded with that of the gap in Outcome I – or Chapter 5 for that matter. This is because counter to intuition, the secular resonance – despite its condition being met – does not seem to 'act' in Outcome II, at least not to excite planetesimal eccentricities to values of unity (Figure 6.17). Instead, the origin of the underdense band is a direct consequence of the fact that both the inner and outer disc parts are offset in the same direction relative to the star in different magnitudes, and the transition between them occurs via the 'spiky' e-behavior at around 65 au. It is for these reasons that I refer to the axisymmetric 'depletion' at 65 au in Figure 6.17 as an "underdense band" and not a gap – although for practical purposes, i.e., from an observational point of view, they are practically the same.

Before moving on, I also wish to clarify that qualitatively similar behaviour was recovered with an initially more eccentric planet, say by a factor of 10 than in `Model B` (i.e., $e_p(0) = 0.50$). Indeed, increasing the planetary eccentricity did not help in exciting the planetesimal eccentricities to values larger than $e_{m,p}$ (Equation 5.13) at the location where the resonance condition is satisfied[8]. Instead, such an exercise only contributed to (i) enhancing the level of apocentre glow in the inner disc parts, as one would already expect based on models of massless discs (e.g. Wyatt et al., 1999; Pan & Wu, 2016, see also Chapter 2); and (ii) causing the underdense band to become ever so slightly broader (i.e., less than a few au) and non-axisymmetric.

6.5.3 Intermediate Outcome

Before closing this section, it is worthwhile to mention that the transition between the two outcomes, i.e. Outcome I and II, appears to not be sudden at some critical value for the planetary semimajor axis, but rather gradual. In other words, it seems that there is an intermediate stage, whereby by increasing the planetary semimajor axis from slightly less to slightly more than $a_p^c \approx 12$ au (which, admittedly, is defined rather loosely), the planet and the disc start to behave similar to Outcome I in certain aspects and similar to Outcome II in others. For purposes of clarity, I differentiate this outcome from the two previous ones by referring to it as the "intermediate" outcome.

[8]This also provides another confirmation that the observed behaviour in Outcome II does not follow from the fact that the simulations were stopped earlier than supposed (recall that in principle $\tau \propto 1/e_p$, e.g. Chapter 5).

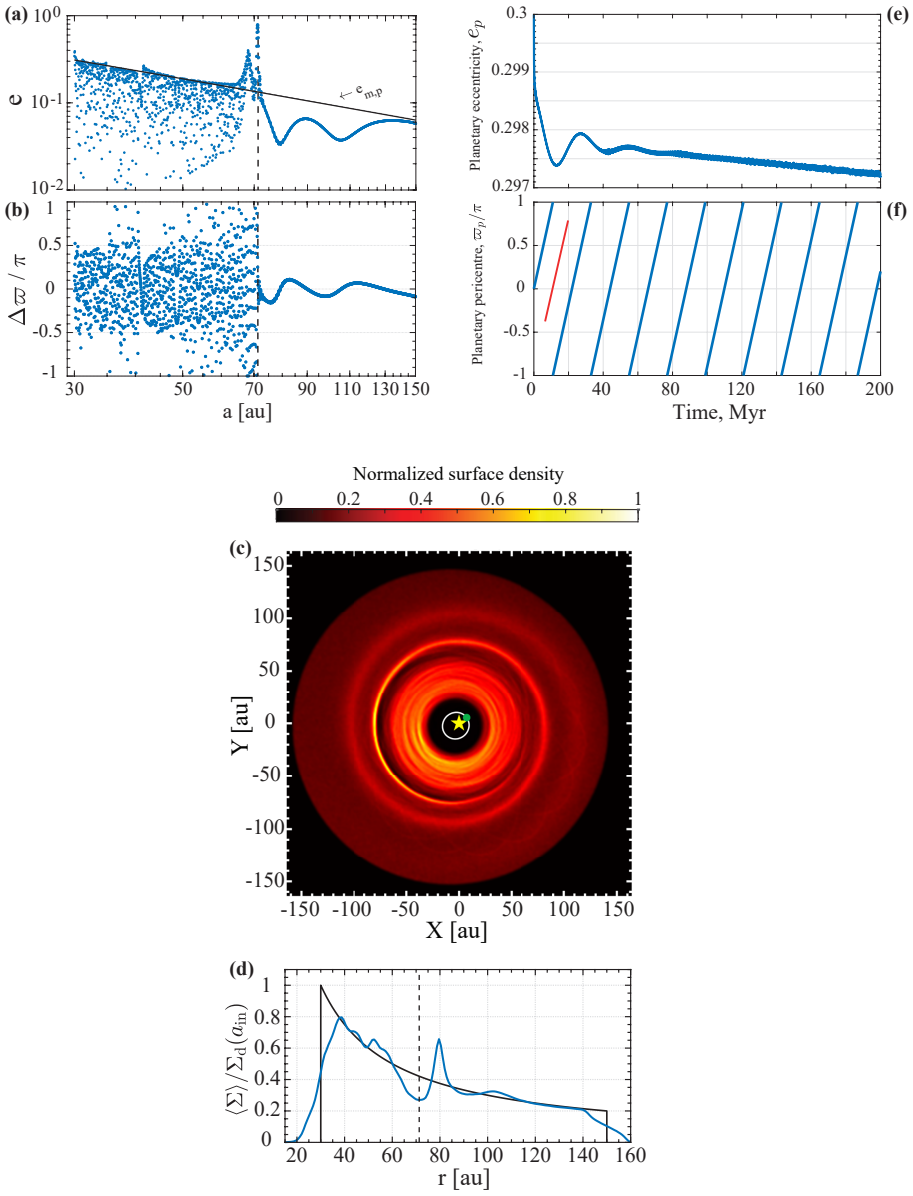

Figure 6.18. Summary of the results obtained by running a full N-ring simulation for model number 40 ($a_p = 12.8$ au, Table H.1), as an example of the general behaviour observed in simulations resulting in an intermediary stage between Outcomes I and II (Section 6.5.3). Panels (a) and (b) show the dynamical state of the planetesimals in terms of eccentricity and apsidal angles, respectively, at $t = 200$ Myr. The corresponding snapshots of the disc surface density and its azimuthally averaged radial profile are shown in panels (c) and (d), respectively. Panels (e) and (f), on the other hand, show the time evolution of the planetary eccentricity and apsidal angle, respectively, over a period of 200 Myr. The red line in panel (f) shows the slope of $\varpi_p(t)$ extracted from the nominal simulation of the same planet–disc model. The dashed vertical lines in panels (a), (b), and (d) mark the location of the secular resonance ($a_{\rm res} \approx 71.3$ au). See the text (Section 6.5.3) for details.

Analysing the outcomes of various simulations of systems whereby the planetary semimajor axis is varied within the range 10 to 15 au (see Table H.1), I have made the following observations. First, for planets with $a_p/a_{\rm in} \sim 0.4$, planetesimal eccentricities at the exact location of the resonance do start to be pumped up to relatively large values – i.e., $e(a_{\rm res}) \to 1$, as in Outcome I. Note that this seems to happen for only a few particles at $a_{\rm res}$, and not even slightly away from that location. At the same time, the 'spiky' e-behaviour described in Outcome II does still appear at $a \lesssim a_{\rm res}$. Additionally, planetesimal eccentricities in the outer disc parts behave more as in Outcome I, rather than oscillating within a narrow range as in Outcome II. Second, as this happens, the planetary eccentricity no longer remains constant in time as in Outcome II, but rather declines. However, the decline is not as significant as in Outcome I: instead, e_p drops only marginally, by about less than a few percent relative to its initial value. Once again, and similar to Outcomes I and II, the planetary precession rate is well-reproduced by Equation (6.17). Third, and finally, the outer parts of the disc develop a spiral wave with 2 windings that co-precess with the planet, similar to Outcome I. This happens even though unlike in Outcome I, the planetesimal apsidal angles $\Delta\varpi(a)$ do not cover the entire range of $[-\pi, \pi]$, but rather remain constrained to a narrow range around $\Delta\varpi = 0$, somewhat akin to Outcome II.

The features mentioned above can be seen by looking at Figure 6.18, which summarises the snapshots of the simulation labelled as '40' in Table H.1 in which the planet's semimajor axis is $a_p = 12.8$ au, comparable to a_p^c. This figure shows the dynamical state of the planetesimal orbits and the disc surface density at 200 Myr into the system's evolution, together with the evolution of the planetary orbit over a period of 200 Myr. Note that the initial planetary eccentricity was chosen to be relatively large (Table H.1), namely $e_p(0) = 0.3$, to help bring out the appearance of a spiral beyond the gap. Looking at this figure, one can see that many of the features appearing in Outcomes I and II, separately, manifest themselves in this simulation in some modified combination as discussed above.

This completes my description of the outcomes found in my 'full' N-ring simulations (Table H.1). Finally, I close this section by noting that the results presented here are preliminary in that I do not have a theoretically satisfying understanding of the physics underlying many of the features described throughout this section. As such, I refrain myself from commenting any further on the results presented here. This said, however, I plan on further investigation in the near future.

6.6 Discussion and Summary

In this chapter, I investigated the secular interactions between a single eccentric planet and an external, self-gravitating debris disc. In doing so, I have gone further than the simplified analytic model of Chapter 5, having accounted for the full secular gravitational potential of the disc, i.e., including both its axi- and non-axisymmetric components.

This was achieved by the N-ring model outlined in Section 6.2, which allows one to simulate the secular evolution of gravitating disc-planet systems using the Lagrange planetary equations (Equation 6.9). The model treats the disc as a collection of geometrically spaced massive rings, interacting both with each other and the planet via softened gravity. To this end, I adopted the softening prescription of Hahn (2003) (see also Chapter 4), which could be thought of as arising due to the disc's aspect ratio that is a consequence of the inherent velocity dispersion of the constituent particles. Note that the N-ring model employed in this chapter is not novel per se: indeed, it has been used previously in a series of papers to study the secular evolution of the primordial Kuiper belt as well as planetary rings e.g. around Saturn (Hahn, 2003, 2007, 2008). However, to the best of my knowledge, this is the first time that it has been applied to examine the secular evolution of gravitating debris-planet systems.

Results of previous sections show that introducing the non-axisymmetric component of disc gravity into the simplified scenario considered in Chapter 5 leads to a rich phenomenology. Namely, I found that there are two distinct outcomes in terms of the planet–debris disc evolution, depending on the planetary semimajor axis. One outcome, referred to as Outcome I, occurred for planets orbiting closer to the disc than to the star, namely, when $0.4 \lesssim a_p/a_{\rm in} \lesssim 1$ (Section 6.5.1). In this case, the disc developed a clear depletion in its surface density at the location of the secular resonance, in addition to a long, one-armed spiral density wave beyond that region. At the same time, the planetary orbit circularised due to resonant friction (more on this later) as it precessed under the action of the disc gravity. The features of the gap, such as its depth and width, are somewhat similar to those seen in Chapter 5, being asymmetric and depleted partially (by about a factor of two relative to the initial density profile). As to the spiral, I found that its pattern speed is to a good approximation equal to the planetary precession rate. The second outcome, referred to as Outcome II, occurred for planets situated closer to the central star than to the disc, namely, when $a_p \lesssim 0.4 a_{\rm in}$ (Section 6.5.2). In this case, the evolution was strikingly different than that in the previous regime. Namely, the disc remained coherent and did not develop any potentially observable structure. This happened as a result of

suppressed planetesimal eccentricities at and beyond the location where the resonance condition was met. Additionally, in this regime, the planet's eccentricity remained constant as its longitude of pericentre precessed. It was also shown that the transition between Outcomes I and II is gradual with variations in planetary semimajor axis, and that there exists an intermediate stage between the two (Section 6.5.3). Based on the simulations performed in this chapter with $a_{\rm in} = 30$ au (see Table H.1), I found that the critical value of a_p at which the transition from Outcome II to I occurs is $a_p^c \sim 12$ au.

The inhibition of eccentricity excitation at the site where the resonance condition is met in Outcome II is rather interesting and counter-intuitive. At the moment of writing, however, I do not have a satisfying explanation as to why this happens, but plan on further investigation in the near future. At this stage, however, it is clear that this behaviour is somehow related to the disc's non-axisymmetric component of self-gravity, since when the latter is switched off – as in the nominal N-ring simulations described in Sections 6.3 and 6.4 – the expected behaviour at the secular resonance (i.e., $e(a_{\rm res}) \to 1$) is recovered for all values of a_p; see Section 6.3. This leads me to speculate that perhaps it is the relative strength of the planet's non-axisymmetric potential with that of the disc that dictates – partly or completely – which outcome occurs for a given system. To assess the validity of this line of thought, I performed the following exercise. I simulated the evolution of both models `A` and `B` – which, I remind, exemplify Outcomes I and II, respectively (Sections 6.5.1 and 6.5.2) – but this time by modifying *by hand* the magnitude of the A_{jk} terms in Equation (6.9) that represent the non-axisymmetric perturbations amongst the disc rings (i.e., for all but $j = 0$ and $k = 0$). Doing so, I found that the evolution of `Model A` leads to Outcome II (and not I) if the terms $A_{j\neq0,k\neq0}$ are scaled up by a factor of ~ 5. Conversely, when the terms $A_{j\neq0,k\neq0}$ were scaled down by a factor of ~ 5, `Model B` resulted in Outcome I (and not II).

Further support to this line of thought can be sought as follows. Consider the ratio of the non-axisymmetric gravity experienced by planetesimals due to the planet $B_p(a)$ (Equation 5.6) to that due to the disc, which I denote by $B_d(a)$. For simplicity, let us further assume that the disc is apse-aligned and characterised by an eccentricity profile forced by the planet alone, $e_d(a) \propto 1/a$ (Equation 5.13). Then, one can neglect the evolution of the disc rings and calculate the ratio $B_p(a)/B_d(a)$ using only the non-diagonal terms of the matrix $\mathbf{A}$ (Equation 6.4). Doing such a calculation for the planet–disc parameters considered in this chapter, I found that the disc and the planet cancel each other's non-axisymmetric perturbations at the exact location where the

resonance condition is met for $a_p \approx 17$ au – which, surprisingly, is comparable to the critical value of $a_p^c \sim 12$ au reported in Section 6.5. Below and above this value of a_p, the cancellation occurs at a semimajor axis smaller and larger than the resonance location, respectively. This suggests that perhaps Outcome II is related to the fact that the disc's non-axisymmetric perturbations 'inhibit' planetesimal excitation at the resonance as it dominates or, perhaps, cancels the net non-axisymmetric perturbations around that region. A more detailed calculation of $B_d(a)$ is necessary to make a definitive statement. If true, however, this suggests the presence of planet- and disc-dominated regimes in terms of the planetesimal eccentricity excitation, similar to the case of precession rates discussed in detail in Section 5.3. Interestingly, such a classification has been made before for planetesimal dynamics in stellar binaries by Silsbee & Rafikov (2015a). Finally and admittedly, the discussion presented here does not constitute a definite explanation for the transition between Outcomes I and II, but rather a suggestive one which deserves further investigation.

Apart from being of academic interest, a full understanding of the causes behind the outcomes found in the full N-ring simulations is also essential for refining my predictions in Chapter 5 for the combinations of planet and disc parameters that could explain the observed structures in HD 107146, HD 92945, and HD 206893. For now, however, it is obvious that the introduction of the disc non-axisymmetric potential imposes additional constraints to the parameter space identified as 'allowed' in Figures 5.8 and 5.13 for HD 107146 and HD 92945, respectively. The most obvious constraint follows from the emergence of Outcome II: since in this case the disc does not develop a gap, the parameter spaces in Figures 5.8 and 5.13 corresponding to $a_p \lesssim a_p^c \sim 0.4 a_{\rm in}$ should be ruled out. Additionally, based on the nominal simulations results presented in Section 6.3, the values of m_p and M_d in the remaining portion of the parameter space should be modified as well. First, as we saw, the disc non-axisymmetric torque on the planet shifts the location of the secular resonance to a value larger than $a_{\rm res} = 70$ au (Section 6.4.3). To correct for this shift, the ratio M_d/m_p should be larger than that in Chapter 5 by a factor of $\sim 2-3$; see Figure 6.11. This could be done, for instance, by keeping M_d unchanged and adopting values of m_p smaller by $\sim 2-3$ than those in Figures 5.8 and 5.13. Doing so, however, would extend the timescales τ (since $\tau \propto 1/m_p$, Equation 5.15), in addition to the delay in $e(a_{\rm res})$ hitting unity due to the effects of the disc's non-axisymmetric gravity; see e.g. Section 6.4.4 and Figure 6.12. However, it is unlikely that this would affect the parameter space in Figures 5.8 and 5.13 any further, since the gap in Outcome I is already well developed by $t \sim 2\ \tau_{\rm sec}$, which is generally much smaller than τ; see e.g. Section 6.5.1. Given this, it is perhaps

better to use $\tau_{\rm sec}$ as a metric[9], rather than τ. I plan on further investigating the points raised here in the near future.

Here, it is also worth pointing out that the HD 92945 disc appears to feature an asymmetric gap, with potential spiral features (Marino et al., 2019) – somewhat akin to the disc morphology of Outcome I. It would be interesting to investigate in the future whether my model can reproduce the ALMA images of that disc (Figure 2.6). Additionally, another potential application of my model would be to the disc of HD 141569 (Perrot et al., 2016), which also features a 'gap+spiral' structure (see Section 2.3.1) as in Outcome I. The applicability of my model, however, is less clear in this case, mainly because of the presence of gas in the system (Zuckerman et al., 1995). Finally, I have also checked that the statements made in Section 5.7.2.2 of Chapter 5 regarding HD 206893 are not affected by the introduction of the disc non-axisymmetric potential. Namely, if the HD 206893 disc is indeed $\approx 170 M_{\oplus}$, then the brown dwarf observed in that system can sculpt a gap in the disc. This happens such that the brown dwarf–debris disc system evolve according to the intermediate outcome described in Section 6.5.3.

Another important finding in this chapter, as already pointed out above, is that secular planet–debris disc interactions can lead to the damping of the planet's eccentricity, without affecting its orbital semimajor axis. As shown in Sections 6.3 and 6.5, this process of resonant friction ensues from the gravitational coupling of the planet and the disc at a distance via apsidal secular resonances (Tremaine, 1998), without the need for the planet to be crossing the disc along its orbit. This is distinct from other widely cited circularising processes in the literature – namely, scattering of planetesimals (e.g. Pearce & Wyatt, 2015) and dynamical friction (e.g. Takeda, 1988; Kominami & Ida, 2002; Eriksson et al., 2018) – and should be considered as a viable process to circularise planetary orbits. By running a suite of nominal N-ring simulations, i.e., by ignoring the non-axisymmetric component of the disc self-gravity, I also showed that the simulation results agree very well with the theoretical expectations as derived by Tremaine (1998) within the same limit. To the best of my knowledge, this constitutes the first ever numerical verification of the theoretical results of Tremaine (1998). In the future, it would be interesting to derive analytical expressions for the decay rate of the planetary eccentricity by accounting for the full secular gravitational potential of the disc. Such an exercise would also lend more predictive power to the evolution observed in Outcome I (Section 6.5.1), both in terms of predicting the decay rate of e_p

[9]In retrospect, this comment also applies to the results of Chapter 5.

and the timescale associated with the excitation of planetesimal eccentricities at the resonance (similar to the analysis of the nominal simulations in Section 6.3).

Here, it is also worthwhile to point out that a previous study of planet–debris disc interactions might have stumbled upon evidence of resonant friction, without a rigorous identification of the process. In particular, in N-body simulations of planets and debris discs – where only the back reaction of the disc was accounted for and not the disc self-gravity – Pearce & Wyatt (2015) found that the planet is often circularised without undergoing any migration. According to their figure 5, this decay follows an exponential behaviour, especially at late times (see Stage 3 in their figure). Pearce & Wyatt (2015) interpreted this phenomenon as resulting from planetesimal scattering. While this might indeed be the case at early times when their planet is highly eccentric ($e_p \sim 0.5$) and thus crosses the disc when at its apocentre, it is very much likely that the decay following the shrinking of the planetary apocentre to values smaller than the inner disc edge results from resonant friction. This speculation could be tested by comparing the simulation data of Pearce & Wyatt (2015) with the decay rate of Equation (6.18). If true, however, this implies the presence of a secular resonance in the disc in Pearce & Wyatt (2015), without which resonant friction does not occur (Section 6.4.2). This is indeed the case in Pearce & Wyatt (2015), but their resonance condition is $A_p(a) = \dot{\varpi}_p$, i.e., Equation (6.23) with $A_d(a) = 0$ – see e.g. their figure 2. This also complements the explanation provided in Pearce & Wyatt (2015) regarding the appearance of a gap within the disc in their simulations.

The occurrence of resonant friction in interactions between planets and discs also highlights an important caveat related to the dynamical modelling of debris discs in general. My results show that resonant friction is robust over a broad range of planet–disc parameters provided that $a_p/a_{\rm in} \gtrsim 0.4$ (Outcome I; Section 6.5.1), and is also effective, in the sense that the ensuing eccentricity-damping timescale τ_D can be less than around the typical ages of observed debris discs – see e.g. Equation (6.21) and Figure 6.9. In the extreme limit, this leads to the counter-intuitive possibility for a planetary system to be comprised of a planet with an eccentricity which is incompatible – based on massless disc models – with the radial substructure of its debris disc. This is an interesting possibility which should be considered when interpreting observed disc structures. For instance, the incompatibility of a planet's current eccentricity with the disc structure that it supposedly sculpted could be used to reveal the dynamical past of the planet itself, e.g. the initial eccentricity along with the details of its decay. This could also be useful to indirectly measure the total mass of gapped debris discs. This is because the decay rate depends on the planet's mass (Equation 6.20), which –

provided the gap location is constrained – can be directly related to the disc mass (e.g. Equation 5.18).

This said, however, a word of caution is necessary at this stage: my simulation results are valid only to second order in eccentricities e (Section 6.2; see also Chapter 5). Thus, one concern is that the strength of resonant friction would in reality be limited when $e(a_{\rm res}) \to 1$ and the effects of higher-order terms in the secular disturbing function become non-negligible (Section 5.8.3). This deserves further investigation in the future. Nevertheless, such details are unlikely to negate the main message here: a planet could sculpt an observable radial structure (e.g. a gap and/or spiral pattern) in an external debris disc, and yet have its orbital eccentricity reduced below the initial value.

Finally, I close this chapter by noting that the flexibility of the N-ring code could also be exploited to model a range of debris disc scenarios in the future: e.g., discs with planets on inclined orbits, discs in multi-planet systems, or even discs perturbed by migrating planets.

Chapter 7

Conclusion

This chapter concludes the dissertation by first summarising the key research findings (Section 7.1), and then proposing directions of possible future work (Section 7.2).

7.1 Summary

In this dissertation, I have presented a theoretical investigation into the dynamical role of disc (self-)gravity in the long-term, secular interactions between planets and debris discs. This work would have been less relevant if not for the development of new observatories such as ALMA in the last decade, which have allowed for the high-resolution imaging of debris discs and the unprecedented characterisation of their substructures, such as gaps and spiral arms. Analogous to the studies of the asteroid and Kuiper belts in our own solar system, investigating the structure of debris discs can provide unique insights into the architecture and dynamical history of exo-planetary systems. For any accurate interpretation of debris disc observations, however, a model that is based on an understanding of the various processes affecting the debris disc's evolution, and thus its structure, is necessary – with planet–disc interactions being one of them. This rationale – coupled with the possibility that debris discs could be as massive as tens (if not hundreds) of Earth masses – provided much of the motivation for the investigation undertaken in this dissertation.

7.1.1 Softened Potentials of Astrophysical Discs

One of the key challenges of celestial mechanics and dynamical astronomy is the characterisation of the long-term, secular dynamical effects of self-gravitating discs, both on the constituents of the disc as well as external objects (e.g. a planet). By and

large, this follows from the fact that the classical Laplace–Lagrange theory for planetary orbits is mathematically ill-posed if and when applied to continuum discs. Thus, before diving into the problem of secular interactions between planets and self-gravitating debris discs, I first conducted a comparative study of the methods existing in the literature for the computation of the orbit-averaged gravitational potential due to a flat disc. This was the subject of Chapter 4. The main aim of this chapter was to investigate the questions of how well and under what conditions can the methods relying on potential softening – i.e., spatially smoothing the Newtonian point-mass potential – reproduce the expected secular dynamics driven by the unsoftened disc potential. To answer these questions, I focused on the four different softening prescriptions that have been previously proposed in the literature – namely, those in Tremaine (1998), Touma (2002), Hahn (2003), and Teyssandier & Ogilvie (2016) – and compared the resultant dynamical behaviour with that obtained based on the unsoftened method of Heppenheimer (1980).

Analysing the secular dynamics in several disc models with different surface density and eccentricity profiles, I arrived at the following conclusions. First, the softening prescriptions of both Touma (2002) and Hahn (2003) correctly reproduce the unsoftened calculations when taken in the limit of zero softening. In the same limit, on the other hand, it was found that the results ensuing from the softening prescription of Tremaine (1998) converge to the unsoftened calculations only approximately, with quantitative differences being on the order of $\sim 20 - 30$ per cent. As to the softening prescription of Teyssandier & Ogilvie (2016), it was found that it leads to the divergence of the disc potential, and thus does not reproduce the expected results.

To understand this variation of outcomes, I developed a generalised Laplace–Lagrange theory for flat discs which is applicable for a wide variety of softening prescriptions. This framework in itself represents a standalone result of Chapter 4. The main takeaway from this calculation is the following: for any form of softening, the development of the disc's orbit-averaged potential must be performed rigorously. More specifically, a direct replacement of the Laplace coefficients appearing in the classical Laplace–Lagrange theory is not sufficient for obtaining a well-behaved disturbing function for self-gravitating discs. To further validate this point, I showed that an accurate implementation of the softening prescriptions suggested in Tremaine (1998) and Teyssandier & Ogilvie (2016) does lead to the recovery of the expected dynamical behaviour in the limit of zero softening.

Finally, I examined the implications of potential softening for numerical treatments of self-gravitating discs, i.e., when discs are treated as a series of $N \gg 1$ massive rings

interacting with each other via softened forms of gravity. I found that such discretised treatments of discs must obey important constraints. Namely, a fine numerical sampling is required to ensure that the disc's secular potential is captured properly. Indeed, for a given value of the softening parameter $\varsigma \ll 1$, the minimum number of modelled rings should be $N \sim \varsigma^{-\chi}$, with χ ranging from ~ 1.5 to 2 depending on the adopted distribution of surface density. Thus, a relatively large number of disc rings is required to ensure that the correct secular behaviour is captured by such N-ring calculations, particularly when ς is small. This finding among others provides the bedrock upon which the numerical tool developed later in Chapter 6 is built.

7.1.2 Eccentric Planet–Debris Disc Interactions

In the second part of the dissertation, I examined the secular interaction between an eccentric planet and an external, self-gravitating debris disc – with both lying in the same plane. The aim of this part of the dissertation was to try and answer the four related questions posed in the Introduction (Section 1.3), which I reproduce here:

1. What is the dynamical role of disc (self-)gravity in planet– debris disc interactions?
2. What are the observational signatures of long-term interactions between planets and *massive* debris discs?
3. Can we infer the presence and evolution of planets in systems that might otherwise remain undetected?
4. What are the insights that we can draw about the masses of debris discs based on planet-disc interactions?

In order to explore these questions, I approached the problem of planet–debris disc interactions in a progression of complexity. In Chapter 5, I first approached the problem using a simplified analytical model based on the unsoftened treatment of disc gravity discussed in Chapter 4. The model was simplified in the sense that it only accounted for the axisymmetric components of both the disc self-gravity and its back reaction on the planet, ignoring the corresponding non-axisymmetric components. I found that even when the debris disc is less massive than the planet, the system may feature secular apsidal resonances within the disc – contrary to what may be naively expected. The result of this is that there are two dynamical regimes for the planetesimals: (1) a planet-dominated regime in the inner disc parts, where planetesimals by and large evolve as if the disc were massless; and (2) a disc-dominated regime in outer disc parts,

where the disc affects the dynamics significantly. The secular resonance – which causes planetesimal eccentricities to be significantly excited – occurs at the transition between these two regimes.

Based on this outcome, I proposed that gaps in debris discs, such as those around HD 107146 and HD 92945, could be produced by the interactions of the debris disc with a yet-undetected planet interior to it (and not within the gap itself). I also characterised the dependence of the properties of the secular resonances – i.e., locations, timescales, and widths – on the planet and disc parameters, finding that the mechanism is robust over a broad region of parameter space. As exemplary cases, I applied my results to HD 107146 and HD 92945, and demonstrated how my results can be used to set constraints on the planet–disc parameters that could explain the observations. By investigating the secular evolution of planet–debris disc systems using some of these parameters, I showed that the proposed mechanism readily produces $\mathcal{O}(10)$ au wide non-axisymmetric gaps. I also showed that this happens over timescales comparable to the period of planetary precession cycle (driven by the disc gravity); which, typically, is on the order of tens of megayears. Finally, I demonstrated that my results may be used to set constraints on the total masses of gapped debris discs. As an example, based on the properties of the observed gap in the HD 206893 disc and the directly imaged companion interior to it, I inferred a value of $\approx 170 M_{\oplus}$ for the total disc mass – assuming the gap is sculpted by the mechanism proposed here.

In Chapter 6, I continued and expanded on the investigation initiated in the previous chapter by accounting for the full secular gravitational potential of the debris disc, i.e., both axi- and non-axisymmetric components. This was achieved by the aid of a numerical tool – the "N-ring model" – which I developed here based on the results of Chapter 4 concerning the softening prescription of Hahn (2003). The model treats the disc as a series of gravitating rings, each with prescribed spacing (justified by the secular approximation), that interact both with each other and the planet via softened gravity. To build a comprehensive understanding of the dynamical role of the disc's non-axisymmetric perturbations, I tackled the problem in two steps: first by allowing for the non-axisymmetric torque on the planet but not on the disc particles themselves (Section 6.3), and then accounting for the full disc (self-)gravity (Section 6.5). This allowed for the development of approximate quantitative explanations for several (but not all) dynamical effects seen in my simulations with the full disc (self-)gravity.

The main result of Chapter 6 was that the secular evolution of the planet and the debris particles followed one of two regimes, depending on the orbital semimajor axis of the planet relative to the disc's inner edge. For planets orbiting closer to

the disc than to the star, the disc developed two features at late times: (i) a gap at the location of the secular resonance; and (ii) a long, one-armed spiral density wave beyond that region. It was also shown that the pattern speed of the spiral wave is equal to the precession rate of the planetary orbit. In this regime, the planetary orbit circularised over time due to a process known as resonant friction, which does not require the planet to cross the disc along its orbit. The details of this process have been elucidated using a simplified model of the planet–disc coupling in Section 6.3. In the other regime, which occurred for planets orbiting closer to the star than to the disc, the disc maintained a coherent structure, i.e., without developing any discernible structure, and the planetary orbit only precessed, i.e., without any changes in its eccentricity. I also found that the transition between these two regimes occurs via an intermediary stage, with the evolution of the planet–disc system showing parallels with either of the two regimes. Finally, the theoretical and observational implications of these results were discussed, both in the specific context of Chapter 5 and the general contexts of dynamical modelling and interpretation of debris discs.

In summary, if there is one takeaway from the work presented here, it is this: the gravitational potential of debris discs can have a notable effect on the secular evolution of both its constituent debris particles and planets orbiting it, and this even if the disc is less massive than the planet. The inclusion of disc self-gravity in studies of planet–debris disc interactions should be considered as it may impose modifications on the orbital parameters and masses, if not numbers, of planets invoked to reproduce observations.

7.2 Future Work

The investigation undertaken in this dissertation has uncovered some important – and previously neglected – aspects of interactions between planets and debris discs. However, I humbly admit that this work brings to light but a small aspect of the problem. Indeed, there are multitude of ways in which the work presented here can and should be extended. Below, I outline some potentially fruitful avenues to this end, apart from those already discussed at the end of Chapter 6, i.e., Section 6.6.

7.2.1 Collisional Depletion of Planetesimals

Throughout this dissertation, I modelled the debris disc as an ensemble of collisionless planetesimals. In practice, however, collisions are inevitable and can themselves affect

the disc evolution. Indeed, once the disc is sufficiently stirred, planetesimals collide and break up into smaller fragments, initiating a collisional cascade (Section 2.2). In this process, colliding planetesimals are gradually ground to dust until they are removed from the system by radiation effects. This causes the total disc mass to collisionally deplete over time (e.g. Equation 2.12), attenuating the gravitational role of the disc. This begs the question: how would the coupled dynamical and collisional evolution of debris particles affect the disc morphology?

Admittedly, this is a wide-open and challenging avenue of research, as it would require e.g. adapting existing collisional models – which can already handle perturbations due to planets – to account for the effects of disc gravity (e.g., the SMACK code of Nesvold et al. (2013), or the ACE code of Krivov et al. (2005); Sende & Löhne (2019)). However, a small step towards addressing the question above can be taken by introducing some modifications into the N-ring model employed in this dissertation. For instance, one can use the semi-analytical collisional model of Wyatt et al. (2010) to compute the collisional rate of a given particle in the disc (based on the dynamical state of the disc), and thus estimate how much mass is lost by a given particle at a given time (see e.g. their equations (33) and (34)). This approach could easily be integrated into my N-ring model (e.g., by updating the coefficients of the matrix $\mathbf{A}$ at every time step of the simulation, Section 6.2).

Speaking strictly within the context of Chapters 5 and 6, I expect collisions to preferentially deplete the disc surface density around the secular resonance (where $e \to 1$ and relative velocities are high) – in addition to the dynamical depletion due to the resonance. This may enhance the gap depths arising from my collisionless model. Collisional evolution may also contribute to widening the gaps resulting from my model. This can be understood as follows: as the total disc mass is depleted over time, the system's precession frequencies get altered, affecting the location of the secular resonances in a time-dependent way[1] (see also e.g., Heppenheimer, 1980; Ward, 1981; Nagasawa & Ida, 2000). If one naively assumes that the depletion is uniform across the disc (e.g., all disc rings lose mass at the same rate), the results of Figure 5.5 would then suggest that the resonance would sweep through the disc outwards as M_d decreases, potentially producing a wider gap than in my model. This is because planetesimal eccentricities could now be excited over a larger range in semimajor axis. Of course, this possibility is subject to the adiabaticity condition that the disc mass decays on a timescale that is longer than the secular dynamical timescales. If true, however, this

[1] I note that this could also happen if the planet migrates, either inwards or outwards, due to some physical process not considered here.

could be important, e.g., for the HD 107146 disc, for which my collisionless model produces gaps that are narrower than observed. Additionally, the shape of the resulting gap would provide information on the initial and final disc masses, along with the history of mass loss. I plan on exploring these avenues in future work.

7.2.2 Extension to Mutually Inclined Planets and Debris Discs

Another assumption of the work presented here is the coplanarity of the planetary systems. Such an assumption is expected to be valid for the vast majority of well studied multi-planetary systems (e.g. those discovered by the Kepler spacecraft, Winn & Fabrycky (2015)), including the solar system, which are statistically consistent with having low mutual inclinations (i.e., $\lesssim 5°$). Given this, it is perhaps natural to assume that the debris discs are also coplanar in such systems (e.g. due to dissipation within the primordial gaseous disc). However, to date, there are several systems known to harbour mutually inclined planets (e.g. Mills & Fabrycky, 2017; Xuan & Wyatt, 2020), and several others which are comprised of debris discs revealing vertical structures such as warps (see e.g. Section 2.3.3). The most notable example, to my mind, is the β Pic system. As discussed in Chapter 2, this system is comprised of two nearly coplanar planets (Lagrange et al., 2010; Nowak et al., 2020) orbiting interior to a warped debris disc (e.g. Heap et al., 2000, see also Figure 2.5(e)). Given this, it would be interesting to relax the coplanarity assumption inherent in the N-ring model of Chapter 6, and study the evolution of mutually inclined planet–debris disc systems in a future work – even if composed of a single planet for a start (see Section 5.7.3 for a discussion concerning multi-planet systems). This could be readily achieved by generalising the techniques developed in Chapter 4 (namely, Appendix A) for the calculation of the disturbing function due to softened discs to cases where the rings are mutually inclined (see also, Hahn 2003).

Generally speaking, I believe that a small but nonzero relative inclination (e.g., $\lesssim 5°$) between the planet and disc particles would not affect my results for eccentricity dynamics described in Chapters 5 and 6 (see also, Pearce & Wyatt, 2014). This is because within the context of Laplace–Lagrange theory, the evolution of eccentricities e and inclinations I are decoupled from each other when $e, I \ll 1$, and thus the eccentricity and inclination degrees of freedom may be treated in isolation (Murray & Dermott, 1999). This said, however, it is possible for planetesimal inclinations – similar to eccentricities – to be excited significantly at *inclination resonances* (e.g., Hahn, 2003, 2007), where the precession rates of both planet's and planetesimal's longitudes of ascending node are commensurate. In principle, this could happen when

the planet is initially inclined with respect to a razor-thin disc, or when the planet lies in the mid-plane of a puffed-up disc that is populated by planetesimals with nonzero inclination dispersion.

The occurrence of such a secular inclination resonance could affect the disc morphology in various ways. Firstly, if the disc is face-on, similar to HD 107146, and the planet is both eccentric and inclined, it would be possible for a single planet to carve two or a single but broader gap in the disc (if massive enough), depending on the locations of the eccentricity and inclination resonances. If the planet is circular but inclined, on the other hand, a face-on disc could still feature a gap due to the inclination resonance, provided that planetesimal inclinations are excited significantly at and around the site of the resonance. If true, this would be particularly appealing, given the fact that identifying vertical structures in nearly face-on debris discs is difficult (Hughes et al., 2018). Secondly, if the disc is nearly edge-on, the secular inclination resonance could sculpt a warp within the disc akin to that observed in β Pic. This could also be accompanied with a resonant friction in terms of the planetary inclination (e.g. Ward & Hahn, 2003; Hahn, 2007) – similar to the eccentricity case identified in Chapter 6. Exploring these avenues in a future work would be helpful to not only characterise the observational signatures resulting from planet–debris disc interactions, but also to be able to link such signatures to the total masses of debris discs.

7.2.3 Stirring in Debris Discs

Throughout this dissertation, I only focused on the dynamical evolution of debris discs which are less massive than the planets. This was motivated by the fact that such systems readily establish a secular resonance within the disc, leading to the formation of a gap such as those observed in HD 107146, HD 92945, and HD 206893. However, a natural question that arises is the following: how would planetesimals comprising a disc which is comparable to or more massive than the planet evolve? As I discuss below, such a situation may have important consequences for stirring of debris discs.

Results of Chapters 5 and 6 show that planetesimal dynamics in the outer parts of the disc is dominated by the disc's gravity, rather than that of the planet. The transition from planet- to disc-dominated regimes occurs at the site of the secular resonance. Given this, one can imagine a scenario whereby the secular resonance is designed to occur somewhere in between the planetary orbit and the disc's inner edge (but not within the disc), and so planetesimal dynamics would be disc-dominated throughout the whole disc. According to Equation (5.18), the disc-to-planet mass ratio

required for this to occur would be

$$\frac{M_d}{m_p} \gtrsim 3\left(\frac{a_p}{20 \text{ au}}\right)^2. \tag{7.1}$$

Provided that this condition is satisfied, i.e., $a_p \lesssim a_{\rm res} \lesssim a_{\rm in}$, planetesimals throughout the whole disc could evolve similar to those at large distances in Outcome II of Chapter 6. In other words, planetesimal eccentricities would be trapped to small-amplitude oscillations in time, with an average value which is (roughly) independent of semimajor axis (e.g. panel a of Figure 6.17). At the same time, planetesimals would remain roughly apse-aligned with respect to each other (and the planet, see e.g. panel b of Figure 6.17). This dynamical behaviour would translate into relative velocities of planetesimals throughout the whole disc that are lower than what would be expected from massless disc models (e.g. Mustill & Wyatt, 2009). In the extreme limit, this could lead to the counter-intuitive result that a relatively massive disc orbiting a single planet would produce less dust than if the disc were modelled as massless, as if the disc were gravitationally rigid to planetary perturbations. This hypothesis deserves further investigation, as it could at least impose modifications on the masses and orbital parameters of planets invoked to reproduce observed levels of stirring. Conversely, if the planets are detected, the results of such a study would be helpful in constraining the total masses of debris discs.

This said, however, it is important to note that the above hypothesis would not necessarily hold true if (i) the secular resonance occurs at $a_{\rm res} \gtrsim a_{\rm in}$ (and not at $a_{\rm res} \lesssim a_{\rm in}$), and (ii) planetesimals throughout the whole disc evolve as in the outer disc parts of Outcome I of Chapter 6. Instead, when these two conditions are met, the evolution could lead to the possibility of sculpting a long-lived spiral pattern that extends from the inner to the outer edge of the disc (see e.g. Figures 6.14(a),(b)), without an observable gap (see also Hahn, 2003, 2008). This is because according to Chapter 5, the width of a secular resonance occurring at $\approx a_{\rm in}$ would be very narrow. Interestingly, this would in principle also shorten the timescale at which the planet's eccentricity damps since $\tau_D \propto a_{\rm res}^{15/2}$ (Equation 6.21). If this occurs, then it would be natural to expect a (nearly) circular planet to be orbited by a disc characterised by a long-lived spiral density wave, a possibility which is otherwise counter-intuitive if the disc were massless. I plan on exploring these avenues in future work, once a more complete (semi-)analytic understanding of the role of the non-axisymmetric component of disc self-gravity is developed (see the detailed discussion in Section 6.6).

* * *

I leave the reader with a quote by Carl Sagan (1997): "There is much that science doesn't understand, many mysteries still to be resolved. In a Universe tens of billions of light years across and some ten or fifteen billion years old, this may be the case forever." In the vastness of space and the immensity of time, it is my joy to have contributed – even if ever so slightly – to the field of planetary science and I look forward to the surprises that the Universe holds in store.

* * *

References

ALMA Partnership et al., 2015, ApJ, 808, L3
Absil O., et al., 2006, A&A, 452, 237
Acke B., et al., 2012, A&A, 540, A125
Adachi I., Hayashi C., Nakazawa K., 1976, Progress of Theoretical Physics, 56, 1756
Allen C. W., 1973, Astrophysical quantities. London: University of London, Athlone Press
Andrews S. M., 2020, ARA&A, 58, 483
Andrews S. M., et al., 2016, ApJ, 820, L40
Andrews S. M., et al., 2018, ApJ, 869, L41
Apai D., et al., 2008, ApJ, 672, 1196
Apai D., Schneider G., Grady C. A., Wyatt M. C., Lagrange A.-M., Kuchner M. J., Stark C. J., Lubow S. H., 2015, ApJ, 800, 136
Armitage P. J., 2010, Astrophysics of Planet Formation. Cambridge University Press
Augereau J. C., Nelson R. P., Lagrange A. M., Papaloizou J. C. B., Mouillet D., 2001, A&A, 370, 447
Aumann H. H., et al., 1984, ApJ, 278, L23
Backman D. E., Paresce F., 1993, in Levy E. H., Lunine J. I., eds, Protostars and Planets III. Tucson, AZ: Univ. Arizona Press, p. 1253
Bai X.-N., Stone J. M., 2010, ApJ, 722, 1437
Bailey V., et al., 2014, ApJ, 780, L4
Ballering N. P., Rieke G. H., Gáspár A., 2014, ApJ, 793, 57
Batygin K., 2012, Nature, 491, 418
Batygin K., 2018, MNRAS, 475, 5070
Batygin K., Morbidelli A., Tsiganis K., 2011, A&A, 533, A7
Batygin K., Adams F. C., Brown M. E., Becker J. C., 2019, Phys. Rep., 805, 1
Beichman C. A., et al., 2006, ApJ, 652, 1674
Belyaev M. A., Rafikov R. R., 2011, Icarus, 214, 179
Beust H., Lagrange-Henri A. M., Vidal-Madjar A., Ferlet R., 1990, A&A, 236, 202
Beust H., et al., 2014, A&A, 561, A43
Biller B. A., et al., 2013, ApJ, 777, 160
Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition. Princeton University Press
Blum J., 2018, Space Sci. Rev., 214, 52
Boccaletti A., et al., 2019, A&A, 625, A21
Boley A. C., 2009, ApJ, 695, L53
Boley A. C., Morris M. A., Ford E. B., 2014, ApJ, 792, L27
Bonnefoy M., et al., 2017, A&A, 597, L7
Booth M., et al., 2013, MNRAS, 428, 1263
Booth M., et al., 2016, MNRAS, 460, L10
Booth M., et al., 2017, MNRAS, 469, 3200
Boss A. P., 1997, Science, 276, 1836
Boss A. P., 2011, ApJ, 731, 74

References

Bottke William F. J., Vokrouhlický D., Rubincam D. P., Nesvorný D., 2006, Annual Review of Earth and Planetary Sciences, 34, 157
Bowler B. P., 2016, PASP, 128, 102001
Bowler B. P., Nielsen E. L., 2018, Occurrence Rates from Direct Imaging Surveys. Springer International Publishing, p. 155, doi:10.1007/978-3-319-55333-7_155
Brouwer D., Clemence G. M., 1961, Methods of celestial mechanics. Academic Press: New York
Burns J. A., Lamy P. L., Soter S., 1979, Icarus, 40, 1
Burrows C. J., Krist J. E., Stapelfeldt K. R., WFPC2 Investigation Definition Team 1995, in American Astronomical Society Meeting Abstracts. p. 32.05
Carpenter J. M., et al., 2009a, ApJS, 181, 197
Carpenter J. M., Mamajek E. E., Hillenbrand L. A., Meyer M. R., 2009b, ApJ, 705, 1646
Carrera D., Gorti U., Johansen A., Davies M. B., 2017, ApJ, 839, 16
Carroll B. W., Ostlie D. A., 1996, An Introduction to Modern Astrophysics. Reading, MA: Addison-Wesley
Chambers J. E., 1999, MNRAS, 304, 793
Chapman C. R., Williams J. G., Hartmann W. K., 1978, Annual Review of Astronomy and Astrophysics, 16, 33
Chavez-Dagostino M., et al., 2016, MNRAS, 462, 2285
Chiang E., Fung J., 2017, ApJ, 848, 4
Chiang E. I., Goldreich P., 2000, ApJ, 540, 1084
Chiang E., Kite E., Kalas P., Graham J. R., Clampin M., 2009, ApJ, 693, 734
Cieza L. A., et al., 2016, Nature, 535, 258
Clampin M., et al., 2003, AJ, 126, 385
Consolmagno G. J., 1980, Icarus, 43, 203
Crifo F., Vidal-Madjar A., Lallement R., Ferlet R., Gerbaldi M., 1997, A&A, 320, L29
Currie T., Lisse C. M., Kuchner M., Madhusudhan N., Kenyon S. J., Thalmann C., Carson J., Debes J., 2015, ApJ, 807, L7
Cuzzi J. N., Hogan R. C., Paque J. M., Dobrovolskis A. R., 2001, ApJ, 546, 496
Daley C., et al., 2019, ApJ, 875, 87
Davydenkova I., Rafikov R. R., 2018, ApJ, 864, 74
Dawson R. I., Murray-Clay R. A., Fabrycky D. C., 2011, ApJ, 743, L17
DeMeo F. E., Carry B., 2013, Icarus, 226, 723
Delorme P., et al., 2017, A&A, 608, A79
Dent W. R. F., et al., 2014, Science, 343, 1490
Dermott S. F., Jayaraman S., Xu Y. L., Gustafson B. Å. S., Liou J. C., 1994, Nature, 369, 719
Dermott S. F., Grogan K., Durda D. D., Jayaraman S., Kehoe T. J. J., Kortenkamp S. J., Wyatt M. C., 2001, Orbital Evolution of Interplanetary Dust. Berlin: Springer, p. 569
Dick S. J., 1996, The biological universe: the twentieth-century extraterrestrial life debate and the limits of science. Cambridge University Press
Dodson-Robinson S. E., Veras D., Ford E. B., Beichman C. A., 2009, ApJ, 707, 79
Dohnanyi J. S., 1969, J. Geophys. Res., 74, 2531
Dominik C., Decin G., 2003, ApJ, 598, 626
Dominik C., Tielens A. G. G. M., 1997, ApJ, 480, 647
Dong J., Dawson R. I., Shannon A., Morrison S., 2020, ApJ, 889, 47
Draine B. T., Lee H. M., 1984, ApJ, 285, 89
Duncan M., Quinn T., Tremaine S., 1989, Icarus, 82, 402
Eddington A., 1928, The Nature of the Physical World. New York: Macmillan
Edgeworth K. E., 1943, Journal of the British Astronomical Association, 53, 181
Eiroa C., et al., 2013, A&A, 555, A11

Ercolano B., Pascucci I., 2017, Royal Society Open Science, 4, 170114
Eriksson L. E. J., Mustill A. J., Johansen A., 2018, MNRAS, 475, 4609
Ertel S., et al., 2014, A&A, 570, A128
Ertel S., et al., 2020, AJ, 159, 177
Faramaz V., et al., 2014, A&A, 563, A72
Feldt M., et al., 2017, A&A, 601, A7
Fontana A., Marzari F., 2016, A&A, 589, A133
Fraser W. C., Brown M. E., Morbidelli A., Parker A., Batygin K., 2014, ApJ, 782, 100
Friebe M. F., Pearce T. D., Löhne T., 2022, MNRAS, 512, 4441
Gaia Collaboration 2018, VizieR Online Data Catalog, p. I/345
Gáspár A., Psaltis D., Rieke G. H., Özel F., 2012, ApJ, 754, 74
Gáspár A., Rieke G. H., Balog Z., 2013, ApJ, 768, 25
Gladman B., Volk K., 2021, Annual Review of Astronomy and Astrophysics, 59, 203
Gladman B., Kavelaars J. J., Petit J.-M., Morbidelli A., Holman M. J., Loredo T., 2001, AJ, 122, 1051
Goldreich P., Sari R., 2003, ApJ, 585, 1024
Goldreich P., Tremaine S., 1979, AJ, 84, 1638
Goldreich P., Tremaine S., 1980, ApJ, 241, 425
Golimowski D. A., et al., 2011, AJ, 142, 30
Gomes R., Levison H. F., Tsiganis K., Morbidelli A., 2005, Nature, 435, 466
Grandjean A., et al., 2019, A&A, 627, L9
Graps A. L., Showalter M. R., Lissauer J. J., Kary D. M., 1995, AJ, 109, 2262
Greaves J. S., Holland W. S., Jayawardhana R., Wyatt M. C., Dent W. R. F., 2004, MNRAS, 348, 1097
Greaves J. S., et al., 2005, The Astrophysical Journal, 619, L187
Greenberg R., Wacker J. F., Hartmann W. K., Chapman C. R., 1978, Icarus, 35, 1
Gulati M., Saini T. D., Sridhar S., 2012, MNRAS, 424, 348
Gustafson B. A. S., 1994, Annual Review of Earth and Planetary Sciences, 22, 553
Güttler C., Blum J., Zsom A., Ormel C. W., Dullemond C. P., 2009, Geochimica et Cosmochimica Acta Supplement, 73, A482
Habing H. J., et al., 2001, A&A, 365, 545
Haffert S. Y., Bohn A. J., de Boer J., Snellen I. A. G., Brinchmann J., Girard J. H., Keller C. U., Bacon R., 2019, Nature Astronomy, 3, 749
Hahn J. M., 2003, ApJ, 595, 531
Hahn J. M., 2007, ApJ, 665, 856
Hahn J. M., 2008, ApJ, 680, 1569
Hamilton D. P., Grun E., Baguhl M., 1996, in Gustafson B. A. S., Hanner M. S., eds, Astronomical Society of the Pacific Conference Series Vol. 104, IAU Colloq. 150: Physics, Chemistry, and Dynamics of Interplanetary Dust. p. 31
Hayashi C., 1981, Progress of Theoretical Physics Supplement, 70, 35
Heap S. R., Lindler D. J., Lanz T. M., Cornett R. H., Hubeny I., Maran S. P., Woodgate B., 2000, ApJ, 539, 435
Heppenheimer T. A., 1980, Icarus, 41, 76
Holland W. S., et al., 2017, MNRAS, 470, 3606
Horanyi M., 1996, ARA&A, 34, 383
Hughes A. M., Wilner D. J., Andrews S. M., Williams J. P., Su K. Y. L., Murray-Clay R. A., Qi C., 2011, ApJ, 740, 38
Hughes A. M., et al., 2012, ApJ, 750, 82
Hughes A. M., Duchêne G., Matthews B. C., 2018, ARA&A, 56, 541
Ida S., Makino J., 1993, Icarus, 106, 210
Ida S., Bryden G., Lin D. N. C., Tanaka H., 2000, ApJ, 534, 428

References

Isella A., et al., 2016, Phys. Rev. Lett., 117, 251101
Isella A., Benisty M., Teague R., Bae J., Keppler M., Facchini S., Pérez L., 2019, ApJ, 879, L25
Jacobs V., Sellwood J. A., 2001, ApJ, 555, L25
Jalali M. A., Tremaine S., 2012, MNRAS, 421, 2368
Jeans J. H., 1919, Problems of cosmogony and stellar dynamics. Cambridge University Press
Jewitt D., Luu J., 1993, Nature, 362, 730
Jílková L., Portegies Zwart S., 2015, MNRAS, 451, 804
Johansen A., Oishi J. S., Mac Low M.-M., Klahr H., Henning T., Youdin A., 2007, Nature, 448, 1022
Johansen A., Youdin A., Mac Low M.-M., 2009, ApJ, 704, L75
Kaib N. A., White E. B., Izidoro A., 2018, MNRAS, 473, 470
Kalas P., Jewitt D., 1995, AJ, 110, 794
Kalas P., Graham J. R., Fitzgerald M. P., Clampin M., 2013, ApJ, 775, 56
Kalas P. G., et al., 2015, ApJ, 814, 32
Kasper M., Apai D., Wagner K., Robberto M., 2015, ApJ, 812, L33
Kataoka A., Tanaka H., Okuzumi S., Wada K., 2013, A&A, 557, L4
Kazandjian M. V., Touma J. R., 2013, MNRAS, 430, 2732
Kennedy G. M., 2020, Royal Society Open Science, 7, 200063
Kennedy G. M., Piette A., 2015, MNRAS, 449, 2304
Kennedy G. M., Wyatt M. C., 2010, MNRAS, 405, 1253
Kennedy G., Wyatt M., 2013, in Protostars and Planets VI Posters. (arXiv:1305.6607)
Kennedy G. M., Wyatt M. C., 2014, MNRAS, 444, 3164
Kennedy G. M., Marino S., Matrà L., Panić O., Wilner D., Wyatt M. C., Yelverton B., 2018, MNRAS, 475, 4924
Kenyon S. J., Bromley B. C., 2008, ApJS, 179, 451
Keppler M., et al., 2018, A&A, 617, A44
Kessler M. F., et al., 1996, A&A, 500, 493
Kimura H., Ishimoto H., Mukai T., 1997, A&A, 326, 263
Kocsis B., Tremaine S., 2011, MNRAS, 412, 187
Kominami J., Ida S., 2002, Icarus, 157, 43
Konishi M., et al., 2016, ApJ, 818, L23
Kóspál Á., et al., 2013, ApJ, 776, 77
Kowal C. T., Liller W., Marsden B. G., 1979, in Duncombe R. L., ed., Dynamics of the Solar System Vol. 81, Dynamics of the Solar System. p. 245
Kral Q., et al., 2017, The Astronomical Review, 13, 69
Kral Q., Marino S., Wyatt M. C., Kama M., Matrà L., 2019, MNRAS, 489, 3670
Krist J. E., et al., 2005, AJ, 129, 1008
Krivov A. V., 2010, Research in Astronomy and Astrophysics, 10, 383
Krivov A. V., Wyatt M. C., 2021, MNRAS, 500, 718
Krivov A., Kimura H., Mann I., 1998, Icarus, 134, 311
Krivov A. V., Sremčević M., Spahn F., 2005, Icarus, 174, 105
Krivov A. V., Löhne T., Sremčević M., 2006, A&A, 455, 509
Krivov A. V., Ide A., Löhne T., Johansen A., Blum J., 2018, MNRAS, 474, 2564
Kuiper G. P., 1951a, in Hynek J. A., ed., 50th Anniversary of the Yerkes Observatory and Half a Century of Progress in Astrophysics. p. 357
Kuiper G. P., 1951b, Proceedings of the National Academy of Science, 37, 1
Lagrange A. M., et al., 2010, Science, 329, 57
Lagrange A. M., et al., 2016, A&A, 586, L8
Lagrange A. M., et al., 2019, A&A, 621, L8
Laskar J., Boué G., 2010, A&A, 522, A60

Latter H. N., Ogilvie G. I., Rein H., 2017, preprint, (arXiv:1701.04312)
Lazzoni C., et al., 2018, A&A, 611, A43
Lebreton J., et al., 2012, A&A, 539, A17
Lee E. J., Chiang E., 2016, ApJ, 827, 125
Lee W.-K., Dempsey A. M., Lithwick Y., 2019, ApJ, 872, 184
Levison H. F., Agnor C., 2003, AJ, 125, 2692
Levison H. F., Duncan M. J., 1997, Icarus, 127, 13
Levison H. F., Morbidelli A., Van Laerhoven C., Gomes R., Tsiganis K., 2008, Icarus, 196, 258
Levison H. F., Thommes E., Duncan M. J., 2010, AJ, 139, 1297
Lin J. W., Chiang E., 2019, ApJ, 883, 68
Lissauer J. J., 1993, ARA&A, 31, 129
Lisse C. M., Beichman C. A., Bryden G., Wyatt M. C., 2007, ApJ, 658, 584
Lodders K., 2003, ApJ, 591, 1220
Long F., et al., 2018, ApJ, 869, 17
Lovell J. B., et al., 2021, MNRAS, 500, 4878
Lynch E. M., Lovell J. B., 2021, MNRAS,
Lyra W., Kuchner M., 2013, Nature, 499, 184
MacGregor M. A., et al., 2017, ApJ, 842, 8
MacGregor M. A., et al., 2018, ApJ, 869, 75
MacGregor M. A., et al., 2019, ApJ, 877, L32
Malhotra R., 1993, Nature, 365, 819
Malhotra R., 1998, Orbital Resonances and Chaos in the Solar System. San Francisco, CA: ASP, p. 37
Mannings V., Barlow M. J., 1998, ApJ, 497, 330
Mardling R. A., 2013, MNRAS, 435, 2187
Marino S., 2021, MNRAS, 503, 5100
Marino S., et al., 2018a, MNRAS, 479, 5423
Marino S., Bonsor A., Wyatt M. C., Kral Q., 2018b, MNRAS, 479, 1651
Marino S., Yelverton B., Booth M., Faramaz V., Kennedy G. M., Matrà L., Wyatt M. C., 2019, MNRAS,
Marino S., et al., 2020, Monthly Notices of the Royal Astronomical Society, 498, 1319
Marois C., Macintosh B., Barman T., Zuckerman B., Song I., Patience J., Lafrenière D., Doyon R., 2008, Science, 322, 1348
Marois C., Zuckerman B., Konopacky Q. M., Macintosh B., Barman T., 2010, Nature, 468, 1080
Marshall J. P., et al., 2014, A&A, 565, A15
Marzari F., Scholl H., 2000, ApJ, 543, 328
Matrà L., et al., 2017, ApJ, 842, 9
Matrà L., Wyatt M. C., Wilner D. J., Dent W. R. F., Marino S., Kennedy G. M., Milli J., 2019, AJ, 157, 135
Matsuo T., Shibai H., Ootsubo T., Tamura M., 2007, ApJ, 662, 1282
Matthews B. C., Krivov A. V., Wyatt M. C., Bryden G., Eiroa C., 2014a, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, Protostars and Planets VI. p. 521 (arXiv:1401.0743), doi:10.2458/azu_uapress_9780816531240-ch023
Matthews B., Kennedy G., Sibthorpe B., Booth M., Wyatt M., Broekhoven-Fiene H., Macintosh B., Marois C., 2014b, ApJ, 780, 97
Mayor M., Queloz D., 1995, Nature, 378, 355
Mayor M., et al., 2011, arXiv e-prints, p. arXiv:1109.2497
Meschiari S., 2014, ApJ, 790, 41
Meshkat T., et al., 2017, AJ, 154, 245

References

Meyer M. R., 2006, in Armus L., Reach W. T., eds, Astronomical Society of the Pacific Conference Series Vol. 357, The Spitzer Space Telescope: New Views of the Cosmos. p. 51 (arXiv:astro-ph/0701058)
Michel A., van der Marel N., Matthews B. C., 2021, ApJ, 921, 72
Millar-Blanchaer M. A., et al., 2015, ApJ, 811, 18
Milli J., et al., 2017a, A&A, 597, L2
Milli J., et al., 2017b, A&A, 599, A108
Mills S. M., Fabrycky D. C., 2017, AJ, 153, 45
Montesinos B., et al., 2016, A&A, 593, A51
Moór A., et al., 2017, ApJ, 849, 123
Morbidelli A., Raymond S. N., 2016, Journal of Geophysical Research (Planets), 121, 1962
Morbidelli A., Levison H. F., Tsiganis K., Gomes R., 2005, Nature, 435, 462
Morbidelli A., Levison H. F., Gomes R., 2008, The Dynamical Structure of the Kuiper Belt and Its Primordial Origin. University of Arizona Press, p. 275
Morbidelli A., Lunine J. I., O'Brien D. P., Raymond S. N., Walsh K. J., 2012, Annual Review of Earth and Planetary Sciences, 40, 251
Morfill G. E., Gruen E., 1979, Planet. Space Sci., 27, 1269
Moro-Martín A., Malhotra R., 2002, AJ, 124, 2305
Moro-Martín A., et al., 2007, ApJ, 658, 1312
Moro-Martín A., et al., 2015, ApJ, 801, 143
Morrison S. J., Kratter K. M., 2018, MNRAS, 481, 5180
Mouillet D., Larwood J. D., Papaloizou J. C. B., Lagrange A. M., 1997, MNRAS, 292, 896
Mouillet D., Lagrange A. M., Augereau J. C., Ménard F., 2001, A&A, 372, L61
Muñoz-Gutiérrez M. A., Pichardo B., Reyes-Ruiz M., Peimbert A., 2015, ApJ, 811, L21
Müller A., et al., 2018, A&A, 617, L2
Murray C. D., Dermott S. F., 1999, Solar system dynamics. Cambridge University Press
Mustill A. J., Wyatt M. C., 2009, MNRAS, 399, 1403
Mustill A. J., Wyatt M. C., 2012, MNRAS, 419, 3074
Nagasawa M., Ida S., 2000, AJ, 120, 3311
Nagasawa M., Ida S., Tanaka H., 2001, Earth, Planets and Space, 53, 1085
Najita J. R., Kenyon S. J., Bromley B. C., 2021, arXiv e-prints, p. arXiv:2111.06406
Nederlander A., et al., 2021, ApJ, 917, 5
Nesvold E. R., Kuchner M. J., 2015a, ApJ, 798, 83
Nesvold E. R., Kuchner M. J., 2015b, ApJ, 815, 61
Nesvold E. R., Kuchner M. J., Rein H., Pan M., 2013, ApJ, 777, 144
Nesvold E. R., Naoz S., Vican L., Farr W. M., 2016, ApJ, 826, 19
Nesvold E. R., Naoz S., Fitzgerald M. P., 2017, ApJ, 837, L6
Nesvorný D., Jenniskens P., Levison H. F., Bottke W. F., Vokrouhlický D., Gounelle M., 2010, ApJ, 713, 816
Nguyen M. M., De Rosa R. J., Kalas P., 2021, AJ, 161, 22
Nielsen E. L., et al., 2016, AJ, 152, 175
Norfolk B. J., et al., 2021, MNRAS,
Nowak M., et al., 2020, A&A, 642, L2
O'Brien D. P., Greenberg R., 2005, Icarus, 178, 179
O'Keefe J. A., 1976, Tektites and Their Origin. Elsevier Sci., New York
Olofsson J., Juhász A., Henning T., Mutschke H., Tamanai A., Moór A., Ábrahám P., 2012, A&A, 542, A90
Olofsson J., et al., 2018, A&A, 617, A109
Olofsson J., et al., 2022, MNRAS, 513, 713
Oudmaijer R. D., van der Veen W. E. C. J., Waters L. B. F. M., Trams N. R., Waelkens C., Engelsman E., 1992, A&AS, 96, 625

Paddack S. J., 1969, Journal of Geophysical Research (1896-1977), 74, 4379
Pan M., Sari R., 2005, Icarus, 173, 342
Pan M., Schlichting H. E., 2012, ApJ, 747, 113
Pan M., Wu Y., 2016, ApJ, 821, 18
Pan M., Nesvold E. R., Kuchner M. J., 2016, ApJ, 832, 81
Panić O., et al., 2013, MNRAS, 435, 1037
Patience J., et al., 2011, A&A, 531, L17
Pawellek N., 2016, PhD thesis, Friedrich-Schiller-Universität Jena
Pawellek N., Krivov A. V., 2015, MNRAS, 454, 3207
Pawellek N., Krivov A. V., Marshall J. P., Montesinos B., Ábrahám P., Moór A., Bryden G., Eiroa C., 2014, ApJ, 792, 65
Pearce T. D., Wyatt M. C., 2014, MNRAS, 443, 2541
Pearce T. D., Wyatt M. C., 2015, MNRAS, 453, 3329
Pearce T. D., Krivov A. V., Booth M., 2020, MNRAS, 498, 2798
Pearce T. D., Beust H., Faramaz V., Booth M., Krivov A. V., Löhne T., Poblete P. P., 2021, MNRAS, 503, 4767
Pecaut M. J., Mamajek E. E., 2016, MNRAS, 461, 794
Perrot C., et al., 2016, A&A, 590, L7
Peterson C., 1976, Icarus, 29, 91
Petrovich C., Wu Y., Ali-Dib M., 2019, AJ, 157, 5
Pilbratt G. L., et al., 2010, A&A, 518, L1
Plavchan P., Werner M. W., Chen C. H., Stapelfeldt K. R., Su K. Y. L., Stauffer J. R., Song I., 2009, ApJ, 698, 1068
Plummer H. C. K., 1918, An introductory treatise on dynamical astronomy. Cambridge University Press
Pollack J. B., Hubickyj O., Bodenheimer P., Lissauer J. J., Podolak M., Greenzweig Y., 1996, Icarus, 124, 62
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 2002, Numerical recipes in C++ : the art of scientific computing. Cambridge University Press
Quillen A. C., 2006, MNRAS, 372, L14
Quillen A. C., Faber P., 2006, MNRAS, 373, 1245
Radzievskii V. V., 1952, Azh, 29, 162
Rafikov R. R., 2004, AJ, 128, 1348
Rafikov R. R., 2005, ApJ, 621, L69
Rafikov R. R., 2011, ApJ, 727, 86
Rafikov R. R., 2013, ApJ, 764, L16
Read M. J., Wyatt M. C., Marino S., Kennedy G. M., 2018, MNRAS, 475, 4953
Reche R., Beust H., Augereau J. C., Absil O., 2008, A&A, 480, 551
Regály Z., Dencs Z., Moór A., Kovács T., 2018, MNRAS, 473, 3547
Reidemeister M., Krivov A. V., Stark C. C., Augereau J. C., Löhne T., Müller S., 2011, A&A, 527, A57
Rein H., Liu S. F., 2012, A&A, 537, A128
Ricci L., Carpenter J. M., Fu B., Hughes A. M., Corder S., Isella A., 2015, ApJ, 798, 124
Rieke G. H., Gáspár A., Ballering N. P., 2016, ApJ, 816, 50
Rigley J. K., Wyatt M. C., 2020, MNRAS, 497, 1143
Riviere-Marichalar P., et al., 2013, A&A, 555, A67
Roberge A., Weinberger A. J., Malumuth E. M., 2005, ApJ, 622, 1171
Rodet L., Beust H., Bonnefoy M., Lagrange A. M., Galli P. A. B., Ducourant C., Teixeira R., 2017, A&A, 602, A12
Sadakane K., Nishida M., 1986, PASP, 98, 685
Safronov V. S., 1991, Icarus, 94, 260

References

Sagan C., 1997, The demon-haunted world : science as a candle in the dark / Carl Sagan., 1st ballantine books ed. edn. Ballantine Books, New York
Schneider G., et al., 2006, ApJ, 650, 414
Schneider G., et al., 2014, AJ, 148, 59
Schneider G., et al., 2016, AJ, 152, 64
Schneider G., et al., 2018, AJ, 155, 77
Schüppler C., et al., 2015, A&A, 581, A97
Schüppler C., Krivov A. V., Löhne T., Booth M., Kirchschlager F., Wolf S., 2016, MNRAS, 461, 2146
Scicluna P., Rosotti G., Dale J. E., Testi L., 2014, A&A, 566, L3
Scotti J. V., et al., 1992, IAU Circ., 5434, 1
Sefilian A. A., Rafikov R. R., 2019, MNRAS, 489, 4176
Sefilian A. A., Touma J. R., 2019, AJ, 157, 59
Sefilian A. A., Rafikov R. R., Wyatt M. C., 2021, ApJ, 910, 13
Sende J. A., Löhne T., 2019, A&A, 631, A141
Shannon A., Clarke C., Wyatt M., 2014, MNRAS, 442, 142
Shannon A., Bonsor A., Kral Q., Matthews E., 2016, MNRAS, 462, L116
Shu F. H., Adams F. C., Lizano S., 1987, ARA&A, 25, 23
Sibthorpe B., Kennedy G. M., Wyatt M. C., Lestrade J. F., Greaves J. S., Matthews B. C., Duchêne G., 2018, MNRAS, 475, 3046
Silsbee K., Rafikov R. R., 2015a, ApJ, 798, 71
Silsbee K., Rafikov R. R., 2015b, ApJ, 808, 58
Silsbee K., Rafikov R. R., 2021, A&A, 652, A104
Simon J. B., Armitage P. J., Li R., Youdin A. N., 2016, ApJ, 822, 55
Smith B. A., Terrile R. J., 1984, Science, 226, 1421
Sridhar S., Touma J., 1999, MNRAS, 303, 483
Statler T. S., 2001, AJ, 122, 2257
Su K. Y. L., Rieke G. H., 2014, in Booth M., Matthews B. C., Graham J. R., eds, IAU Symposium Vol. 299, Exploring the Formation and Evolution of Planetary Systems. Cambridge University Press, pp 318–321, doi:10.1017/S1743921313008764
Su K. Y. L., et al., 2006, ApJ, 653, 675
Su K. Y. L., et al., 2009, ApJ, 705, 314
Tabeshian M., Wiegert P. A., 2016, ApJ, 818, 159
Takeda H., 1988, Progress of Theoretical Physics Supplement, 96, 196
Teyssandier J., Lai D., 2019, MNRAS, 490, 4353
Teyssandier J., Ogilvie G. I., 2016, MNRAS, 458, 3221
Teyssandier J., Terquem C., Papaloizou J. C. B., 2013, MNRAS, 428, 658
Thébault P., Augereau J. C., 2007, A&A, 472, 169
Thureau N. D., et al., 2014, MNRAS, 445, 2558
Tiscareno M. S., 2013, Planetary Rings. Springer, p. 309, doi:10.1007/978-94-007-5606-9_7
Touma J. R., 2002, MNRAS, 333, 583
Touma J. R., Sridhar S., 2012, MNRAS, 423, 2083
Touma J. R., Tremaine S., Kazandjian M. V., 2009, MNRAS, 394, 1085
Tremaine S., 1993, in Phillips J. A., Thorsett S. E., Kulkarni S. R., eds, Astronomical Society of the Pacific Conference Series Vol. 36, Planets Around Pulsars. pp 335–344
Tremaine S., 1998, AJ, 116, 2015
Tremaine S., 2001, AJ, 121, 1776
Tsiganis K., Gomes R., Morbidelli A., Levison H. F., 2005, Nature, 435, 459
Veras D., Jacobson S. A., Gänsicke B. T., 2014, MNRAS, 445, 2794
Vitense C., Krivov A. V., Kobayashi H., Löhne T., 2012, A&A, 540, A30

Vokrouhlický D., Bottke W. F., Chesley S. R., Scheeres D. J., Statler T. S., 2015, in , Asteroids IV. University of Arizona Press, pp 509–531, doi:10.2458/azu_uapress_9780816532131-ch027
Wang J. J., et al., 2015, ApJ, 811, L19
Ward W. R., 1981, Icarus, 47, 234
Ward W. R., 1989, ApJ, 336, 526
Ward W. R., Hahn J. M., 1998a, AJ, 116, 489
Ward W. R., Hahn J. M., 1998b, Science, 280, 2104
Ward W. R., Hahn J. M., 2000, in Mannings V., Boss A. P., Russell S. S., eds, Protostars and Planets IV. Tucson, AZ: Univ. Arizona Press, p. 1135
Ward W. R., Hahn J. M., 2003, AJ, 125, 3389
Watson C. A., Littlefair S. P., Diamond C., Collier Cameron A., Fitzsimmons A., Simpson E., Moulds V., Pollacco D., 2011, MNRAS, 413, L71
Weidenschilling S. J., 1977, Ap&SS, 51, 153
Weidenschilling S. J., 1980, Icarus, 44, 172
Weidenschilling S. J., Cuzzi J. N., 1993, in Levy E. H., Lunine J. I., eds, Protostars and Planets III. p. 1031
Weissman P. R., 1995, ARA&A, 33, 327
Werner M. W., et al., 2004, ApJS, 154, 1
Williams J. P., Cieza L. A., 2011, ARA&A, 49, 67
Williams J. P., Najita J., Liu M. C., Bottinelli S., Carpenter J. M., Hillenbrand L. A., Meyer M. R., Soderblom D. R., 2004, ApJ, 604, 414
Wilner D. J., Holman M. J., Kuchner M. J., Ho P. T. P., 2002, ApJ, 569, L115
Wilner D. J., Andrews S. M., Hughes A. M., 2011, ApJ, 727, L42
Winn J. N., 2018, Planet Occurrence: Doppler and Transit Surveys. Springer, p. 195, doi:10.1007/978-3-319-55333-7_195
Winn J. N., Fabrycky D. C., 2015, ARA&A, 53, 409
Wisdom J., 1980, AJ, 85, 1122
Wolszczan A., Frail D. A., 1992, Nature, 355, 145
Wu Y.-L., et al., 2016, ApJ, 823, 24
Wyatt M. C., 2003, ApJ, 598, 1321
Wyatt M. C., 2005a, A&A, 433, 1007
Wyatt M. C., 2005b, A&A, 440, 937
Wyatt M. C., 2006, ApJ, 639, 1153
Wyatt M. C., 2008, ARA&A, 46, 339
Wyatt M. C., 2018, Debris Disks: Probing Planet Formation. Springer International Publishing, Cham, pp 2543–2568, doi:10.1007/978-3-319-55333-7_146, https://doi.org/10.1007/978-3-319-55333-7_146
Wyatt M. C., 2020, Extrasolar Kuiper belts. Oxford: Elsevier, pp 351–376, doi:10.1016/B978-0-12-816490-7.00016-3
Wyatt M. C., Dent W. R. F., 2002, MNRAS, 334, 589
Wyatt M. C., Dermott S. F., Telesco C. M., Fisher R. S., Grogan K., Holmes E. K., Piña R. K., 1999, ApJ, 527, 918
Wyatt M. C., Dent W. R. F., Greaves J. S., 2003, MNRAS, 342, 876
Wyatt M. C., Smith R., Greaves J. S., Beichman C. A., Bryden G., Lisse C. M., 2007, ApJ, 658, 569
Wyatt M. C., Booth M., Payne M. J., Churcher L. J., 2010, MNRAS, 402, 657
Wyatt M. C., et al., 2012, MNRAS, 424, 1206
Wyatt M. C., Panić O., Kennedy G. M., Matrà L., 2015, Ap&SS, 357, 103
Wyatt M. C., Bonsor A., Jackson A. P., Marino S., Shannon A., 2017, MNRAS, 464, 3385
Xu Y.-B., Zhou L.-Y., Lhotka C., Ip W.-H., 2020, MNRAS, 493, 1447

References

Xuan J. W., Wyatt M. C., 2020, MNRAS, 497, 2096
Yang C.-C., Johansen A., Carrera D., 2017, A&A, 606, A80
Yelverton B., Kennedy G. M., 2018, MNRAS, 479, 2673
Yelverton B., Kennedy G. M., Su K. Y. L., 2020, MNRAS, 495, 1943
Youdin A. N., Goodman J., 2005, ApJ, 620, 459
Zheng X., Lin D. N. C., Kouwenhoven M. B. N., Mao S., Zhang X., 2017, ApJ, 849, 98
Zsom A., Dullemond C. P., 2008, A&A, 489, 931
Zuckerman B., Forveille T., Kastner J. H., 1995, Nature, 373, 494
de Muizon M. J., 2005, Space Sci. Rev., 119, 201
de Pater I., Lissauer J. J., 2010, Planetary Sciences. Cambridge University Press
van der Marel N., Dong R., di Francesco J., Williams J. P., Tobin J., 2019, ApJ, 872, 112

Appendices

Appendix A

Calculation of the Secular Ring–Ring Interaction

In this Appendix, I present a calculation of the secular disturbing function due to two coplanar rings interacting with each other via softened gravity in the form given by Equation (4.10). I do not assume any specific form for the softening function $\mathcal{F}$, apart from requiring it to be a function of the instantaneous positions of interacting particles with respect to the centre of the system (i.e., r_1 and r_2). I first write the ring–ring interaction function as[1]

$$\begin{aligned} \Psi &= \left[(\mathbf{r}_1 - \mathbf{r}_2)^2 + \mathcal{F}(r_1, r_2) \right]^{-1/2}, \\ &= \left[r_1^2 + r_2^2 - 2 r_1 r_2 \cos(f_1 - f_2 + \varpi_1 - \varpi_2) + \mathcal{F}(r_1, r_2) \right]^{-1/2}, \end{aligned} \tag{A.1}$$

where $\mathcal{F}(r_1, r_2)$ is an arbitrary softening function introduced to cushion the singularity which arises otherwise at null inter-particle separations. In the above expression, f_i is the true anomaly of the i^{th} ring, ϖ_i is its longitude of pericentre and r_i is its instantaneous position, $i = 1, 2$. My goal is to obtain the orbit-averaged expansion of Ψ to second order in eccentricities e_i valid for arbitrary $\mathcal{F}(r_1, r_2)$.

[1] Note that I do not deal with the indirect part of the potential – which is left unsoftened – as it contains only periodic terms and does not affect the secular dynamics (Murray & Dermott, 1999).

A.1 Expansion of the Interaction Function Ψ around Small Eccentricities

Following the classical techniques of celestial mechanics (see, Plummer, 1918, Ch. XVI), I start by expanding Ψ around circular orbits. Using Taylor expansion I write

$$\begin{aligned} \Psi &= \exp\left\{ \log\left(\frac{r_1}{a_1}\right) D_1 + \log\left(\frac{r_2}{a_2}\right) D_2 + (f_1 - M_1) D_3 + (f_2 - M_2) D_4 \right\} \Psi_0, \\ &\equiv \mathbb{T}\Psi_0 \end{aligned} \tag{A.2}$$

with

$$\Psi_0 = \left[a_1^2 + a_2^2 - 2a_1 a_2 \cos\theta + \mathcal{F}(a_1, a_2) \right]^{-1/2}, \tag{A.3}$$

where $\theta = M_1 - M_2 + \varpi_1 - \varpi_2$, M_i represents the mean anomaly of the i^{th} ring characterized with semimajor axis a_i, and the linear operators D_k are given by (Plummer, 1918)

$$D_1 = a_1 \frac{\partial}{\partial a_1} \equiv a_1 \partial_1 \quad , \qquad D_2 = a_2 \frac{\partial}{\partial a_2} \equiv a_2 \partial_2 \quad , \qquad \text{and} \qquad D_3 = -D_4 = \frac{\partial}{\partial \theta}. \tag{A.4}$$

Note that this expansion, as well as subsequent steps, is completely symmetric with respect to interchanging the particle indices.

Next, in order to calculate the action of the operator $\mathbb{T}$ defined by Equation (A.2) on the disturbing function of circular softened rings Ψ_0 (Equation A.3), I make use of the elliptical expansions of r/a and $f - M$,

$$\begin{aligned} (a^{-1} r)^D &= 1 - e\cos M \cdot D + \frac{1}{2} e^2 [1 - \cos(2M)] \cdot D \\ &\quad + \frac{1}{4} e^2 [1 + \cos(2M)] \cdot D(D-1) + \mathcal{O}(e^3), \end{aligned} \tag{A.5}$$

$$\begin{aligned} \exp\{(f - M) D\} &= 1 + 2e \sin M \cdot D + \frac{5}{4} e^2 \sin(2M) \cdot D \\ &\quad + e^2 [1 - \cos(2M)] \cdot D^2 + \mathcal{O}(e^3) \end{aligned} \tag{A.6}$$

to multiply individual terms appearing in $\mathbb{T}$, keep the ones up to second order in eccentricities, and drop all terms which do not contain the difference of mean anomalies, $k(M_1 - M_2)$, as they are evidently periodic and vanish upon orbit-averaging. Performing this procedure and dropping an irrelevant constant term, one can demonstrate that Ψ

reduces to

$$\Psi \;=\; \mathbb{T}\Psi_0 \equiv \mathbb{A}\Psi_0\; e_1^2 + \mathbb{B}\Psi_0\; e_2^2 + \mathbb{C}\Psi_0\; e_1 e_2 \cos(\varpi_1 - \varpi_2), \tag{A.7}$$

where the operators $\mathbb{A}$, $\mathbb{B}$ and $\mathbb{C}$ acting on Ψ_0 are defined as

$$\mathbb{A} \;\equiv\; D_3^2 + \frac{1}{4}D_1(D_1+1), \qquad \mathbb{B} \equiv D_4^2 + \frac{1}{4}D_2(D_2+1), \tag{A.8}$$

$$\mathbb{C} \;\equiv\; \cos\theta\left(2D_3D_4 + \frac{1}{2}D_1D_2\right) - \sin\theta(D_2D_3 - D_1D_4). \tag{A.9}$$

Note that in arriving at Equations (A.7) – (A.9), I have used the fact that $\cos(M_1 - M_2) = \cos\theta\cos(\varpi_1 - \varpi_2)$ and $\sin(M_1 - M_2) = \sin\theta\cos(\varpi_1 - \varpi_2)$ in the secular regime (see, e.g., Plummer, 1918).

A.2 Computation of the Action of Relevant Operators

Equipped with the expression (A.7) for Ψ, I proceed to compute the action of operator $\mathbb{T}$ on Ψ_0 prior to orbit-averaging the resultant expression. With this in mind, I compute the action of several operators appearing in the definitions of $\mathbb{A}$, $\mathbb{B}$ and $\mathbb{C}$ on Ψ_0 and list them below:

$$D_3^2\;\Psi_0 = D_4^2\;\Psi_0 \;\;=\;\; 3a_1^2a_2^2\sin^2\theta\;\Psi_0^5 - a_1a_2\cos\theta\;\Psi_0^3, \tag{A.10}$$

$$\begin{aligned} D_1D_2\;\Psi_0 = a_1a_2\left(\cos\theta - \frac{1}{2}\partial_1\partial_2\mathcal{F}\right)\;\Psi_0^3 \\ + 3\left(a_2^2 - a_1a_2\cos\theta + \frac{a_2}{2}\partial_2\mathcal{F}\right)\left(a_1^2 - a_1a_2\cos\theta + \frac{a_1}{2}\partial_1\mathcal{F}\right)\;\Psi_0^5, \end{aligned} \tag{A.11}$$

$$D_2D_3\;\Psi_0 = -a_1a_2\sin\theta\;\Psi_0^3 + 3a_1a_2\sin\theta\left(a_2^2 - a_1a_2\cos\theta + \frac{a_2}{2}\partial_2\mathcal{F}\right)\;\Psi_0^5, \tag{A.12}$$

$$D_1D_4\;\Psi_0 = a_1a_2\sin\theta\;\Psi_0^3 - 3a_1a_2\sin\theta\left(a_1^2 - a_1a_2\cos\theta + \frac{a_1}{2}\partial_1\mathcal{F}\right)\;\Psi_0^5, \tag{A.13}$$

$$D_1\;\Psi_0^3 = -3\left(a_1^2 - a_1a_2\cos\theta + \frac{a_1}{2}\partial_1\mathcal{F}\right)\;\Psi_0^5, \tag{A.14}$$

$$D_2\;\Psi_0^3 = -3\left(a_2^2 - a_1a_2\cos\theta + \frac{a_2}{2}\partial_2\mathcal{F}\right)\;\Psi_0^5, \tag{A.15}$$

where, for conciseness, I have written $\mathcal{F}$ instead of $\mathcal{F}(a_1, a_2)$. Here, it is worthwhile to mention that, as far as the expansion technique is concerned, the terms $\partial_i \mathcal{F}$ (with $i = 1,\ 2$) appearing in the above expressions are the only difference brought upon by softening the Newtonian point-mass interaction (Equation A.1).

Another set of operators useful in computing $\mathbb{T}\Psi_0$ is the following:

$$D_1(D_1+1)\ \Psi_0 \ = \ -D_1 D_2\ \Psi_0 + \frac{1}{2} D_1 \left[\left(2\mathcal{F} - a_1 \partial_1 \mathcal{F} - a_2 \partial_2 \mathcal{F} \right)\ \Psi_0^3 \right], \quad \text{(A.16)}$$

$$D_2(D_2+1)\ \Psi_0 \ = \ -D_1 D_2\ \Psi_0 + \frac{1}{2} D_2 \left[\left(2\mathcal{F} - a_1 \partial_1 \mathcal{F} - a_2 \partial_2 \mathcal{F} \right)\ \Psi_0^3 \right], \quad \text{(A.17)}$$

which can be obtained by making use of the identity $(D_1 + D_2 + 1)\Psi_0 = \frac{1}{2}(2\mathcal{F} - a_1\partial_1\mathcal{F} - a_2\partial_2\mathcal{F})\Psi_0^3$. Here, I note that for all softening functions $\mathcal{F}$ for which $2\mathcal{F} - a_1\partial_1\mathcal{F} - a_2\partial_2\mathcal{F} = 0$, one finds $D_1 + D_2 = -1$. Consequently, in such cases, the operators $D_1(D_1+1)$ and $D_2(D_2+1)$ become identical rendering $\mathbb{A}\Psi_0 = \mathbb{B}\Psi_0$ (since $D_3^2 = D_4^2$, see Equations (A.8) and (A.10)). As a result, the resultant orbit-averaged disturbing function (Equation (A.7)) is symmetric in e_1 and e_2, similar to the case of classical Laplace–Lagrange theory (Murray & Dermott, 1999). This is not true in general, for instance, when $\mathcal{F}(r_1, r_2) = \text{const} \neq 0$.

A.3 Orbit-averaging the Interaction Function Ψ

Expressions (A.10)-(A.17) allow the computation of $\Psi = \mathbb{T}\Psi_0$, which needs to be time-averaged in order to recover the secular disturbing function. For the sake of conciseness, here I do not show the cumbersome collated expression for $\mathbb{T}\Psi_0$ and proceed to the final step of orbit-averaging, which will conclude my derivation. In short, my goal is to compute

$$\langle \Psi \rangle = \langle \mathbb{T}\Psi_0 \rangle = \frac{1}{2\pi} \int_0^{2\pi} \mathbb{T}\Psi_0 \ d\theta, \quad \text{(A.18)}$$

which essentially reduces to computing the individual terms $\langle \mathbb{A}\Psi_0 \rangle$, $\langle \mathbb{B}\Psi_0 \rangle$ and $\langle \mathbb{C}\Psi_0 \rangle$.

At the outset, it is important to note that each of the terms appearing in $\mathbb{T}\Psi_0$ (through $\mathbb{A}\Psi_0$, $\mathbb{B}\Psi_0$ and $\mathbb{C}\Psi_0$, or the operators they entail) are proportional to $\cos(m\theta)\Psi_0^{2s}$. By making use of $\alpha = a_</a_>$, where $a_< = \min(a_1, a_2)$ and $a_> =$

$\max(a_1, a_2)$, this combination can be reduced to

$$\cos(m\theta)\Psi_0^{2s} = a_>^{-2s}\cos(m\theta)\Big[1+\alpha^2-2\alpha\cos\theta+a_>^{-2}\mathcal{F}(a_1,a_2)\Big]^{-s}. \tag{A.19}$$

For that reason, calculation of the orbit-averaged Ψ (by integrating over $d\theta$) yields integrals of the form

$$\mathcal{B}_s^{(m)}(\alpha) \equiv \frac{2}{\pi}\int_0^\pi \cos(m\theta)\Big[1+\alpha^2-2\alpha\cos\theta+\epsilon^2(\alpha)\Big]^{-s}d\theta, \tag{A.20}$$

which is the generalization of the classical Laplace coefficients $b_s^{(m)}$ (recovered when $\mathcal{F}(a_1,a_2)=0$, see Equation 4.2) with the dimensionless softening parameter

$$\epsilon^2(\alpha) \equiv a_>^{-2}\mathcal{F}(a_1,a_2), \tag{A.21}$$

see Equation (4.7).

Employing this notation, I now present the simplified expressions of $\langle\mathbb{A}\Psi_0\rangle$, $\langle\mathbb{B}\Psi_0\rangle$, and $\langle\mathbb{C}\Psi_0\rangle$ obtained as a result of orbit-averaging:

$$\begin{aligned}
a_> \langle\mathbb{A}\Psi_0\rangle(\alpha) &\equiv \phi_{11}(\alpha)\\
&= \frac{\alpha}{2}\Bigg\{-\frac{5}{4}\mathcal{B}_{3/2}^{(1)}+\frac{3}{8}\alpha\mathcal{B}_{5/2}^{(0)}+\frac{3}{4}(1+\alpha^2)\mathcal{B}_{5/2}^{(1)}-\frac{15}{8}\alpha\mathcal{B}_{5/2}^{(2)}+\frac{3}{8}T_2\mathcal{B}_{5/2}^{(1)}-\frac{3}{16}T_5\mathcal{B}_{5/2}^{(0)}\\
&\quad+\frac{1}{8}\Big(T_3+\alpha^{-1}T_4\Big)\mathcal{B}_{3/2}^{(0)}-\frac{3}{8}T_1\Big(\frac{a_1}{a_2}\mathcal{B}_{5/2}^{(0)}-\mathcal{B}_{5/2}^{(1)}+\frac{1}{2}T_7\mathcal{B}_{5/2}^{(0)}\Big)\Bigg\}, \qquad \text{(A.22)}\\
a_> \langle\mathbb{B}\Psi_0\rangle(\alpha) &\equiv \phi_{22}(\alpha)\\
&= \frac{\alpha}{2}\Bigg\{-\frac{5}{4}\mathcal{B}_{3/2}^{(1)}+\frac{3}{8}\alpha\mathcal{B}_{5/2}^{(0)}+\frac{3}{4}(1+\alpha^2)\mathcal{B}_{5/2}^{(1)}-\frac{15}{8}\alpha\mathcal{B}_{5/2}^{(2)}+\frac{3}{8}T_2\mathcal{B}_{5/2}^{(1)}-\frac{3}{16}T_5\mathcal{B}_{5/2}^{(0)}\\
&\quad+\frac{1}{8}\Big(T_3+\alpha^{-1}T_6\Big)\mathcal{B}_{3/2}^{(0)}-\frac{3}{8}T_1\Big(\frac{a_2}{a_1}\mathcal{B}_{5/2}^{(0)}-\mathcal{B}_{5/2}^{(1)}+\frac{1}{2}T_8\mathcal{B}_{5/2}^{(0)}\Big)\Bigg\}, \qquad \text{(A.23)}\\
a_> \langle\mathbb{C}\Psi_0\rangle(\alpha) &\equiv \phi_{12}(\alpha)\\
&= \frac{\alpha}{2}\Bigg\{\frac{9}{4}\mathcal{B}_{3/2}^{(0)}+\frac{1}{4}\mathcal{B}_{3/2}^{(2)}+\frac{3}{8}\alpha\mathcal{B}_{5/2}^{(3)}+\frac{21}{8}\alpha\mathcal{B}_{5/2}^{(1)}+\frac{3}{4}(1+\alpha^2)\mathcal{B}_{5/2}^{(2)}-\frac{9}{4}(1+\alpha^2)\mathcal{B}_{5/2}^{(0)}\\
&\quad-\frac{1}{4}T_3\mathcal{B}_{3/2}^{(1)}-\frac{9}{8}T_2\mathcal{B}_{5/2}^{(0)}+\frac{3}{8}T_5\mathcal{B}_{5/2}^{(1)}+\frac{3}{8}T_2\mathcal{B}_{5/2}^{(2)}\Bigg\}. \qquad \text{(A.24)}
\end{aligned}$$

In Equations (A.22)–(A.24), I have defined the dimensionless functions $T_i(\alpha)$ such that

$$T_1 = a_>^{-2}\Big(2\mathcal{F} - a_1\partial_1\mathcal{F} - a_2\partial_2\mathcal{F}\Big), \qquad T_2 = \alpha\left(\frac{\partial_1\mathcal{F}}{a_2} + \frac{\partial_2\mathcal{F}}{a_1}\right), \qquad T_3 = \partial_1\partial_2\mathcal{F}, \quad \text{(A.25)}$$

$$T_4 = \frac{a_1}{a_>^2}\partial_1[2\mathcal{F} - a_1\partial_1\mathcal{F} - a_2\partial_2\mathcal{F}], \qquad T_5 = \alpha\left(2\frac{\partial_1\mathcal{F}}{a_1} + 2\frac{\partial_2\mathcal{F}}{a_2} + \frac{\partial_1\mathcal{F}}{a_1}\frac{\partial_2\mathcal{F}}{a_2}\right), \quad \text{(A.26)}$$

$$T_6 = \frac{a_2}{a_>^2}\partial_2[2\mathcal{F} - a_1\partial_1\mathcal{F} - a_2\partial_2\mathcal{F}], \qquad T_7 = a_2^{-1}\partial_1\mathcal{F}, \qquad T_8 = a_1^{-1}\partial_2\mathcal{F}, \quad \text{(A.27)}$$

where, as before, $\mathcal{F} \equiv \mathcal{F}(a_1, a_2)$, $\alpha = a_</a_>$ and $\partial_i \equiv \partial/\partial a_i$. Note that the expressions for ϕ_{11} and ϕ_{22} swap definitions upon replacing a_1 by a_2, whilst keeping $\alpha < 1$ by construction. This can be understood by first noting that functions T_i with $i = 1,\ 2,\ 3,$ and 5 are invariant under $a_1 \leftrightharpoons a_2$ while, at the same time, T_4 and T_7 (appearing in the second line of Equation (A.22)) translate to T_6 and T_8 (appearing in the second line of Equation (A.23)); and vice versa.

These identities, when combined, yield the desired expression of $\langle\Psi\rangle = \langle\mathbb{T}\Psi_0\rangle$; see Equations (A.7)–(A.9). Subsequently, the softened ring-ring disturbing function in the form given by Equation (4.11) is recovered, with the coefficients ϕ_{ij} defined by Equations (A.22) – (A.24). This completes my calculation of the secular ring–ring interaction between two softened coplanar rings, up to second order in eccentricity and valid for arbitrary softening functions $\mathcal{F}(r_1, r_2)$.

Note that in the absence of softening (i.e., $\mathcal{F}(r_1, r_2) = 0$), we have $T_i = 0$ for all i (i.e., $i = 1..8$), and the classical expressions for $\phi_{11}^{\rm LL}$, $\phi_{22}^{\rm LL}$ and $\phi_{12}^{\rm LL}$ — i.e., Equations (4.8)–(4.9) — are recovered. Finally, it is worthwhile to note that the expansion technique exploited here can be used to recover the orbit-averaged disturbing function valid to *arbitrary* order in eccentricities, as well as inclinations.

A.4 Alternative Calculation: Secular Disc–Particle Interaction

Calculations presented above describe the orbit-averaged coupling between two individual disc annuli, which subsequently need to be integrated over the semimajor axes of the disc elements to represent the effect of a continuous disc. In principle, one can also arrive at the expressions (4.4) by assuming a continuous mass distribution in the disc from the start and performing a calculation similar to that in Davydenkova & Rafikov (2018). Namely, one would need to compute $R_d = \langle G\int_S \Sigma(\mathbf{r}_d)\Phi(\mathbf{r}_d, \mathbf{r}_p)dS\rangle$, where Φ is the interaction potential given by Equation (4.10), angle brackets indicate

averaging over the orbit of the test particle given by $\mathbf{r}_p$, and integration is carried out over the full surface of the disc S with $\mathbf{r}_d$ denoting the location of a disc element. To obtain the expression for R_d accurate to second order in eccentricities, one would need to expand $\Phi(\mathbf{r}_d, \mathbf{r}_p)$ to second order in particle and disc eccentricities by, e.g., writing $r_p = a_p(1 - e_p \cos E_p)$, where E_p is the eccentric anomaly of the particle orbit. This expansion should explicitly account for the dependence of $\mathcal{F}$ on $\mathbf{r}_d$ and $\mathbf{r}_p$. Averaging the resulting expressions over E_p, one would arrive at the proper expression for R_d in the form given by Equation (4.4).

In particular, after a lengthy but straightforward calculation, this method gives the following expression for the disc-driven precession rate:

$$\begin{aligned}\mathsf{A}_d = \frac{\pi G}{2 n_p a_p^2} \int \frac{a \Sigma(a) da}{a_>} \times \\ \left\{ \frac{1}{4} \left[3 a_p \mathcal{F}' \left(\mathcal{F}' + 4 a_p \right) - 2 \left(2\mathcal{F}' + a_p \mathcal{F}'' \right) \left(a_p^2 + a^2 + \mathcal{F} \right) - 12 a_p \mathcal{F} \right] \frac{a_p \mathcal{B}^{(0)}_{5/2}(\alpha)}{a_>^4} \right. \\ \left. + \alpha \mathcal{B}^{(1)}_{3/2}(\alpha) - \left(\mathcal{F}' - a_p \mathcal{F}'' \right) \frac{a_p \alpha \mathcal{B}^{(1)}_{5/2}(\alpha)}{a_>^2} \right\},\end{aligned} \tag{A.28}$$

where prime denotes differentiation with respect to a_p (e.g., $\mathcal{F}' = \partial \mathcal{F} / \partial a_p$), $a_> = \max(a_p, a)$, $\alpha = \min(a_p, a) / \max(a_p, a)$, and integration is done over the semimajor axis a of the disc elements. Calculation of the non-axisymmetric part of R_d resulting from non-zero disc eccentricity (i.e., $\mathbf{B}_d$) is somewhat more tedious but can nevertheless be done similar to Davydenkova & Rafikov (2018).

Appendix B

Specific Cases of $\mathcal{F}(r_1, r_2)$

As already mentioned in Appendix A, the functions $\phi_{ij}(\alpha)$ with $i, j = 1, 2$ (Equations (A.22)–(A.24)) entering in the ring–ring secular disturbing function (Equation 4.11) reduce to the classical Laplace–Lagrange expressions for ϕ_{11}^{LL}, ϕ_{22}^{LL}, and ϕ_{12}^{LL} (Equations (4.8)–(4.9)) in the absence of softening, i.e., $\mathcal{F}(r_1, r_2) = 0$.

In addition to this, however, the general framework developed in Appendix A allows one to recover the expressions of ϕ_{ij} arrived at by Touma (2002) and Hahn (2003) upon specifying certain functional forms of $\mathcal{F}(r_1, r_2)$. Indeed, Touma (2002) performed the same calculations as those presented in Appendix A for the case of Plummer potential – $\mathcal{F}(r_1, r_2) = b_c^2$ – to second order in eccentricities, and later to fourth order in eccentricities (Touma & Sridhar, 2012). Furthermore, I find that the results obtained by Hahn (2003) can be recovered from my general framework by setting $\mathcal{F}(r_1, r_2) = H^2(r_1^2 + r_2^2)$. For reference, the functional forms of T_i for these forms of $\mathcal{F}(r_1, r_2)$, along with their softening parameters $\epsilon^2(\alpha)$, are summarized in Table B.1, which can be used to show that Equations (A.22)–(A.24) reduce to those in Table 4.1 after some algebra with the aid of the recursive relationships for $\mathcal{B}_s^{(m)}$ presented in Appendix C.

As to the formalism of Teyssandier & Ogilvie (2016), by using their softening prescription of $\mathcal{F}(r_1, r_2) = S^2 r_1 r_2$, I find that my general framework yields ϕ_{ij} expressions different from those reported by Teyssandier & Ogilvie (2016). To demonstrate this, I first note that in this case one has $T_1 = T_4 = T_6 = 0$ (Table B.1), which renders the

expressions of ϕ_{11} and ϕ_{22} identical such that

$$\phi_{11} = \phi_{22} = \frac{\alpha}{8}\Bigg\{ -5\mathcal{B}^{(1),\mathrm{TO}}_{3/2} + \frac{3}{2}\alpha\mathcal{B}^{(0),\mathrm{TO}}_{5/2} + 3(1+\alpha^2+S^2\alpha)\mathcal{B}^{(1),\mathrm{TO}}_{5/2} - \frac{15}{2}\alpha\mathcal{B}^{(2),\mathrm{TO}}_{5/2} - \frac{3}{4}S^2(S^2\alpha+2\alpha^2+2)\mathcal{B}^{(0),\mathrm{TO}}_{5/2} + \frac{1}{2}S^2\mathcal{B}^{(0),\mathrm{TO}}_{3/2}\Bigg\}. \quad \text{(B.1)}$$

The above expression can be simplified further using the recursive relationships listed in Appendix C.1. Indeed, Equation (C.2) with $m = 1$ and $s = 5/2$ and Equation (C.1) with $m = 1$ and $s = 3/2$ read as

$$3(1+\alpha^2+S^2\alpha)\mathcal{B}^{(1),\mathrm{TO}}_{5/2} = -\frac{3\alpha}{2}\mathcal{B}^{(2),\mathrm{TO}}_{5/2} + \frac{15}{2}\alpha\mathcal{B}^{(0),\mathrm{TO}}_{5/2}, \quad \text{(B.2)}$$

$$-6\mathcal{B}^{(1),\mathrm{TO}}_{3/2} = 9\alpha\left(\mathcal{B}^{(2),\mathrm{TO}}_{5/2} - \mathcal{B}^{(0),\mathrm{TO}}_{5/2}\right), \quad \text{(B.3)}$$

respectively. Inserting the above two identities in Equation (B.1) one then arrives at Equation (4.32). Similarly, the expression of ϕ_{12} (Equation A.24) can be simplified with the aid of Equation (C.3) (with $m = 0,\ s = 3/2$), Equation (C.2) (with $m = 2,\ s = 5/2$), and Equation (C.1) (with $m = 2,\ s = 3/2$) resulting in Equation (4.33) after some straightforward algebra. As discussed in Section 4.6.2, the terms in Equations (4.32) and (4.33) which are explicitly proportional to S^2 are absent in the original formulation of Teyssandier & Ogilvie (2016); see Table 4.1.

Similarly, for the formalism of Tremaine (1998), propagating their functional form of $\mathcal{F}(r_1, r_2) = \beta_c^2 \max(r_1^2, r_2^2)$ through my general framework, I arrive at expressions for $\phi_{ij}(\alpha)$ differing from those reported in Tremaine (1998) in a very special way: I find ϕ_{ij} to contain additional terms proportional to $T_3(\alpha) \sim \delta(\alpha - 1)$, where $\delta(x)$ is the Dirac delta-function. Such terms are absent in the original formulation of Tremaine (1998); see Tables 4.1 and B.1. Emergence of these terms can be easily demonstrated by first noting that in this case $\phi_{11} = \phi_{22}$ (as $T_1 = T_4 = T_6 = 0$), employing the recursive relationships for Laplace coefficients (in a similar order as done above for Teyssandier & Ogilvie (2016)) to simplify the general expressions of $\phi_{11}(= \phi_{22})$ and ϕ_{12}, and finally arriving at Equations (4.26) and (4.27). The ramifications of this finding is discussed in Section 4.6.1.

Table B.1. The functional forms of the coefficients $T_i(\alpha)$ given by Equations (A.25)–(A.27) appearing in the orbit-averaged disturbing function due to two coplanar (arbitrarily) softened rings (Equations A.22–A.24) such that $\alpha \equiv a_</a_> \leq 1$.

Method	$\mathcal{F}(r_1, r_2)$	$\epsilon^2(\alpha)$	$T_1(\alpha)$	$T_2(\alpha)$	$T_3(\alpha)$	$T_4(\alpha)$	$T_5(\alpha)$	$T_6(\alpha)$	$T_7(\alpha)$	$T_8(\alpha)$
Hahn (2003)	$H^2(r_1^2 + r_2^2)$	$H^2(1+\alpha^2)$	0	$2H^2(1+\alpha^2)$	0	0	$4\alpha H^2(2+H^2)$	0	$2H^2\frac{a_1}{a_2}$	$2H^2\frac{a_2}{a_1}$
Touma (2002)	b_c^2	$\beta^2 = (b_c/a_>)^2$	$2\beta^2$	0	0	0	0	0	0	0
Tremaine (1998)	$\beta_c^2 \max(\mathrm{r}_1^2, \mathrm{r}_2^2)$	β_c^2	0	$2\beta_c^2$	$-2\beta_c^2\delta(\alpha-1)$	0	$4\alpha\beta_c^2$	0	$\frac{2\beta_c^2}{\alpha}\Theta(a_1-a_2)$	$\frac{2\beta_c^2}{\alpha}\Theta(a_2-a_1)$
Teyssandier & Ogilvie (2016)	$S^2 r_1 r_2$	$S^2\alpha$	0	$2\alpha S^2$	S^2	0	$S^2(S^2\alpha + 2\alpha^2 + 2)$	0	S^2	S^2

Note. The first column lists the softening prescriptions analysed in this work (see Section 4.2.1), while the second column shows the specific forms of the softening function $\mathcal{F}(r_1, r_2)$ in Equation (A.1). For reference, the corresponding expressions for the dimensionless softening parameters $\epsilon^2(\alpha) = a_>^{-2}\mathcal{F}(a_1, a_2)$ (Equation A.21) entering in the definition of softened Laplace coefficients (Equation A.20) are also shown. Note that $\Theta(x)$ represents the Heaviside step function and $\delta(x) = d\Theta(x)/dx$ stands for Dirac delta-function.

Appendix C

Generalized Laplace Coefficients

In Appendix A, I demonstrated that softening the Newtonian point-mass potential by an arbitrary function $\mathcal{F}(r_1, r_2)$ modifies the definition of the classical Laplace coefficients – as shown by Equations (4.7) and (A.20) – by the introduction of a softening parameter $\epsilon^2(\alpha) = a_>^{-2}\mathcal{F}(a_1, a_2)$ (Equation A.21), $0 \leq \alpha = a_</a_> \leq 1$. In this Appendix, I present (i) some useful recursive relationships amongst different generalized Laplace coefficients $\mathcal{B}_s^{(m)}(\alpha)$, (ii) along with their asymptotic behaviour in the limits of $\alpha \to 0$ and $\alpha \to 1$, (iii) as well as their relationship to complete elliptic integrals.

C.1 Recursive Relations

Generalizing the results for the usual (unsoftened) Laplace coefficients $b_s^{(m)}$ (e.g., Plummer, 1918, p. 159), the following relationships can be easily obtained for the generalized Laplace coefficients defined by Equations (4.7) and (A.20):

$$m\mathcal{B}_s^{(m)} = s\alpha\mathcal{B}_{s+1}^{(m-1)} - s\alpha\mathcal{B}_{s+1}^{(m+1)}, \tag{C.1}$$

$$m(1+\alpha^2+\epsilon^2)\mathcal{B}_s^{(m)} = \alpha(m+1-s)\mathcal{B}_s^{(m+1)} + \alpha(m+s-1)\mathcal{B}_s^{(m-1)}, \tag{C.2}$$

$$(m+s)\mathcal{B}_s^{(m)} = s(1+\alpha^2+\epsilon^2)\mathcal{B}_{s+1}^{(m)} - 2s\alpha\mathcal{B}_{s+1}^{(m+1)}. \tag{C.3}$$

The difference with the classical recursive relations for $b_s^{(m)}$ amounts to substituting the combination $1+\alpha^2$ appearing in the case of ordinary Laplace coefficients with $1+\alpha^2+\epsilon^2(\alpha)$.

Another useful expression relating the generalized Laplace coefficients of arguments α and α^{-1} is

$$\mathcal{B}_s^{(m)}(\alpha^{-1}) = \alpha^{2s}\mathcal{B}_s^{(m)}(\alpha). \tag{C.4}$$

Note that the above relationship is valid only as long as the softening parameter satisfies $\alpha^2\epsilon^2(1/\alpha) = \epsilon^2(\alpha)$. For instance, this condition is violated when the softening parameter ϵ has no dependence on α (e.g., that of Tremaine (1998), see Table 4.1).

C.2 Asymptotic Behaviour

Here I derive approximate expressions for $\mathcal{B}_s^{(m)}$ in the asymptotic limits; for $\alpha \to 0$ and $\alpha \to 1$.

Case 1: In the limit of $\alpha \approx 0$, one can factor out the term $1 + \alpha^2 + \epsilon^2(\alpha)$ from the integrand of $\mathcal{B}_s^{(m)}$ to expand the denominator around $\gamma^{-1} \approx 0$, where $\gamma = (2\alpha)^{-1}[1 + \alpha^2 + \epsilon^2(\alpha)]$. This allows one to approximate $\mathcal{B}_s^{(m)}$ as

$$B_s^{(m)}(\alpha) \approx \frac{2}{\pi(2\alpha\gamma)^s} \int_0^\pi \cos(m\theta) \times \left[1 + \frac{s}{\gamma}\cos\theta + \frac{s(s+1)}{2\gamma^2}\cos^2\theta + \frac{s(s+1)(s+2)}{6\gamma^3}\cos^3\theta\right] d\theta. \tag{C.5}$$

Using the orthogonality of the cosine functions, it is straightforward to show that

$$\mathcal{B}_s^{(m)} \approx \frac{\alpha^m F_m}{(2\alpha\gamma)^{s+m}} \quad \text{as} \quad \alpha \to 0, \quad \text{where} \quad F_m = \begin{cases} 2 & \text{if } m = 0 \\ 2s & \text{if } m = 1 \\ s(s+1) & \text{if } m = 2 \\ \frac{1}{3}s(s+1)(s+2) & \text{if } m = 3 \end{cases} \tag{C.6}$$

Case 2: In the opposite limit of $x = 1 - \alpha \approx 0$, the dominant contribution to $\mathcal{B}_s^{(m)}$ comes from $\theta \ll 1$ (Goldreich & Tremaine, 1980). Thus one can set $\cos(m\theta) \to 1$ in the numerator, approximate $\cos\theta \approx 1 - \theta^2/2$ in the denominator and extend the integration limit to infinity. Furthermore, setting $\alpha = 1$ (i.e. $x = 0$) everywhere except when it appears in the combination $1 - \alpha$, the generalized Laplace coefficient can be approximated as

$$\mathcal{B}_s^{(m)} \approx \frac{2}{\pi} \int_0^\infty \frac{d\theta}{\left[x^2 + \theta^2 + \epsilon_{\alpha=1}^2\right]^s} = \frac{2}{\pi} \begin{cases} (x^2 + \epsilon_{\alpha=1}^2)^{-1} & \text{if } s = 3/2 \\ (2/3)(x^2 + \epsilon_{\alpha=1}^2)^{-2} & \text{if } s = 5/2 \end{cases} \tag{C.7}$$

where $\epsilon_{\alpha=1}^2$ is the softening parameter evaluated at $\alpha = 1$.

C.3 Relationship to Elliptic Integrals

Here I express the generalized Laplace coefficients $\mathcal{B}_s^{(m)}$ in terms of complete elliptic integrals. These expressions can be used for rapid numerical evaluation of the generalized Laplace coefficients without relying on numerical integration of Equation (A.20) (or Equation (4.7)). Let us write, as before, $2\alpha\gamma = 1 + \alpha^2 + \epsilon^2(\alpha)$ and define $\chi = \sqrt{2/(\gamma+1)}$ such that, for any general softening parameter $\epsilon^2(\alpha)$, we have $0 \leq \chi \leq 1$ and $\gamma \geq 1$. Now let us express $\mathcal{B}_s^{(m)}$ in terms of γ to write

$$\mathcal{B}_s^{(m)} = \frac{2^{1-s}}{\pi\alpha^s} \int_0^\pi \frac{\cos(m\theta)}{(\gamma - \cos\theta)^s} d\theta. \tag{C.8}$$

Introducing complete elliptic integrals $\mathbf{K}(\chi) = \int_0^{\pi/2} (1 - \chi^2 \sin^2\phi)^{-1/2}\, d\phi$ and $\mathbf{E}(\chi) = \int_0^{\pi/2} (1 - \chi^2 \sin^2\phi)^{1/2}\, d\phi$, it is then easy to demonstrate that, for relevant values of s and m (see Appendix A), $\mathcal{B}_s^{(m)}$ can be written as follows:

$$\mathcal{B}_{3/2}^{(0)} = \frac{2\mathbf{E}(\chi)}{\pi\alpha(\gamma-1)\sqrt{2\alpha(\gamma+1)}}, \tag{C.9}$$

$$\mathcal{B}_{3/2}^{(1)} = \frac{2\Big[-(\gamma-1)\mathbf{K}(\chi) + \gamma\mathbf{E}(\chi)\Big]}{\pi\alpha(\gamma-1)\sqrt{2\alpha(\gamma+1)}}, \tag{C.10}$$

$$\mathcal{B}_{3/2}^{(2)} = \frac{2\Big[-4\gamma(\gamma-1)\mathbf{K}(\chi) + (4\gamma^2-3)\mathbf{E}(\chi)\Big]}{\pi\alpha(\gamma-1)\sqrt{2\alpha(\gamma+1)}}, \tag{C.11}$$

$$\mathcal{B}_{3/2}^{(3)} = \frac{2}{3}\frac{\Big[-(\gamma-1)(32\gamma^2-5)\mathbf{K}(\chi) + \gamma(32\gamma^2-29)\mathbf{E}(\chi)\Big]}{\pi\alpha(\gamma-1)\sqrt{2\alpha(\gamma+1)}}, \tag{C.12}$$

$$\mathcal{B}_{5/2}^{(0)} = \frac{4\Big[-(\gamma-1)\mathbf{K}(\chi) + 4\gamma\mathbf{E}(\chi)\Big]}{3\pi(2\alpha)^{5/2}(\gamma+1)^{3/2}(\gamma-1)^2}, \tag{C.13}$$

$$\mathcal{B}_{5/2}^{(1)} = \frac{4\Big[-\gamma(\gamma-1)\mathbf{K}(\chi) + (\gamma^2+3)\mathbf{E}(\chi)\Big]}{3\pi(2\alpha)^{5/2}(\gamma+1)^{3/2}(\gamma-1)^2}, \tag{C.14}$$

$$\mathcal{B}_{5/2}^{(2)} = \frac{4\Big[(\gamma-1)(4\gamma^2-5)\mathbf{K}(\chi) - 4\gamma(\gamma^2-2)\mathbf{E}(\chi)\Big]}{3\pi(2\alpha)^{5/2}(\gamma+1)^{3/2}(\gamma-1)^2}, \tag{C.15}$$

$$\mathcal{B}^{(3)}_{5/2} = \frac{4\Big[\gamma(\gamma-1)(32\gamma^2-33)\mathbf{K}(\chi) - (32\gamma^4-57\gamma^2+21)\mathbf{E}(\chi)\Big]}{3\pi(2\alpha)^{5/2}(\gamma+1)^{3/2}(\gamma-1)^2}. \tag{C.16}$$

The expressions given above permit for an efficient numerical evaluation of arbitrarily softened Laplace coefficients as functions of α, since effective algorithms for computing $\mathbf{K}$ and $\mathbf{E}$ already exist in the literature (e.g., Press et al., 2002).

Appendix D

Convergence Criterion for the Pre-factors of Power-Law Discs

Astrophysical discs often extend over a few orders of magnitude in radius so that $a_{\rm out}/a_{\rm in} \gg 1$. In such situations, far from the disc edges, one can take the limit of both $\alpha_1 = a_{\rm in}/a_p$ and $\alpha_2 = a_p/a_{\rm out}$ going to zero, provided that the gravitational potential of a power-law disc is insensitive to the locations of the disc boundaries (see Equations 4.18 and 4.19). Then the pre-factors ψ_1 and ψ_2 of the disturbing function converge to values depending only on the power-law indices p and $p+q$ respectively, as well as on the adopted softening prescription.

The conditions on the values of p and q which guarantee this convergence can be determined by expanding the coefficients $\phi_{ij}(\alpha)$, which appear in the integrands of each of ψ_1 and ψ_2, in the limit of $\alpha \approx 0$. Using the Taylor expansions of softened Laplace coefficients $\mathcal{B}_s^{(m)}$, we determined that both ψ_1 and ψ_2 calculated using the softening methods of Hahn (2003) and Tremaine (1998) (as well as its rectified version) are convergent as long as $-1 < p < 4$ and $-2 < p+q < 5$, respectively, for all values of softening (i.e, H and β_c, respectively). This follows from the fact that for both Hahn (2003) and Tremaine (1998) we have $\phi_{11} = \phi_{22} \sim \alpha^2$ and $\phi_{12} \sim \alpha^3$ to lowest order in α. These ranges of p and $p+q$ are in line with the findings of Silsbee & Rafikov (2015a).

As to the (rectified) softening model of Teyssandier & Ogilvie (2016), a similar exercise yields that $\phi_{11} = \phi_{22} \approx -\frac{1}{4}S^2\alpha + \frac{3}{8}(1+\frac{3}{2}S^4)\alpha^2$ and $\phi_{12} \approx \frac{3}{2}S^2\alpha^2 - \frac{15}{16}(1+5S^4)\alpha^3$ which, in the limit of $S \to 0$, translate to the same ranges for ψ_1 and ψ_2 convergence as Silsbee & Rafikov (2015a). However, when S is relatively large, it is trivial to show that ψ_1 and ψ_2 are convergent over limited ranges of $0 < p < 3$ and $-1 < p+q < 4$, respectively. A similar analysis for the softening method of Touma (2002) reveals that the ranges for ψ_1 and ψ_2 convergence are in line with the findings of Silsbee &

Rafikov (2015a) when the corresponding softening parameter $b_c \to 0$. However, when b_c is non-zero, the ranges are narrowed down to $-1 < p < 2$ and $-2 < p + q < 3$ respectively.

Appendix E

Disturbing Function of Planet due to Disc Gravity

To calculate the secular disturbing function $R_{d,p}$ of a planet due to an external disc, I use Equations (4.4), (4.5), and (4.6) from Chapter 4 for the case of unsoftened gravity. Strictly speaking, these equations represent the continuum version of the classical Laplace-Lagrange theory (e.g., Murray & Dermott, 1999), and are valid for arbitrary profiles of disc surface density $\Sigma_d(a)$, eccentricity $e_d(a)$, and apsidal angle $\varpi_d(a)$; see Chapter 4.

For the purposes of Chapter 5, I consider the disc to be apse-aligned (i.e., $d\varpi_d/da = 0$) and have surface density $\Sigma_d(a)$ given by Equation (5.1). For completeness, and use in Chapter 6, I also adopt a power-law scaling for the disc eccentricity given by

$$e_d(a) = e_0 \left(\frac{a_{\rm out}}{a} \right)^q \tag{E.1}$$

for $a_{\rm in} \leq a \leq a_{\rm out}$. Plugging these ansatzes into Equations (4.4)–(4.6) of Chapter 4, it can be shown, after some algebra, that $R_{d,p}$ is given by:

$$R_{d,p} = n_p a_p^2 \left[\frac{1}{2} A_{d,p} e_p^2 + B_{d,p} e_p \cos\left(\varpi_p - \varpi_d\right) \right], \tag{E.2}$$

with

$$A_{d,p}(a_p) = 2\pi \frac{G\Sigma_d(a_{\rm in})}{n_p a_p} \frac{a_{\rm in}}{a_p} \phi_1, \tag{E.3}$$

$$B_{d,p}(a_p) = \pi \frac{G\Sigma_d(a_{\rm in})}{n_p a_p} \frac{a_{\rm in}}{a_p} e_d(a_{\rm in}) \phi_2. \tag{E.4}$$

Here $A_{d,p}$ represents the free precession rate of the planetary orbit in the disc potential, while $B_{d,p}$ represents the torque exerted on the planet by the non-axisymmetric component of the disc gravity (which I have neglected in Chapter 5; see Section 5.8.2). The effects of the latter is explored in Chapter 6. Note that the coefficients $A_{d,p}$ and $B_{d,p}$ can be easily expressed in terms of the total disc mass M_d using Equation (5.2).

The coefficients ϕ_1 and ϕ_2 appearing in Equations (E.3) and (E.4), respectively, are given by:

$$\begin{aligned}
\phi_1 &= \frac{1}{4}\left(\frac{a_p}{a_{\rm in}}\right)^{1-p} \int_{a_p/a_{\rm out}}^{a_p/a_{\rm in}} \alpha^{p-1} b_{3/2}^{(1)}(\alpha) d\alpha, \\
&= \frac{3}{4}\left(\frac{a_p}{a_{\rm in}}\right)^{2} \frac{1-\delta^{-1-p}}{p+1} \phi_1^c, \qquad \text{(E.5)} \\
\phi_2 &= -\frac{1}{2}\left(\frac{a_p}{a_{\rm in}}\right)^{1-p-q} \int_{a_p/a_{\rm out}}^{a_p/a_{\rm in}} \alpha^{p+q-1} b_{3/2}^{(2)}(\alpha) d\alpha, \\
&= -\frac{15}{8}\left(\frac{a_p}{a_{\rm in}}\right)^{3} \frac{1-\delta^{-2-p-q}}{p+q+2} \phi_2^c, \qquad \text{(E.6)}
\end{aligned}$$

where $\delta \equiv a_{\rm out}/a_{\rm in}$ as in Chapter 5.

Here, the second lines in both Equations (E.5) and (E.6) are obtained by performing the integrals appearing in the definitions of ϕ_1 and ϕ_2 assuming $\alpha \to 0$; that is, $b_{3/2}^{(1)}(\alpha) \approx 3\alpha$ and $b_{3/2}^{(2)}(\alpha) \approx (15/4)\alpha^2$ which are valid to the lowest order in α. Thus, the coefficients ϕ_1^c and ϕ_2^c in Equations (E.5) and (E.6) represent correction factors accounting for the contribution of disc annuli close to the planet, i.e., higher-order terms in $b_{3/2}^{(m)}(\alpha)$. It is straightforward to show that

$$\begin{aligned}
\phi_1^c &= \frac{1}{3}\frac{p+1}{1-\delta^{-p-1}}\frac{a_{\rm in}}{a_p}\int_1^\delta u^{-p-1} b_{3/2}^{(1)}\left(\frac{1}{u}\frac{a_p}{a_{\rm in}}\right) du, \qquad \text{(E.7)} \\
\phi_2^c &= \frac{4}{15}\frac{p+q+2}{1-\delta^{-p-q-2}}\left(\frac{a_{\rm in}}{a_p}\right)^2 \int_1^\delta u^{-p-q-1} b_{3/2}^{(2)}\left(\frac{1}{u}\frac{a_p}{a_{\rm in}}\right) du. \qquad \text{(E.8)}
\end{aligned}$$

Figure E.1 shows the behavior of ϕ_1^c and ϕ_2^c as a function of $a_p/a_{\rm in}$, computed for different values of p, q, and δ. For clarity, I have plotted the curves of ϕ_1^c and ϕ_2^c in separate panels. It can be seen that ϕ_i^c $(i = 1,\ 2)$ mainly depend on $a_p/a_{\rm in}$, showing weak dependence on the disc model. Indeed, regardless of (p, q, δ), I find that $\phi_i^c \to 1$

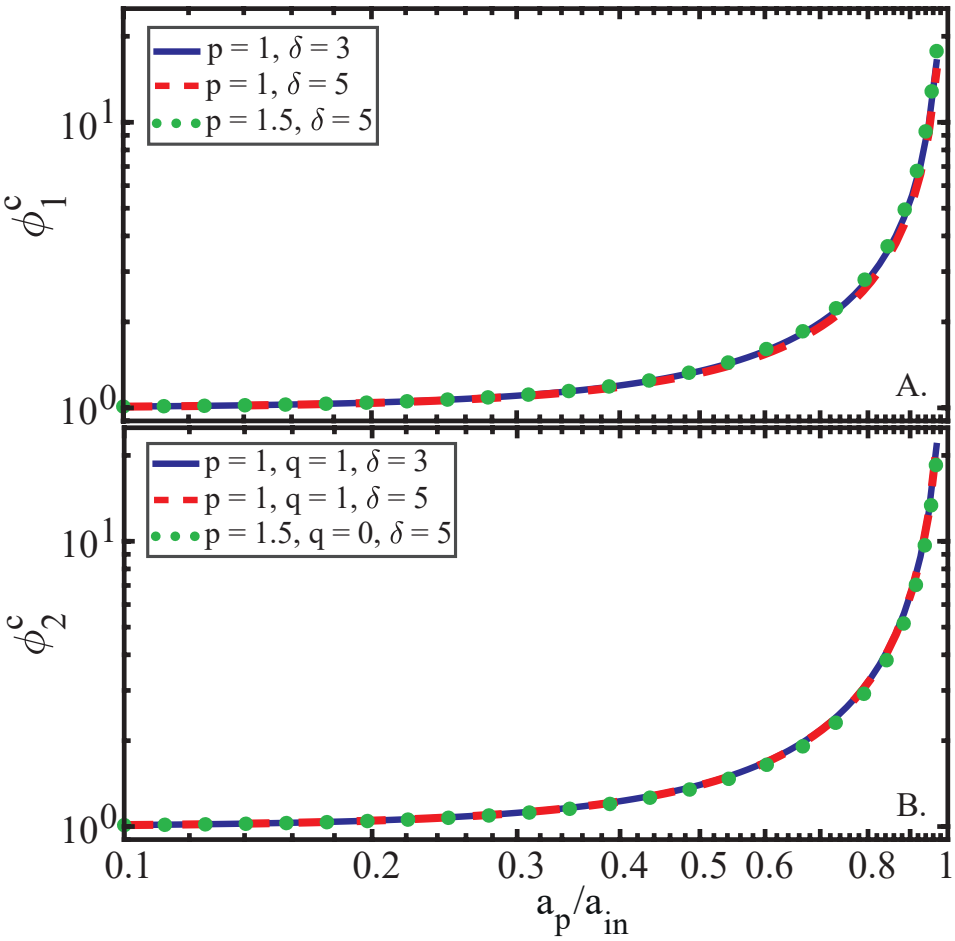

Figure E.1. The behavior of the correction factors ϕ_1^c (panel A, Equation (E.7)) and ϕ_2^c (panel B, Equation (E.8)) as a function of $a_p/a_{\rm in}$. The calculations assume different disc models specified by the values of p, q, and $\delta \equiv a_{\rm out}/a_{\rm in}$ as explained in the legend. Both ϕ_1^c and ϕ_2^c approach unity as $a_p/a_{\rm in} \to 0$, and they diverge as $a_p/a_{\rm in} \to 1$.

for $a_p/a_{\rm in} \to 0$, while in the limit $a_p/a_{\rm in} \to 1$, I find that ϕ_i^c diverge. This divergence follows from the fact that $b_{3/2}^{(m)}(\alpha) \to (1-\alpha)^{-2}$ when $\alpha \to 1$.

Finally, I note that inserting Equations (5.2) and (E.5) into Equation (E.3) results in the expression for $A_{d,p}$ given by Equation (5.7). A similar expression was found by Petrovich et al. (2019) (see also Ward, 1981; Rafikov, 2013).

Appendix F

Analytic Expression for Resonance Widths

The width w of a given resonance at $a = a_{\rm res}$ can be approximated by using the fact that

$$A\left(a_{\rm res} + \frac{w}{2}\right) - A\left(a_{\rm res} - \frac{w}{2}\right) \approx w \times \left.\frac{dA}{da}\right|_{a_{\rm res}} . \tag{F.1}$$

Additionally, Equation (5.19) allows one to write

$$A\left(a_{\rm res} \pm w/2\right) \approx A_{d,p} \mp \tilde{e}^{-1} B_p(a_{\rm res}) \times {\rm sgn}\left[dA/da\right]_{a_{\rm res}} , \tag{F.2}$$

where ${\rm sgn}(x) = x/|x|$ is the sign function introduced to account for the fact that resonances occurring at $\simeq a_{\rm in}$ have $dA/da > 0$, while those farther away have $dA/da < 0$; see Figure 5.2. Substituting Equation (F.2) into Equation (F.1), I thus arrive at

$$\frac{w}{a_{\rm in}} \approx \frac{2}{a_{\rm in}} \left.\left|\frac{B_p(a)\tilde{e}^{-1}}{dA/da}\right|\right|_{a_{\rm res}} . \tag{F.3}$$

The above expression can be further simplified by considering the approximate forms of A_p and A_d in the limits of $a_p/a_{\rm res} \to 0$ and $a_{\rm in} \ll a_{\rm res} \ll a_{\rm out}$, respectively. In this case, I can approximate the derivative of $A = A_p + A_d$ in the following fashion

$$\left.\frac{dA_p}{da}\right|_{a_p \ll a} = \frac{-7}{2a} A_p, \quad \left.\frac{dA_d}{da}\right|_{\psi_1 = const} = \frac{1-2p}{2a} A_d, \tag{F.4}$$

and expression (F.3) reduces to

$$\frac{w}{a_{\rm in}} \approx 4\frac{a_{\rm res}}{a_{\rm in}}\left|\frac{B_p(a_{\rm res})\tilde{e}^{-1}}{7A_p(a_{\rm res})+(2p-1)A_d(a_{\rm res})}\right|. \tag{F.5}$$

Inserting the condition for secular resonances, i.e., Equation (5.16) or Equation (5.18), into the above expression for $p = 1$, and taking the limits $a_p/a_{\rm in} \to 0$ (so one can use the asymptotic behavior of $b_s^{(m)}(\alpha)$) and $a_{\rm in} \ll a_{\rm res} \ll a_{\rm out}$, I arrive at the scaling relationship given by Equation (5.20).

Appendix G

Constructing Maps of Disc Surface Density

Here, I provide some technical details about how I convert the eccentricity–apsidal angle distribution of planetesimals into maps of disc surface density in Chapters 5.

I first begin by assigning a mass m_i to each considered planetesimal in a given annulus of the disc (which, for instance, in Chapter 5 are $N = 5000$ in number; see Section 5.2.3). Given that in my calculations in Chapters 5 and 6 the planetesimals are initiated on circular orbits, the planetesimal masses can be determined from their initial semimajor axis distribution – which remains constant in the secular approximation (Murray & Dermott, 1999, see also Chapter 3). This can be done by using the relationship $dm(a) = 2\pi a \Sigma_d(a) da$ (Statler, 2001; Davydenkova & Rafikov, 2018) relating the mass distribution per unit semimajor axis to the density distribution (which, e.g., in Chapters 5 and 6 is given by Equation 5.1 with $p = 1$; see Section 5.2.1). The self-consistency of this initial mass assignment to planetesimals – which are essentially treated as massless particles in the analytical model described in Chapter 5 (see, e.g., Section 5.2) – is discussed in Section 5.8.1. Note that for the softened N-ring simulations presented later in Chapter 6, the assignment of masses m_i to the disc rings is already well justified; see Section 6.2 for details.

At a given time of the evolution, I then populate every planetesimal's orbit with N_{np} new particles (taken to be, e.g., $N_{np} = 10^4$ in Chapter 5: each with mass m_i/N_{np}, orbital elements similar to the parent planetesimal, but with randomly distributed mean anomalies l between 0 and 2π. This procedure is motivated by the orbit-averaging principle (Murray & Dermott, 1999). I also note that this procedure effectively increases the number of evolved planetesimals (from N to $N \times N_{np}$), enhancing the quality of the resultant maps of disc surface density. Next, I numerically solve for each new particle's

eccentric anomaly ϵ using Kepler's equation (Murray & Dermott, 1999),

$$l = \epsilon - e \sin \epsilon, \tag{G.1}$$

and compute the heliocentric position (X, Y) of each particle along its orbit via (Sridhar & Touma, 1999; Binney & Tremaine, 2008):

$$\begin{pmatrix} X \\ Y \end{pmatrix} = a \begin{pmatrix} \cos \varpi & -\sin \varpi \\ \sin \varpi & \cos \varpi \end{pmatrix} \cdot \begin{pmatrix} \cos \epsilon - e \\ \sqrt{1 - e^2} \sin \epsilon \end{pmatrix}. \tag{G.2}$$

Finally, I bin the positions of all $N \times N_{np}$ particles in the Cartesian system centred at the host star (e.g., with a resolution of 400×400 pixels in Chapter 5), compute the total mass per bin, and divide by its area to arrive at the disc surface density distribution, Σ, at a given time. Note that this also allows me to trivially obtain the azimuthally averaged surface density profile $\langle \Sigma \rangle$ as a function of radial distance r, where $r = \sqrt{X^2 + Y^2} = a(1 - e \cos \epsilon)$, by splitting the disc into annular bins.

Appendix H

Parameters of N-ring Simulations

The results and analysis presented in Chapter 6 are based on the outcomes of a set of 67 simulations carried out using the softened N-ring model described in Section 6.2. The set of parameters describing the disc–planet systems considered is summarised in Table H.1. In each simulation, except otherwise stated, (i) the central star's mass was taken to be $M_c = 1.09 M_\odot$; (ii) the planet's initial apsidal angle was set to $\varpi_p(0) = 0$; (iii) the planetesimals' initial eccentricities and apsidal angles were set to $e(0) = 10^{-4}$ and $\varpi(0) = -\pi/2$, respectively; (iv) the disc was modelled as a collection of $N = 5000$ softened rings, each with a softening parameter of $H = 0.1$; (v) the ring representing the planet was left unsoftened; and (vi) integrations were stopped after 200 Myr.

Table H.1. Parameters of the disc–planet systems considered in Chapter 6. The combinations of M_d, m_p, and a_p are chosen from the parameter space portrayed in Figure 5.8, both from within and outside the allowed region. Outcomes 1 and 0 signify whether the planetesimal eccentricities achieved a minimum value of 1 at the resonance (i.e., $e(a_{\rm res}) \geq 1$) or were in the process of doing so, respectively, by the time the simulation was stopped. Note that this only applies to the 'nominal' N-ring simulations and not to the 'full' N-ring simulations; see Section 6.2.2.3.

Simulation	$M_d(M_\oplus)$	$m_p(M_J)$	$a_p(au)$	$e_p(0)$	Outcome	Notes
1	10	0.31	25.96	0.05	0	
2	...	...	...	0.10	1	
3	20	0.59	25.38	0.05	0	
4	...	...	...	0.10	1	
5	50	1.40	24.46	0.05	1	
6	...	...	...	0.10	1	

Continued on next page

Table H.1 – *Continued from previous page*

Simulation	$M_d(M_\oplus)$	$m_p(M_J)$	$a_p(au)$	$e_p(0)$	Outcome	Notes
7	100	2.74	23.63	0.05	1	
8	...	...	...	0.10	1	
9	10	0.43	15	0.05	0	
10	...	...	...	0.10	0	
11	...	...	...	0.30	1	Stopped at 150 Myr
12	20	0.86	...	0.05	0	
13	...	...	...	0.10	1	
14	...	...	...	0.30	1	Stopped at 150 Myr
15	50	2.15	...	0.05	1	
16	...	...	...	0.10	1	
17	...	...	...	0.30	1	Stopped at 150 Myr
18	100	4.30	...	0.05	1	
19	...	...	...	0.10	1	
20	...	...	...	0.30	1	Stopped at 150 Myr
21	10	0.3	20	0.05	0	
22	...	...	...	0.10	0	
23	...	...	...	0.30	1	Stopped at 100 Myr
24	20	0.6	...	0.05	1	**`Model A` of Chapter 5** Stopped at 500 Myr.
25	...	...	...	0.10	1	
26	...	...	...	0.30	1	Stopped at 100 Myr
27	50	1.5	...	0.05	1	
28	...	...	...	0.10	1	
29	...	...	...	0.30	1	Stopped at 100 Myr
30	100	3	...	0.05	1	
31	...	...	...	0.10	1	Stopped at 150 Myr
32	...	...	...	0.30	1	Stopped at 100 Myr
33	30	1.68	12.8	0.05	0	
34	...	...	...	0.10	1	Stopped at 150 Myr

Continued on next page

Table H.1 – *Continued from previous page*

Simulation	$M_d(M_\oplus)$	$m_p(M_J)$	$a_p(au)$	$e_p(0)$	Outcome	Notes
35	...	...	...	0.20	1	Stopped at 150 Myr
36	...	...	...	0.30	1	Stopped at 150 Myr
37	100	5.59	...	0.05	1	
38	...	...	...	0.10	1	
39	...	...	...	0.20	1	
40	...	...	...	0.30	1	
41	21	0.81	16.1	0.05	0	
42	...	...	...	0.10	1	Stopped at 150 Myr
43	...	...	...	0.20	1	Stopped at 150 Myr
44	...	...	...	0.30	1	Stopped at 150 Myr
45	70	2.72	...	0.05	1	
46	...	...	...	0.10	1	
47	...	...	...	0.20	1	
48	...	...	...	0.30	1	
49	10.5	0.29	22.04	0.05	0	
50	...	...	...	0.10	0	Stopped at 150 Myr
51	...	...	...	0.20	1	Stopped at 150 Myr
52	...	...	...	0.30	1	Stopped at 150 Myr
53	35	0.97	...	0.05	1	
54	...	...	...	0.10	1	
55	...	...	...	0.20	1	
56	...	...	...	0.30	1	
57	121.2	12	9.26	...	1	Stopped at 100 Myr
58	50.5	5	...	...	1	Stopped at 100 Myr
59	10	0.99	...	...	1	Stopped at 300 Myr
60	64.3	12	6.59	...	1	
61	20	3.73	...	...	1	
62	10	1.87	...	...	1	Stopped at 400 Myr
63	50	25.1	3.94	...	1	Stopped at 300 Myr

Continued on next page

Table H.1 – *Continued from previous page*

Simulation	$M_d(M_\oplus)$	$m_p(M_J)$	$a_p(au)$	$e_p(0)$	Outcome	Notes
64	24	12	...	...	1	Stopped at 300 Myr
65	10	5.02	...	...	1	Stopped at 500 Myr
66	95	15.8	7	0.05	1	**`Model B` of Chapter 5**
						Stopped at 400 Myr
67	6	0.2	26.93	...	0	**`Model C` of Chapter 5**
						Stopped at 600 Myr

Appendix I

Planetesimal Evolution in the Case of a Circularising Planet

In this Appendix, I derive explicit expressions describing the secular evolution of planetesimals orbiting in the combined potential of the debris disc and the precessing planet, accounting for the effects of the exponential decay of the planetary eccentricity (i.e., resonant friction, Section 6.4.2) in an *ad hoc* manner.

Unfortunately, a rigorous analytical expression cannot be derived for the generic evolution studied in Chapter 6. Thus, here I make some simplifying assumptions to try and capture the important features of the system's secular evolution. To this end, I make use of the expression of the disturbing function given by Equation (5.8) – which, I remind, was derived for the case of an axisymmetric disc potential (Chapter 5) – and introduce two modifications to account for the effects resulting from the non-axisymmetric torque of the disc on the planet (but not on itself). First, I replace the planetary precession rate $A_{d,p}$ by $\dot{\varpi}_p$, which we saw is a constant depending on the planetary semimajor axis such that $\dot{\varpi}_p/A_{d,p} \leq 1$; see Section 6.4.1. This replacement essentially guarantees that we are still working in a frame co-precessing with the planetary orbit, an assumption inherent in the derivation of Equation (5.8). Second, I multiply the non-axisymmetric component of the planetary potential B_p, which recall is $\propto e_p$ (Equation 5.6), by an exponential factor of $\exp(-Dt/2)$. This modification essentially models the decay of the planetary eccentricity resulting due to the resonant friction; see Section 6.4.2.

Implementing the above two modifications into Equation (5.8), the disturbing function governing the evolution of planetesimals now reads as:

$$R = na^2 \left[\frac{1}{2} \left(A - \dot{\varpi}_p \right) e^2 + B_p(0) \, \exp(-Dt/2) e \cos \Delta\varpi \right], \tag{I.1}$$

where $B_p(0)$ must be understood as being evaluated using the planet's initial eccentricity, $e_p(0)$, and as before, $A = A_d + A_p$ (Equations 5.4 and 5.5). Then the evolution of the planetesimal eccentricity vector $\mathbf{e} = (K, H) = e(\cos\varpi, \sin\varpi)$ is governed by (see Section 5.2.3):

$$\begin{aligned} \frac{dK}{dt} &\approx \frac{-1}{na^2}\frac{\partial R}{\partial H} = -(A - \dot{\varpi}_p)H, \\ \frac{dH}{dt} &\approx \frac{1}{na^2}\frac{\partial R}{\partial K} = (A - \dot{\varpi}_p)K + B_p(0)\,\exp(-Dt/2). \end{aligned} \tag{I.2}$$

Note that when the non-axisymmetric torque of the disc on the planet is ignored, one has $\dot{\varpi}_p \to A_{d,p}$, $D = 0$, and constant e_p, so that the system of equations (I.2) reduce to those in Chapter 5 – namely, Equations (5.9).

Equations (I.2) can be solved using standard methods. Indeed, assuming planetesimals are initiated on circular orbits, i.e., $K(0) = H(0) = 0$, their eccentricities and longitudes of pericentre evolve as follows:

$$e(t) = \sqrt{\frac{4B_p^2(0)}{4(A - \dot{\varpi}_P)^2 + D^2}\left[\exp(-Dt) - 2\,\exp(-Dt/2)\cos[(A - \dot{\varpi}_P)t] + 1\right]}, \tag{I.3}$$

$$\tan\Delta\varpi(t) = \frac{-2(A - \dot{\varpi}_p)\sin[(A - \dot{\varpi}_p)t] + D\Big\{\exp(-Dt/2) - \cos[(A - \dot{\varpi}_p)t]\Big\}}{-2(A - \dot{\varpi}_p)\Big\{\cos[(A - \dot{\varpi}_p)t] - \exp(-Dt/2)\Big\} + D\sin[(A - \dot{\varpi}_p)t]}. \tag{I.4}$$

While seemingly complicated, Equations (I.3) and (I.4) can directly explain several important features of the planetesimal dynamics described in Section 6.3.1 when taken in the limit of $D \approx 0$, in which case one arrives at the following:

$$e(t) \approx \left|\frac{2B_p(0)}{A - \dot{\varpi}_p}\left(1 - \frac{1}{4}Dt\right)\sin\left(\frac{A - \dot{\varpi}_p}{2}t\right)\right| + \mathcal{O}(D^2), \tag{I.5}$$

$$\tan\Delta\varpi(t) \approx \tan\left(\frac{A - \dot{\varpi}_p}{2}t - \frac{\pi}{2}\right) + \mathcal{O}(D). \tag{I.6}$$

This completes my derivation of the general solution describing the evolution of planetesimal orbits in the case of a circularising planet.

Appendix J

Analytic Expression for Planetary Precession Rates due to Eccentric Discs

In this Appendix, I provide the derivation of Equation (6.17) given in Chapter 6 for the apsidal precession rate of the planet due to the gravity of the debris disc, accounting for both the axi- and non-axisymmetric components.

To this end, I revert to the notation of Chapter 5 and assume that the disc is a continuous entity – rather than composed of N softened rings as in Chapter 6. I further assume that the disc is composed of apse-aligned orbits (i.e., $d\varpi_d/da = \text{const}$) and characterised by power-law profiles for the surface density $\Sigma_d(a) \propto a^{-p}$ and eccentricity $e_d(a) \propto a^{-q}$ distributions (Equations 5.1 and E.1). Within this set of assumptions (which I justify below), one can then use the unsoftened disturbing function $R_{d,p}$ given by Equation (E.2) of Appendix E to determine the time evolution of the planetary apsidal line as follows:

$$\dot{\varpi}_p = \frac{1}{n_p a_p^2 e_p} \frac{\partial R_p}{\partial e_p} = A_{d,p} + \frac{B_{d,p}}{e_p} \cos \Delta\varpi. \tag{J.1}$$

where $\Delta\varpi \equiv \varpi_p - \varpi_d$, and the terms $A_{d,p}$ and $B_{d,p}$ are given by Equations (E.3) and (E.4), respectively. Note that the use of the unsoftened disturbing function $R_{d,p}$ (Equation E.2) is justified by the fact that softening the disc potential is not required when studying the dynamics of particles orbiting outside the disc, i.e., where $\Sigma_d(a) = 0$; see Section 4.5. Then, substituting Equations (E.3) and (E.4) into Equation (J.1), one

finds that

$$\dot{\varpi}_p = A_{d,p}\left[1 + \frac{1}{2}\frac{e_d(a_{\rm in})}{e_p}\frac{\phi_2}{\phi_1}\cos\Delta\varpi\right], \tag{J.2}$$

where the coefficients ϕ_1 and ϕ_2 are given by Equations (E.5) and (E.6), respectively, and $e_d(a_{\rm in})$ is the disc eccentricity at the inner disc edge. The above expression can be further simplified by making two assumptions: first, that the disc and the planet remain apse-aligned throughout their evolution, i.e., $\Delta\varpi(t) = 0$; and second, that the disc eccentricity follows the profile forced by the planet in the limit of a massless disc, i.e., $e_d(a) = |e_{\rm forced,p}| = B_p/A_p \propto 1/a$ (Equation 5.14). Inserting these assumptions into Equation (J.2), one then easily arrives at the first line of Equation (6.17). Finally, taking the limit of $a_p/a_{\rm in} \to 0$ – so one can use the asymptotic behaviour of the Laplace coefficients $b_s^{(m)}(\alpha)$ and approximate $\phi_1^c \approx \phi_2^c \approx 1$ (see Figure E.1) – and setting $p = 1$, $\delta \equiv a_{\rm out}/a_{\rm in} = 5$, and $q = 1$, one arrives at the second line of Equation (6.17).

Having completed the derivation of the expression of $\dot{\varpi}_p/A_{d,p}$ as given by Equation (6.17), let us now justify the assumptions that went into it. At the outset, however, it is important to note that the gravitational forces exerted by the disc on the planet is dominated by the disc rings (or particles) that are in close proximity to the planet; simply because of the radial dependence of gravity. In other words, the dynamical state of the disc's outer parts does not significantly affect the gravitational potential exerted by the disc at the location of the planet. Thus, for the purposes of this Appendix, one can approximate the whole disc structure by that of the inner parts, i.e., the region spanning over $a_{\rm in} \leq a \lesssim a_{\rm res}$. Planetesimal orbits in this region tend to precess faster than the planet (e.g., Sections 5.3 and 6.3.1); however, they attain their maximum eccentricity when apse aligned with the planet, and it is at this point that the non-axisymmetric component of the disc gravity becomes important. This justifies the assumption of $\Delta\varpi(t) \approx 0$. Additionally, as the dynamics of planetesimals orbiting at $a_{\rm in} \lesssim a \lesssim a_{\rm res}$ is dominated by the planet's gravity, their eccentricities oscillate between the initial values of 0 and $e_{m,p} = |2e_{\rm forced,p}|$, and so on average $e_d(a) \approx |e_{\rm forced,p}|$, which I assume to be constant in time. While this is not entirely accurate, in the sense that the planetary eccentricity actually decays in time due to resonant friction (see Sections 6.3.1 and 6.4.2), I find that the maximum planetesimal eccentricities scale down together with the planetary eccentricity (see Equation 6.15) so that the ratio $e(t)/e_p(t)$ – which is what appears in Equation (J.2) – can be approximated as a constant. This justifies the adopted assumption about the $e_d(a)$ profile.