DYNAMICS OF SMALL BODIES IN PLANETARY SYSTEM FORMATION

Kedron Silsbee

A DISSERTATION
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

RECOMMENDED FOR ACCEPTANCE
by the Program in
Astrophysical Sciences

Advisors: Roman Rafikov & Scott Tremaine

SEPTEMBER 2017
Abstract

One of the most exciting astronomical developments of the past two decades has been the wealth and diversity of exoplanetary systems. Among the more exotic discoveries is a collection of planets in tight binary star systems. The first three chapters of this thesis focus on planet formation around binary stars. We assume that cores of giant planets form via collisional agglomeration of small planetesimals. A simple-minded estimate suggests that collision velocities between kilometer-sized planetesimals in some planet-hosting binary systems would be too large by a factor of \( \sim 1000 \) for them to grow in mutual collisions rather than being destroyed. To study this issue in more detail, we developed a model for the dynamics of planetesimals in binary systems. Chapter 1 discusses the gravitational effects of the disk on the planetesimals, and derives an expression for the disturbing function of an eccentric disk. Chapter 2 applies the results of chapter 1, as well as some additional work done by Roman Rafikov and myself incorporating the effect of gas drag from the disk, to the particular case of circumbinary planets. Chapter 3 describes our ongoing efforts to simulate the coagulation process, using the rates and collisional outcomes calculated in our other works.

Chapters 4 and 5 address the topic of small bodies in our own solar system. Recent wide-field surveys have discovered a few thousand minor solar-system bodies at tens of AU from the Sun. Upcoming surveys such as LSST should find at least an order of magnitude more. Chapter 4 describes simulations of long-period comet orbits, and predicts the orbital element distribution of the long-period comet population with perihelion between 5 and 45 AU. Chapter 5 investigates what happens if there are several Mars–Earth mass bodies left over after the giant planets are assembled. We find that their influence naturally creates a detached disk (a set of moderately inclined objects with perihelia well beyond the orbit of Neptune, but aphelia inside 1,000 AU), and also suggest the possibility that there could be an undetected Mars–Earth sized planet a few hundred AU from the Sun.
Acknowledgements

I have had the privilege of working with two outstanding advisors — Roman Rafikov and Scott Tremaine. Without their ideas and support, none of the papers in this thesis would have been written, and my understanding of dynamics would be much poorer. Whenever I thought I understood something, I could always count on one of them to probe one level deeper.

I also had the pleasure of working with Bruce Draine on a semester project. While that paper has been excluded from this thesis for topical reasons, it was nevertheless greatly rewarding to work with Bruce, and I learned a lot of physics in his office during our collaboration.

I would like to acknowledge Munan Gong for many engaging discussions of astrophysics, emotional support throughout the five years I was here, and more prosaically, dealing patiently with my many computer issues, including the formatting of this thesis.

Finally, I would like to thank my parents for encouraging me to pursue my esoteric interests without worrying too much were they led, and my grandfather for sharing interesting physics toys and problems with me.
Relation to Published Work

Chapter 1 is published in ApJ \cite{Silsbee2015b}. This is an extension of previous work by Roman Rafikov \cite{Rafikov2013ab}. I presented a poster at the AAS DPS meeting in November of 2014 based in part on this work and in part on work led by Roman Rafikov \cite{Rafikov2015ab}, and gave a Wunch talk about it in the astronomy department at Princeton.

Chapter 2 is published as \cite{Silsbee2015a}. This is an extension of work led by Roman Rafikov \cite{Rafikov2015ab}, and was done in collaboration with Roman Rafikov. I presented on this work at the TRENDY conference in Haifa in June, 2015, and at the SPF1 conference in Tucson in March 2015.

Chapter 3 represents work in progress, and is so far unpublished. I have given talks related to the first three chapters of this work, as well as the work in \cite{Rafikov2015ab} in 2016 at Toronto, Harvard, and UCSC, and in 2017 at Cambridge (DAMTP) and the London DDA meeting.

Chapter 4 is published in the astronomical journal in 2016 \cite{Silsbee2016} and was done in collaboration with Scott Tremaine. I talked about this work at the university of Toronto in April of 2016.

Chapter 5 is unpublished. It was done in collaboration with Scott Tremaine. I presented a poster about this at the 2017 DDA meeting in London. This chapter will soon be submitted to the Astronomical Journal.
Bibliography


Silsbee, K., & Tremaine, S. 2016, AJ, 152, 103
# Contents

| Abstract | iii |
| Acknowledgements | iv |
| Relation to Published Work | v |

## 1 Planet Formation in Binaries: Dynamics of Planetesimals Perturbed by the Eccentric Protoplanetary Disk and the Secondary

1.1 Abstract ................................................................. 1
1.2 Introduction ............................................................ 2
1.3 Problem Setup ........................................................... 4
1.4 Basic Equations .......................................................... 7
  1.4.1 Disturbing Function Due to the Disk ......................... 7
  1.4.2 Disturbing Function Due to the Binary ....................... 11
  1.4.3 Full Planetesimal Disturbing Function ....................... 11
  1.4.4 Evolution Equations and Their Solution .................... 12
  1.4.5 Comparison with Direct Orbit Integrations ................ 14
1.5 Planetesimal Eccentricity Behavior ............................... 17
  1.5.1 Case DD: Disk Dominates Both Precession and Excitation .... 19
  1.5.2 Case BB: Binary Dominates Both Precession and Excitation .... 21
  1.5.3 Case BD: Binary Dominates Precession, Disk Dominates Excitation .. 21
  1.5.4 Case DB: Disk Dominates Precession, Binary Dominates Excitation .... 24
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.3 Radial Behavior of $e_r$</td>
<td>77</td>
</tr>
<tr>
<td>2.6 Dynamics with Disk-Dominated Excitation and Disk Precession</td>
<td>79</td>
</tr>
<tr>
<td>2.6.1 Effects of Disk Precession</td>
<td>81</td>
</tr>
<tr>
<td>2.7 Planetesimal Dynamics Around a Precessing Binary</td>
<td>84</td>
</tr>
<tr>
<td>2.7.1 Binary Precession Rates</td>
<td>85</td>
</tr>
<tr>
<td>2.7.2 Axisymmetric Disk</td>
<td>87</td>
</tr>
<tr>
<td>2.8 Dynamics With Both Binary and Disk Gravity and Both Binary and Disk Precession</td>
<td>88</td>
</tr>
<tr>
<td>2.9 Collisonal Outcomes and Growth of Planetesimals</td>
<td>89</td>
</tr>
<tr>
<td>2.9.1 Planetesimal Growth in the Disk-Dominated Excitation Regime</td>
<td>92</td>
</tr>
<tr>
<td>2.9.2 Collisional Outcomes in an Axisymmetric Disk with Binary Precession</td>
<td>98</td>
</tr>
<tr>
<td>2.9.3 Planetesimal Growth Near the Star in the Limit of Short Stopping Times</td>
<td>100</td>
</tr>
<tr>
<td>2.10 Discussion</td>
<td>103</td>
</tr>
<tr>
<td>2.10.1 Erosion</td>
<td>103</td>
</tr>
<tr>
<td>2.10.2 Growth via “Lucky” Bodies</td>
<td>105</td>
</tr>
<tr>
<td>2.10.3 Limits of the Apsidally Aligned Regime</td>
<td>106</td>
</tr>
<tr>
<td>2.10.4 Disk Density Structure</td>
<td>108</td>
</tr>
<tr>
<td>2.10.5 Non-Secular Terms in the Disturbing Function</td>
<td>110</td>
</tr>
<tr>
<td>2.11 Comparison with Previous Work on Planet Formation in Binaries</td>
<td>111</td>
</tr>
<tr>
<td>2.11.1 Previous Work on Circumbinary Planet Formation</td>
<td>111</td>
</tr>
<tr>
<td>2.11.2 Large Initial Planetesimals</td>
<td>113</td>
</tr>
<tr>
<td>2.11.3 Differences Between P-type and S-type Binaries</td>
<td>114</td>
</tr>
<tr>
<td>2.12 Summary</td>
<td>115</td>
</tr>
<tr>
<td>2.13 Appendix A: Planetesimal Dynamics with Linear Drag</td>
<td>117</td>
</tr>
<tr>
<td>2.14 Appendix B: Binary Precession Rate</td>
<td>118</td>
</tr>
<tr>
<td>2.14.1 General Relativistic Precession</td>
<td>118</td>
</tr>
<tr>
<td>2.14.2 Precession Due to Stellar Quadrupoles Induced by Tides and Rotation</td>
<td>118</td>
</tr>
</tbody>
</table>
3 A Coagulation/Fragmentation Simulation for Planetesimals in Binary Systems

3.1 Abstract
3.2 Introduction
3.3 Model System
3.4 Results for Single-Annulus Coagulation
  3.4.1 Comparison with Simple Heuristics
3.5 Conclusions
3.6 Appendix A: Functionality of the Code
  3.6.1 Planetesimal Collisions - Overview
  3.6.2 Collisional Outcomes
  3.6.3 Determination of Collisional Velocity
  3.6.4 Collision Rate
  3.6.5 Radial Inspiral
  3.6.6 Time Step Determination
3.7 Appendix B: Analytic Tests of the Code
  3.7.1 Tests Against Known Solutions of the Coagulation Equation
  3.7.2 Fragmentation Cascades
  3.7.3 Mass Concentration Due to Radial Inspiral
3.8 Appendix C: Collisions Between Bodies in the Same Size-Bin

4 Modeling the Nearly Isotropic Comet Population in Anticipation of LSST Observations

4.1 Abstract
4.2 Introduction
5.8 Conclusions
Chapter 1

Planet Formation in Binaries: Dynamics of Planetesimals Perturbed by the Eccentric Protoplanetary Disk and the Secondary

1.1 Abstract

Detections of planets in eccentric, close (separations of $\sim 20$ AU) binary systems such as $\alpha$ Cen or $\gamma$ Cep provide an important test of planet formation theories. Gravitational perturbations from the companion are expected to excite high planetesimal eccentricities resulting in destruction, rather than growth, of objects with sizes of up to several hundred km in collisions of similar-size bodies. It was recently suggested that gravity of a massive axisymmetric gaseous disk in which planetesimals are embedded drives rapid precession of their orbits, suppressing eccentricity excitation. However, disks in binaries are themselves expected to be eccentric, leading to additional planetesimal excitation. Here we develop secular theory of eccentricity evolution for planetesimals perturbed by the gravity of an elliptical protoplanetary disk (neglecting gas drag) and the companion. For the first time we derive an expression for the disturbing function due to an eccentric disk, which can be used for a variety of other astrophysical problems. We obtain explicit analytical solutions for planetesimal eccentricity evolution neglecting gas drag and delineate four different regimes of dynamical excitation. We show that in systems with massive ($\gtrsim 10^{-2} M_\odot$) disks, planetesimal
eccentricity is usually determined by the gravity of the eccentric disk alone, and is comparable to the disk eccentricity. As a result, the latter imposes a lower limit on collisional velocities of solids, making their growth problematic. In the absence of gas drag this fragmentation barrier can be alleviated if the gaseous disk rapidly precesses or if its own self-gravity is efficient at lowering disk eccentricity.

1.2 Introduction

Planet-hosting binary systems with separations of several tens of AU present an interesting testbed for planet formation theories. Strong gravitational perturbations induced by the companion excite high eccentricities of planetesimals out of which planets form. Agglomeration of these objects into bigger bodies in mutual collisions, most effective at low relative speeds because of gravitational focussing, may become very ineffective. In a strongly dynamically excited environment planetesimals would destroy each other instead of growing. This fragmentation barrier presents a very serious problem for planetary growth in binaries.

This issue is particularly severe for binaries with small separation. At the moment, we know (Chauvin et al. 2011; Dumusque et al. 2012) of five planet-hosting systems with eccentric companions (eccentricities \( \gtrsim 0.4 \)) and semimajor axes of about 20 AU. Three of them — HD196885, \( \gamma \) Cep, HD 41004 — harbor giant planets with masses above that of the Jupiter at 1.6 – 2.6 AU. At these separations, the eccentricity of a free particle can easily reach 0.1 (Heppenheimer 1978), leading to collisions at speeds of several km s\(^{-1}\) and resulting in destruction of even rather massive (several hundred km in size) objects in collisions, as well as smaller planetesimals. Two other systems — \( \alpha \) Cen and Gl 86 — harbor planets at \( \lesssim 0.1 \) AU but even these objects have likely formed further out and then migrated in.

Planetesimal agglomeration must proceed in gaseous protoplanetary disks. It has long been recognized that gas drag is an important agent of planetesimal dynamics (Marzari & Scholl 2000; Thébault et al. 2004, 2006, 2008, 2009; Paardekooper et al. 2008), helping lower
relative speeds of planetesimals to some extent. Recently it has also been realized that the gravitational field of a massive protoplanetary disk can have a strong effect on planetesimal dynamics. In particular, Rafikov (2013b, hereafter R13) has shown that an axisymmetric, massive gaseous disk drives fast precession of planetesimal orbits by its gravity, which effectively suppresses eccentricity excitation by the companion. This mechanism permits growth of even 10 km planetesimals at 2 AU as long as the disk is massive ($\sim 0.1M_\odot$) and axisymmetric.

At the same time hydrodynamical simulations of protoplanetary disks in binaries always find that disks perturbed by the companion develop some degree of non-axisymmetry (Okazaki et al. 2002; Kley et al. 2008; Marzari et al. 2009; Paardekooper et al. 2008), which usually manifests itself as a non-zero disk eccentricity. Such a disk has a non-axisymmetric component of its gravitational field which affects planetesimals in a way similar to the binary companion. Thus, one expects an eccentric gaseous disk to drive planetesimal eccentricity excitation (in addition to that produced by the binary companion), an effect absent in the case of an axisymmetric disk studied in R13. Recent work of Marzari et al. (2013) supports this expectation by showing this effect to operate in circumbinary disks, which can also develop eccentric structure and drive eccentricity growth by their gravity.

The goal of this work is to analyze dynamics of planetesimals in the presence of gravitational perturbations due to both the binary companion and the eccentric disk. To focus on purely gravitational effects we neglect gas drag in our calculations (it is taken into account in Rafikov & Silsbee 2015a,b). We explore planetesimal dynamics in the secular approximation, neglecting short-period perturbations of planetesimal orbits that average out over the long time intervals. The majority of our results are derived for the case of a non-precessing disk, which is steady with respect to the orientation of the eccentric orbit of the secondary. However, we also explore planetesimal dynamics in the case of precessing disk.

A significant part of this work is a derivation of the disturbing function due to an eccentric disk, which has been carried out for the first time. Because of the technical nature of this
derivation, which we cover in Appendix A, it can be skipped at first reading. The main results are summarized in the main text.

The structure of the paper is as follows. We outline the problem set-up in §1.3 and present basic equations of planetesimal motion and their solutions for the case of non-precessing disk in §2.4. We analyze our solutions and describe four possible dynamical regimes for planetesimal eccentricity excitation in §1.5. Eccentricity behavior as a function of the distance from the primary is discussed in §1.6. In §1.7 we explore the case of a uniformly precessing disk. Our results are discussed in §1.8 where we cover the implications for planetesimal growth (§1.8.1), ways of lowering planetesimal eccentricity (§1.8.2), and comparison with existing numerical results (§1.8.3). Our findings are summarized in §1.9.

1.3 Problem Setup.

We consider a binary star in which the primary and secondary have masses \( M_p \) and \( M_s \), and define \( \nu \equiv M_s/M_p \). The semimajor axis and eccentricity of the binary are \( a_b \) and \( e_b \), and its orientation is specified by apsidal angle \( \varpi_b \).

Coplanar with the binary and orbiting the primary star (this designation is arbitrary) is the eccentric gaseous disk with a non-axisymmetric surface density distribution \( \Sigma(r_d, \phi_d) \). The disk is eccentric in a sense that trajectories of its fluid elements are confocal ellipses, which in general is not equivalent to \( \Sigma \) being constant along these ellipses (see the discussion of this approximation in §1.8). We define \( r_d \) to be the distance from the common focus of the elliptical fluid trajectories, and \( \phi_d \) to be the polar angle with respect to the disk apsidal line, see Figure 1.1 for illustration. For every such gaseous trajectory with semimajor axis \( a_d \) we can define the disk surface density at the periastron \( \Sigma_p(a_d) \) and the eccentricity of the fluid trajectory \( e_d(a_d) \), which we will simply call disk eccentricity. In general both \( \Sigma_p(a_d) \) and \( e_d(a_d) \) can be arbitrary functions of the fluid semi-major axis \( a_d \), as long as \( e_d(a_d) \) varies slowly enough for the particle trajectories to be non-crossing (Ogilvie 2001).
Statler (1999) has given the following expression for the surface density behavior in such a disk, assuming that the lines of apsides of all elliptical trajectories are aligned:

$$\Sigma(a_d, \phi_d) = \Sigma_p(a_d) \frac{1 - e_d^2 - \zeta e_d(1 + e_d)}{1 - e_d^2 - \zeta e_d [e_d + \cos E(\phi_d)]}, \quad (1.1)$$

where $\Sigma_p(a_d)$ is the surface density at the pericenter ($\phi_d = E = 0$), as a function of the semi-major axis $a_d$, $E(\phi_d)$ is the eccentric anomaly (Murray & Dermott 1999) and $\zeta \equiv d \ln e_d(a_d)/d \ln a_d$. Equation (1.1) has been generalized in Statler (2001) and Ogilvie (2001) to the case of the disk apsidal angle $\varpi_d$ varying with $a_d$ but we will not consider this additional complication here as it adds little new to the physics of our problem. Interestingly, Equation (1.1) predicts that surface density is constant along the elliptical fluid trajectory if $e_d$ is not varying with $a_d$, i.e. $\zeta = 0$.

Throughout this work we assume simple power law scalings

$$\Sigma_p(a_d) = \Sigma_0 \left( \frac{a_{\text{out}}}{a_d} \right)^p, \quad e_d(a_d) = e_0 \left( \frac{a_{\text{out}}}{a_d} \right)^q, \quad (1.2)$$

for $a_{\text{in}} < a_d < a_{\text{out}}$, where $a_{\text{in}}$ and $a_{\text{out}}$ are the semi-major axes of the innermost and outermost fluid trajectories, and $\Sigma_0$ and $e_0$ are the pericenter surface density and eccentricity at the outer edge of the disk. If the semi-major axis of the innermost fluid trajectory $a_{\text{in}} \ll a_{\text{out}}$, as expected for realistic disks, then $\Sigma_0$ can be directly related to the disk mass $M_d \approx 2\pi \int_{a_{\text{in}}}^{a_{\text{out}}} \Sigma_p(a_d)a_d da_d$ enclosed within $a_{\text{out}}$ as

$$\Sigma_0 = \frac{2 - p}{2\pi} \frac{M_d}{a_{\text{out}}^2}, \quad (1.3)$$

where we neglected disk ellipticity (see below) and assumed $p < 2$, so that most of the disk mass is concentrated in its outer part.

We will neglect the precession of the binary apsidal line caused by the gravity of the circumprimary disk, as the corresponding precession period is considerably longer than other
Figure 1.1: Geometry of the problem, showing elliptical trajectories of both the planetesimal (red) and a representative fluid element (blue). Their orientation is shown using different polar angles. Dashed circle illustrates our calculation of the disturbing function in Appendix A.
timescales of the problem. We will also focus predominantly on the case of a non-precessing disk. We cover the precessing disk case in Appendix C and §1.7.

Our focus is on the dynamics of planetesimals embedded in the gaseous disk. We characterize planetesimal orbits by semimajor axis $a_p$, eccentricity $e_p$, and apsidal angle $\varpi_p$.

Even though expression (1.1) does not assume $e_d$ to be small, in the rest of the paper we will take both the disk and planetesimal eccentricities to be small, $e_d(r) \ll 1$ and $e_b \ll 1$.

1.4 Basic Equations

We study planetesimal dynamics taking into account gravitational perturbations from both the binary companion and the eccentric disk. We perform calculations in the secular approximation (Murray & Dermott 1999), by averaging the planetesimal disturbing function $R$ over time thus eliminating the short-period terms, and keeping only the slowly varying contributions up to second order in the planetesimal eccentricity $e_p$ and to lowest order in disk eccentricity $e_d$ (in all terms).

1.4.1 Disturbing Function Due to the Disk

In Appendix A we provide a detailed calculation of the planetesimal disturbing function $R_d$ due to a non-axisymmetric disk with surface density and eccentricity distributions given by Equations (1.2). This calculation is very general and can be applied to an arbitrary eccentric disk, not necessarily around one of the components of the binary. In particular it can be used to study planetesimal motion in a circumbinary disk. This calculation thus represents an important stand-alone result of this work.

We show in Appendix A that in the secular approximation and to lowest order in $e_d$ and $e_p$ the disturbing function due to the eccentric disk with orientation $\varpi_d$ (independent of the
distance from the primary) has the form

\[ R_d = a_p^2 n_p \left[ \frac{1}{2} A_d e_p^2 + B_d e_p \cos (\varpi_p - \varpi_d) \right], \quad (1.4) \]

where

\[ A_d = 2\pi G \frac{\Sigma_p(a_p)}{a_p n_p} \psi_1, \quad (1.5) \]

\[ B_d = \pi G \frac{\Sigma_p(a_p)}{a_p n_p} e_d(a_p) \psi_2, \quad (1.6) \]

where \( n_p \equiv \sqrt{GM_p/a_p^3} \) is the planetesimal mean motion, and dimensionless constants \( \psi_1 \) and \( \psi_2 \) are given by Equations (1.74) and (1.75). In deriving this expression for \( R_d \) we used Equation (1.72), in which we dropped the term independent of \( e_p \).

Coefficients \( \psi_1 \) and \( \psi_2 \) are functions of the power law indices \( p, q \), characterizing the disk structure, as well as the distance \( a_p \) with respect to the disk boundaries. Figure 1.10 shows the behavior of \( \psi_1 \) and \( \psi_2 \) for several values of \( p, q \), and different \( \alpha_1 \equiv a_{in}/a_p \leq 1 \), \( \alpha_2 \equiv a_p/a_{out} \leq 1 \) computed according to Equations (1.74)-(1.75). One can see that for the selected values of \( p \) and \( q \), both \( \psi_1 \) and \( \psi_2 \) converge to values depending only on \( p \) and \( q \) in the limit of \( \alpha_1 \to 0, \alpha_2 \to 0 \). Indeed, in Appendix A we show that as long as

\[ -1 < p < 4 \quad \text{and} \quad -2 < p + q < 5 \quad (1.7) \]

the values of \( \psi_1 \) and \( \psi_2 \) are determined locally, by the surface density and \( e_d \) behavior in the vicinity of \( a_p \). In this case, for a disk spanning more than an about order of magnitude in radius and \( a_{in} \lesssim a_p \lesssim a_{out} \) the gravitational effect of disk parts near the boundaries is not important. Then \( \psi_1 \) and \( \psi_2 \) only weakly depend on \( \alpha_{1,2} \) and can be well approximated by Equations (1.78)-(1.79). Their values in this limit are shown in Figure 1.3 as functions of \( p \) and \( p + q \). This is how these coefficients will be often treated (i.e. as constants) in the
Figure 1.2: Illustration of the convergence properties of coefficients $\psi_1$ and $\psi_2$ characterizing disk-driven precession and eccentricity excitation [Equations (1.5) and (1.6)] as a function of the power law indices $p$ and $q$ determining the radial dependence of disk surface density and eccentricity (Equation 1.2). The unshaded region is a part of parameter space where (far from the edges of the disk) the values of $\psi_1$ and $\psi_2$ are determined by the local disk properties at each radius, and is described by the constraint (1.7). Outside of this region the boundary terms must be accounted for in all of the disk, see Appendix A and Figure 1.10.
Figure 1.3: Dependence of the coefficients (a) $\psi_1$ and (b) $\psi_2$ on power law indices $p$ and $p+q$, correspondingly (blue line). Calculation assumes that conditions (1.7) are fulfilled (unshaded region in Figure 1.2) so that values of $\psi_{1,2}$ are determined by the local disk properties at each radius.

following analysis, though as can be seen in Figure 1.10, this approximation breaks down near the boundaries of the disk.

We verified our analytical derivation of $R_d$ given by Equations (1.4)-(1.6) in several different ways. In particular, in the case of an axisymmetric disk $B_d = 0$, we made sure that in this case $R_d$ coincides with the expressions derived in R13 for surface density profile with $p = 1$ and in Rafikov (2013b) for arbitrary $p$, based on the results of Ward (1981). The accuracy of our results in the case of non-axisymmetric disk is verified by direct integration of particle motion discussed in §1.4.5.
1.4.2 Disturbing Function Due to the Binary

Another perturbation to the planetesimal motion is provided by the companion star. For an external binary companion this is given by (Murray & Dermott 1999)

\[ R_b = a_p^2 n_p \left[ \frac{1}{2} A_b e_p^2 + B_b e_p \cos (\varpi_p - \varpi_b) \right], \tag{1.8} \]

where

\[ A_b = \frac{\nu}{4} n_p \alpha_b^2 b_{3/2}^{(1)}(\alpha_b) \approx \frac{3}{4} n_p \nu \left( \frac{a_p}{a_b} \right)^3, \tag{1.9} \]

\[ B_b = -\frac{\nu}{4} n_p \alpha_b^2 b_{3/2}^{(2)}(\alpha_b) e_b \approx -\frac{15}{16} n_p \nu \left( \frac{a_p}{a_b} \right)^4 e_b. \tag{1.10} \]

Here \( \alpha_b \equiv a_p/a_b \) and

\[ b_s^{(j)}(\alpha) = \frac{1}{\pi} \int_0^{2\pi} \frac{\cos (j \theta) d\theta}{(1 - 2\alpha \cos \theta + \alpha^2)^s} \tag{1.11} \]

stands for the standard Laplace coefficient. The approximate expressions assume \( \alpha_b \ll 1 \), which is a reasonable assumption. Equations (1.9)-(1.10) are valid up to the leading order in \( e_b \ll 1 \), more accurate expressions can be found in Heppenheimer (1978) or R13.

1.4.3 Full Planetesimal Disturbing Function

Given that the binary precession due to disk gravity is slow, the orientation of the orbital ellipse of the secondary can be approximated as fixed in time. Then, without loss of generality we may choose the binary apsidal line as the reference direction, in which case \( \varpi_b = 0 \). The total (disk plus star) disturbing function \( R = R_d + R_b \) is then given by

\[ R = a_p^2 n_p \left[ \frac{1}{2} A_d e_p^2 + B_d e_p \cos (\varpi_p - \varpi_d) \right. \]
\[ + \left. B_b e_p \cos \varpi_p \right], \tag{1.12} \]
\[ A = A_d + A_b. \]  
(1.13)

We now introduce planetesimal eccentricity vector \( \mathbf{e}_p = (k_p, h_p) \), where

\[ k_p = e_p \cos \varpi_p, \quad h_p = e_p \sin \varpi_p. \]  
(1.14)

Then \( R \) can be written in terms of \( h_p \) and \( k_p \) as follows:

\[
R = a_p^2 n_p \left[ \frac{1}{2} A (h_p^2 + k_p^2) + (B_b + B_d \cos \varpi_d) k_p + B_d \sin \varpi_d h_p \right].
\]  
(1.15)

### 1.4.4 Evolution Equations and Their Solution

In secular planar approximation only the eccentricity \( e_p \) and apsidal angle \( \varpi_p \) of the planetesimal orbit vary in time. We study this process by following the evolution of \( k_p \) and \( h_p \) using Lagrange equations (Murray & Dermott 1999)

\[
\frac{dk_p}{dt} = - \frac{1}{n_p a_p^2} \frac{\partial R}{\partial h_p}, \quad \frac{dh_p}{dt} = \frac{1}{n_p a_p^2} \frac{\partial R}{\partial k_p}.
\]  
(1.16)

With \( R \) given by the expression (1.15) the evolution equations become

\[
\frac{dk_p}{dt} = - A h_p - B_d \sin \varpi_d, \quad \frac{dh_p}{dt} = A k_p + B_b + B_d \cos \varpi_d.
\]  
(1.17)

This is the key system of equations for our work, valid as long as the orientation of elliptical fluid trajectories, determined by \( \varpi_d \), is independent of radius.
Note that in deriving this system we did not make any assumptions regarding the time behavior of $\varpi_d$. Thus, $\varpi_d$ in Equations (1.17)-(1.18) can be an arbitrary function of time, which makes this system of equations applicable to rigidly precessing disks as well as disks in which the common apsidal line librates around some equilibrium orientation.

However, for simplicity we start with a case when $\varpi_d = \text{const}$, i.e. when the disk shape is fixed in the frame of the binary. The precessing disk case is covered in §1.7. We solve Equations (1.17)-(1.18) assuming an initially circular planetesimal orbit, i.e. $k_p(0) = h_p(0) = 0$. The solution

$$\begin{cases}
k_p(t) \\
h_p(t)
\end{cases} = e_p(t) = e_{\text{forced},b} + e_{\text{forced},d} + e_{\text{free}},$$

$$e_{\text{forced},b} = -\frac{B_b}{A} \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

$$e_{\text{forced},d} = -\frac{B_d}{A} \begin{bmatrix} \cos \varpi_d \\ \sin \varpi_d \end{bmatrix},$$

$$e_{\text{free}}(t) = A^{-1} \sqrt{B_d^2 + 2B_d B_b \cos \varpi_d + B_b^2} \times \begin{bmatrix} \cos(At + \phi) \\ \sin(At + \phi) \end{bmatrix},$$

is decomposed into three distinct contributions: $e_{\text{forced},b}$ is the forced eccentricity due to binary potential, $e_{\text{forced},d}$ is the forced eccentricity due to disk potential, and $e_{\text{free}}(t)$ is the free eccentricity vector rotating at the precession rate $A$, with the phase $\phi$ given by equation

$$\sin \phi = \frac{B_d \sin \varpi_d}{\sqrt{B_d^2 + 2B_d B_b \cos \varpi_d + B_b^2}}.$$
Variation of eccentricity \( e_p = (h_p^2 + k_p^2)^{1/2} \) is given by a simple formula

\[
e_p(t) = \frac{2}{A} \left| \sin \frac{At}{2} \right| \sqrt{B_d^2 + 2B_dB_b \cos \varpi_d + B_b^2}.
\] (1.24)

This result shows that the maximum eccentricity ranges between \( 2|(|B_d| - |B_b|)/A| \) and \( 2(|B_d| + |B_b|)/A| \) depending on the value of \( \varpi_d \).

For the subsequent discussion we will be using a characteristic eccentricity of

\[
e_{\text{char}} = 2\frac{|B_b| + |B_d|}{|A|},
\] (1.25)

which is an upper bound on the \( e_p \). This estimate ignores the dependence of \( e_p \) on \( \varpi_d \) and overlooks some interesting cases when \( e_p \) can be significantly lower than \( e_{\text{char}} \), e.g. when

\[
|B_b| \approx |B_d| \quad \text{and} \quad \cos \varpi_d \approx -\text{sgn}(B_d B_b),
\] (1.26)

(\( \text{sgn}(z) \) is a sign function) a possibility that is discussed in more detail in \( \S \, 1.8.2 \).

### 1.4.5 Comparison with Direct Orbit Integrations

To test our analytical prescription (1.4)-(1.6) for the disk disturbing function \( R_d \), we compared our theory with the results of direct numerical integration of planetesimal motion in the gravitational field of an eccentric disk. We consider a disk extending from \( a_{\text{in}} = 0.1 \) AU to \( a_{\text{out}} = 5 \) AU and having \( \Sigma_p(1\text{AU}) = 100 \text{ g cm}^{-2} \). To isolate effects of the disk gravity we set the mass of the secondary to zero. The details of our numerical calculations are described in Appendix B.

Numerical results were then compared with analytical solutions obtained in the previous section, and the outcomes are shown in Figure 1.4 in the form of planetesimal eccentricity \( e_p \) and apsidal angle \( \varpi_p \) dependence on time. We tried different initial conditions for the
planetesimal orbit but in this figure we concentrate on the case of zero initial planetesimal eccentricity, when the analytical solution is given by Equation $\text{(1.24)}$ with $B_b = 0$.

One can see that irrespective of the parameters of our integrations the agreement between theory and numerical results is very good. The amplitude of $e_p$ variation is always in excellent agreement with theory (the difference being less than a percent), even for the disk eccentricity at the outer edge as high as $e_0 = 0.2$, see panel (b). The period of secular oscillations is within several percent of our analytical prediction $2\pi/A_d$ given by Equation $\text{(1.5)}$ in the high-eccentricity case $e_0 = 0.2$ for a disk with $p = -q = 1$. However, this discrepancy is considerably smaller in other cases shown.

Such deviations between orbit integrations and linear secular theory (although at much larger amplitude), predominantly in periodicity of variation, have been previously documented in the case of perturbation by the eccentric binary companion alone (Thébault et al. 2006; Barnes & Greenberg 2006). Giuppone et al. (2011) find discrepancies in both the amplitude and period of $e_p$ oscillations at the level of $\sim 50\%$ when secular theory predicts $e_p \gtrsim 0.1$. But, as Figure $\text{1.4b,d}$ clearly demonstrates, the agreement between theory and simulations in the case of a disk is much better even when $e_p$ is as high as $0.2 - 0.4$. Most likely this is because the smooth mass distribution of the disk reduces the amplitude of its higher-order gravitational multipoles and allows secular theory, which goes only to octupole order, to better capture the main effects of the disk gravity.

The general conclusion one can draw from the comparisons shown in Figure $\text{1.4}$ is that the secular theory for perturbations due to the disk developed in Appendix A works very well and our analytical results $\text{(1.74)-(1.75)}$ for the behavior of coefficients $\psi_1$ and $\psi_2$ are correct.
Figure 1.4: Verification of analytical calculation of the disk disturbing function \( R_d \) using numerical integrations with MERCURY. Time evolution of planetesimal eccentricity \( e_p \) (left panels) and apsidal angle \( \varpi_p \) (right panels) is shown for different disk parameters. Blue and red curves represent numerical and analytical results. In all cases planetesimals start with zero eccentricity, which explains the discontinuous jumps in \( \varpi_p \); each time the orbit passes through zero eccentricity, \( \varpi_p \) changes by \( \pi \). The disk extends from 0.1 AU to 5 AU and has \( \Sigma_p(1\text{ AU}) = 100\text{ g cm}^{-2} \). (a) Planetesimal motion is shown at \( a_p =1\text{ AU} \) for a disk with \( p = -q = 1 \) and eccentricity at its outer edge \( e_0 = 0.1 \). (b) Same as (a) but at \( a_p =2\text{ AU} \) and \( e_0 = 0.2 \). (c) Same as (b) except that here eccentricity is lowered to \( e_0 = 0.05 \). (d) Here \( a_p =2\text{ AU} \), \( e_0 = 0.1 \), and \( p = 1 \), but the disk eccentricity profile now has \( q = 0 - e_d \) is independent of distance. Apparently, in all cases the agreement between analytical secular theory and direct orbit integrations is very good. See text for more details.
1.5 Planetesimal Eccentricity Behavior

We will now consider different regimes of planetesimal dynamics. We start by using Equation (1.3) to express the disk-related precession rate $A_d$ and eccentricity excitation coefficient $B_d$ via the disk mass $M_d$:

$$A_d = (2 - p) \psi_1 n_p \left( \frac{M_d}{M_p} \left( \frac{a_p}{a_{out}} \right)^{2-p} \right),$$

$$B_d = \frac{2}{2} \psi_2 n_p \left( \frac{M_d}{M_p} \left( \frac{a_p}{a_{out}} \right)^{2-p} \right) e_d(a_p),$$

see Equations (1.5)-(1.6) and $e_d(a_p)$ is given by Equation (1.2).

These expressions show that disk-driven planetesimal eccentricity is determined, in part, by the values of power law indices $p$ and $q$. Unfortunately, these parameters are rather poorly known for real protoplanetary disks. Based on standard accretion disk theory R13 advocated the use of $p \approx 1$ for the circumstellar disks in binaries. However, this choice is subject to uncertainly in our knowledge of the radial behavior of the viscous $\alpha$-parameter, thermal structure of the disk, etc. Thus, in this work we explore a range of values of $p$.

Equally uncertain is the choice of the disk eccentricity slope $q$. If one were to neglect the self-gravity, pressure and viscous forces in the gaseous disk then its fluid elements would behave as free particles perturbed by the binary companion and have their eccentricity scaling linearly with $a_p$ (Heppenheimer 1978; also Equation [1.34]), $e_d \propto a_p$, so that $q = -1$. This behavior is at least approximately supported by the numerical results of Okazaki et al. (2002) and semi-analytical calculations of Paardekooper et al. (2008) within a range of radii. Other authors find $e_d$ to exhibit more complicated, non-power law behavior (Kley et al. 2008; Marzari et al. 2009). Despite that, in this work we will predominantly stick to using $q = -1$, but sometimes we will consider other values of $q < 0$.

All disk models considered in this paper have eccentricity $e_d$ increasing with radius, and surface density decreasing with radius. Under these natural assumptions, the disk should
dominate the motion of planetesimals close to the primary star, since \( A_b \) and \( B_b \) very rapidly grow with \( a_p \) (while \( A_d \) and \( B_d \) can even decay with \( a_p \) for certain values of \( p \) and \( q \)). Similarly, for less massive disks, in the outer parts of the disk the binary dominates both the precession and eccentricity excitation of planetesimals. Then we may ignore the disk-driven perturbations unless the binary orbit is completely circular, in which case eccentricity excitation is solely due to the gravity of elliptical disk.

Using Equations (1.9) and (1.27) we can quantify this logic by forming a ratio

\[
\frac{|A_d|}{A_b} = \frac{4|\psi_1|}{3} \frac{M_d}{\nu M_p} \left( \frac{a_b}{a_{\text{out}}} \right)^{2-p} \times \left( \frac{a_p}{a_b} \right)^{-1/(1+p)}.
\]  

(1.29)

Disk (binary) terms dominate planetesimal precession rate when \( |A_d/A_b| \lesssim 1 \) (\( |A_d/A_b| \gtrsim 1 \)), see Figure 1.5.

We do analogous calculation for eccentricity excitation using Equations (1.10) and (1.28):

\[
\frac{|B_d|}{B_b} = \frac{8|2-p|\psi_2}{15} \frac{e_0 M_d}{e_b \nu M_p} \left( \frac{a_b}{a_{\text{out}}} \right)^{2-p-q} \times \left( \frac{a_p}{a_b} \right)^{-2/(1+p+q)}.
\]  

(1.30)

where \( e_0 \) is the disk eccentricity at its outer edge. Again, disk (binary) dominates planetesimal eccentricity excitation when \( |B_d/B_b| \gtrsim 1 \) (\( |B_d/B_b| \lesssim 1 \)), see Figure 1.5.

Conditions (1.29) & (1.30) define special locations in the disk, where the ratios \( |A_d/A_b| \), \( |B_d/B_b| \) become equal to unity. We find that \( |A_d/A_b| = 1 \) at

\[
a_A = a_b \left[ \frac{4|\psi_1(2-p)|}{3} \frac{M_d}{\nu M_p} \left( \frac{a_b}{a_{\text{out}}} \right)^{2-p} \right]^{1/(1+p)}
\approx 0.16 a_b \left[ \frac{M_d/(\nu M_p)}{0.01} \frac{0.25 a_{\text{out}}}{a_b} \right]^{0.5},
\]

(1.31)
while $|B_d/B_b| = 1$ at

$$a_B = a_b \left[ \frac{8|\psi_2(2 - p)|}{15} \frac{M_d e_0}{\nu M_p e_b} \left( \frac{a_b}{a_{out}} \right)^{2-p-q} \right]^{1/(2+p+q)},$$

where numerical estimates are for a disk model with $p = 1$, $q = -1$ ($|\psi_1(1)| = 0.5$, $|\psi_2(0)| = 1.5$).

For the parameters adopted in these estimates both $a_A$ and $a_B$ lie within the disk, at separations of $2 - 3$ AU for $a_b = 20$ AU (with $a_{out} = 5$ AU), which is outside the semi-major axes of the planets in binaries detected so far. The obvious implication is that these planets have formed in the part of the disk where secular effects were dominated by the disk gravity rather than by the secondary. This suggests that disk gravity plays a decisive role in determining planetesimal dynamics in the planet-building zone.

Using ratios (1.29) & (1.30) we now describe different possible regimes of the planetesimal eccentricity behavior, as illustrated in Figure 1.5. We identify each regime using a two-letter notation in which the first letter describes what dominates planetesimal precession rate $A$, while the second refers to the dominance of eccentricity excitation (e.g. “Case DB” means that $|A_d/A_b| \gtrsim 1$ and $|B_d/B_b| \lesssim 1$). In Figure 1.6 we map out these different dynamical regimes in the space of the scaled disk mass $M_d/(\nu M_p)$ and planetesimal semi-major $a_p/a_b$ for different disk models (combinations of $p$, $q$, $e_0/e_b$).

### 1.5.1 Case DD: Disk Dominates Both Precession and Excitation

At small separations from the primary, $a_p \lesssim a_A, a_B$, the disk dominates both precession and eccentricity excitation of planetesimals, so $A \approx A_d$ and $|B_b| \ll |B_d|$. In this case the
Figure 1.5: Illustration of different regimes of planetesimal eccentricity behavior, based on Equations (1.29) and (1.30). Dynamical regimes are identified using two-letter notation as described in the text. See §§1.5.1-1.5.4 for details.

The characteristic planetesimal eccentricity (1.25) tends to

\[ e_p^{DD}(a_p) \to 2 \left| \frac{B_d}{A_d} \right| = \left| \frac{\psi_2}{\psi_1} \right| e_d(a_p). \] (1.33)

In this regime the maximum planetesimal eccentricity is of order the local disk eccentricity, since \( |\psi_{1,2}| \sim 1 \). For example, ignoring edge effects \( e_p(a_p) \to 3e_d(a_p) \) for a \( p = 1, q = -1 \) disk. Thus, an elliptical disk is capable of exciting planetesimal eccentricity of order of its own eccentricity \( e_d \) purely by its non-axisymmetric gravitational field. In this regime planetesimal
eccentricity should increase with \( a_p \) because \( e_d(a_p) \) is expected to be a growing function of \( a_p \).

Figure 1.6 demonstrates that this dynamical regime is unavoidable for \( a_p \lesssim 1 \) AU even for relatively small disk masses, down to \( M_d \sim 10^{-3}M_\odot \).

### 1.5.2 Case BB: Binary Dominates Both Precession and Excitation

In the opposite limit, far from the primary, as \( a_A, a_B \lesssim a_p \) (which is of course possible only if \( a_A, a_B \lesssim a_{\text{out}} \), planetesimal dynamics is governed completely by the binary potential. The contribution from the disk is insignificant so that both \( A \approx A_b \) and \( |B_d| \ll |B_b| \). This is the limit of planetesimal dynamics in a diskless binary, which has been investigated by Heppenheimer (1978).

In this case planetesimal eccentricity is given by

\[
e_p^{BB}(a_p) \rightarrow 2 \left| \frac{B_b}{A_b} \right| = \frac{5}{2} \frac{a_p}{a_b} e_b, \tag{1.34}
\]

in agreement with Heppenheimer (1978).

Figure 1.6 shows that Case BB is important for a broad range of separations, down to 1 AU, when the disk mass is very small, \( \lesssim 10^{-3}M_\odot \). However, for more massive disks with \( M_d \gtrsim 10^{-2}M_\odot \) this regime never emerges for \( a_p < a_{\text{out}} \). Thus, in compact binaries (\( a_b \sim 20 \) AU) with massive disks the classical result of Heppenheimer (1978) may never actually apply.

### 1.5.3 Case BD: Binary Dominates Precession, Disk Dominates Excitation

In between the two limiting cases covered in §1.5.1 and 1.5.2 there are other dynamical regimes.

Provided that \( a_A < a_B \) there exists a region in the disk with \( a_A \lesssim a_p \lesssim a_B \), where planetesimal precession is dominated by the binary companion \( (A \approx A_b) \), while eccentric-
ity excitation is determined by the disk gravity (|B_d| \gg |B_b|). In this limit planetesimal eccentricity is given by

\[ e_{pBD}(a_p) \rightarrow 2 \frac{|B_d|}{A_b} = \frac{4|\psi_2(2 - p)|}{3} e_d(a_p) \frac{M_d}{\nu M_p} \]

\[ \times \left( \frac{a_b}{a_{out}} \right)^{2-p} \left( \frac{a_p}{a_b} \right)^{-(1+p)} . \] (1.35)

Using this expression and Equations (1.29), (1.30) one can easily show that

\[ e_{pBD}(a_p) = e_{pDD}(a_p) \left( \frac{a_p}{a_A} \right)^{-(1+p)} \]

\[ = e_{BB}(a_p) \left( \frac{a_p}{a_B} \right)^{-2(1+p+q)} . \] (1.36)

Since \( a_A \lesssim a_p \lesssim a_B \) in Case BD, this result implies (for \( p > -1, p + q > -2 \)) that \( e_{pBD}(a_p) \lesssim e_{pDD}(a_p) \lesssim e_{BB}(a_p) \). It is then clear that Case BD requires \( e_{pDD}(a_p) \gtrsim e_{BB}(a_p) \) locally, i.e., according to Equation (1.33), that the disk eccentricity \( e_d(a_p) \) be higher than planetesimal eccentricity \( e_{BB}(a_p) \) in a diskless case for the values of \( p \) and \( q \) explored in this paper.

This situation may not be easy to realize in practice since pressure and viscous forces may tend to reduce (and not increase) eccentricity of fluid elements compared to that expected for test particles (i.e. \( e_{BB} \)).

Figure 1.6 shows that indeed this dynamical regime requires rather special conditions to be realized, such as the relatively high value of the disk eccentricity \( e_0/e_b \). Even then it typically occupies a narrow range of separations, see Figure 1.6a,b. This is because disk models in these two panels have \( e_d(a_p) \approx e_{BB}(a_p) \), essentially eliminating Case BD region. In Figure 1.6c we do display a model with \( e_d(a_p) \gtrsim e_{BB}(a_p) \) close to the primary (we take \( e_d \propto a_p^{1/2} \), while \( e_{BB} \propto a_p \)) so that Case BD emerges at small \( M_d \) and relatively small \( a_p \). However, as we mentioned before, this may not be a typical situation.
Figure 1.6: Map of different dynamical regimes in the space of planetesimal semi-major axis $a_p$ and disk mass $M_d$. Different panels correspond to different disk models, ones on the right have disk eccentricity (indicated on panels together with $p$ and $q$) 10 times lower than the left ones. Disk extends from 0.1 AU to 5 AU, binary semi-major axis 20 AU, eccentricity 0.2, and secondary to primary mass ratio $\nu = 1/2$. Dynamical regimes in each part of the phase space are indicated. Dotted and dashed lines are given by Equations (1.31) and (1.32). One can see that planetesimals are in DD regime in massive disks near the primary, and in BB regime in low-mass disks far from it. Edge effects are ignored in this calculation and we use the values of $\psi_1$ and $\psi_2$ that they take at 1 AU. See text for details.
1.5.4 Case DB: Disk Dominates Precession, Binary Dominates Excitation

Now we look at the opposite case of $a_B < a_A$, which emerges when $e_0/e_b$ is low. Within the range $a_B \lesssim a_p \lesssim a_A$ planetesimal precession is dominated by the disk gravity ($A \approx A_d$), while eccentricity excitation is determined predominantly by the secondary star ($|B_d| \ll |B_b|$). This is the approximation of a massive axisymmetric disk discussed in R13. In agreement with that work we find the maximum eccentricity to follow

$$e_{DB}^p(a_p) \to 2 \left| \frac{B_b}{A_d} \right| = \frac{15}{8|\psi_1(2-p)|} \nu \frac{M_p}{M_d} e_b \left( \frac{a_{out}}{a_b} \right)^{2-p} \left( \frac{a_p}{a_b} \right)^{2+p} .$$

This expression and Equations (1.29), (1.30) imply that

$$e_{DB}^p(a_p) = e_{DD}^p(a_p) \left( \frac{a_p}{a_B} \right)^{2+p+q} = e_{BB}^p(a_p) \left( \frac{a_p}{a_A} \right)^{1+p} .$$

Because now $a_B \lesssim a_p \lesssim a_A$ we see that $e_{DB}^p(a_p) \lesssim e_{DD}^p(a_p) \lesssim e_{BB}^p(a_p)$. Then it follows that DB regime requires $e_d(a_p) \sim e_{DD}^p(a_p) \lesssim e_{BB}^p(a_p)$ in non-pathological cases.

According to Figure 1.6 this dynamical regime is rather common at low $e_0/e_b$, but is difficult to realize inside the disk for higher $e_0/e_b$. For some models (e.g. see Figure 1.6d,e) case DB regime holds within an extended region of the disk.

1.6 Eccentricity Profiles

To illustrate the results of the previous section, in Figure 1.7 we show profiles of planetesimal eccentricity computed for different disk models. For reference, each of the panels displays planetesimal eccentricity for the diskless case ($e_{BB}^p(a_p)$, dark blue, big-dotted) as well as $e_p$
Figure 1.7: Plots of planetesimal eccentricity as a function of $a_p$ for different disk models (values of $p$, $q$, and disk eccentricity at the outer disk edge $e_0$ are shown in panels). For reference the dark blue big-dotted line shows eccentricity in the case of no disk (Equation (1.34)), the red dot-dashed line shows the case of no secondary (disk only, Equation (1.33)), the green dashed line shows the critical eccentricity at fragmentation threshold (Equation (1.41)). Other curves show $e_p$ for a binary ($a_b = 20$ AU, $e_b = 0.2$, $M_p = M_\odot$, and $\nu = 1/2$) with the disk extending from 0.1 AU to 5 AU and having different mass as shown in panels. Note a conspicuous secular resonance around 1.5 AU in models with the low-mass disk. At small separations ($\lesssim 1$ AU) curves of $e_p(a_p)$ converge towards the disk-dominated solution, Equation (1.33). There are deviations of $e_p$ from simple power-law behavior at the inner and outer edges of the disk due to the nontrivial behavior of $\psi_1$ and $\psi_2$ there. See text for more details.
for the case with no secondary \( e_{dd}(a_p) \), red dot-dashed). In the left panels we have chosen disk eccentricity \( e_d(a_p) \) very close to the eccentricity of a free particle in the binary potential, which explains why the curves of \( e_{bb}^{bb}(a_p) \) and \( e_{dd}^{dd}(a_p) \) almost overlap. In the right panels \( e_d \) is reduced by an order of magnitude and the two curves are well separated.

Note that the \( e_{dd}^{dd}(a_p) \) curve does not follow the simple power law in \( a_p \) as one would have expected based on Equation (1.33) and the assumption of \( \psi_1, \psi_2 \) being constant — it clearly deviates from this simple form at the disk edges. This is because near the disk edge, boundary terms neglected in computing the Figure 1.3 start to affect the values of \( \psi_1 \) and \( \psi_2 \) in a non-trivial manner, see Figure 1.10.

We plot eccentricity profiles for different values of the disk mass. At the lowest disk mass, \( M_d = 10^{-3} M_p \), planetesimal eccentricity \( e_p \) starts out very high in the outer disk (in the BB regime, see Figure 1.6), above \( e_{bb}^{bb}(a_p) \). A notable feature of this profile is the secular resonance located at \( \approx 1.5 \) AU and causing \( e_p \) to diverge. Its existence was predicted in R13 and Rafikov (2013a) for the case of circumprimary and circumbinary disks correspondingly. Later Meschiari (2014) confirmed the emergence of this resonance in massive circumbinary disks using numerical simulations of planetesimal dynamics.

The origin of this resonance lies in the fact that \( A_b \) is always positive, whereas for the disks that we are considering, \( A_d \) is negative, see Figure 1.3. This means that at \( a_A \) (see Equation (1.31)), where \( |A_d| = |A_b| \) one actually has \( A = 0 \) and our secular solution (1.3) diverges. Inward of the resonance, \( e_p \) rapidly goes down (in DB and DD regimes) and asymptotically approaches \( e_{dd}^{dd}(a_p) \) for \( a_p \lesssim 0.5 \) AU.

For a somewhat more massive disk \( M_d = 10^{-2} M_p \) the \( e_p \) profile looks very different — it does not exhibit secular resonance (since \( a_A \) is now outside the outer disk edge \( a_{out} \)) and generally features lower values of \( e_p \). This happens because with such a massive disk, planetesimal excitation is never in the BB regime. Disk gravity governs particle dynamics essentially through the whole disk.
This is even more so for the $M_d = 10^{-1} M_p$ disk. At this high mass planetesimal eccentricity curves closely follow $e_p^{DD}(a_p)$ for all $a_p$. As a result, in low-$e_0$ disks $e_p$ can be appreciably lower than what it is if planetesimals are affected by the gravity of the binary companion alone, similar to the case studied in R13. Somewhat counterintuitively, adding an additional perturber — a massive disk — to the system does not heat it up dynamically but in fact reduces planetesimal random velocities.

In all cases we see that $e_p$ is above the smaller of the $e_p^{BB}(a_p)$ and $e_p^{DD}(a_p)$. Thus, nonzero disk eccentricity introduces a lower limit on the $e_p$ value.

### 1.7 Dynamics in the Case of a Precessing Disk

So far we have been dealing with the case of non-precessing disk which keeps its orientation fixed in the frame of the binary orbit. However, simulations often find that gas disks in binaries not only develop a non-zero eccentricity but also precess (Okazaki et al. 2002; Paardekooper et al. 2008; Marzari et al. 2009). Thus it is important to discuss how planetesimal dynamics changes in the case of a precessing disk.

In Appendix C we present the extension of our solutions for the planetesimal eccentricity in §1.4.4 to the case of a disk that precesses as a solid body at a constant rate $\dot{\varpi}_d$. We find that the eccentricity vector can again be separated into three distinct contributions, see Equation (1.83): (1) standard forced eccentricity vector due to binary with amplitude $|e_{\text{forced},b}| = |B_b/A|$, stationary in the binary frame, (2) forced eccentricity vector due to the disk with amplitude $|e_{\text{forced},d}| = |B_d/(A - \dot{\varpi}_d)|$, rotating at the rate $\dot{\varpi}_d$, and (3) the free eccentricity term with amplitude

$$
|e_{\text{free}}| = \frac{1}{|A(A - \dot{\varpi}_d)|} \left[ (AB_d)^2 + (B_b(A - \dot{\varpi}_d))^2 + 2AB_dB_b(A - \dot{\varpi}_d) \cos \varpi_{d0} \right]^{1/2}
$$

\[1.39\]

1\text{Note that } \dot{\varpi}_d\text{ has a meaning different from that in R13, where } \dot{\varpi}_d\text{ was equivalent to } A_d\text{ in our current notation.}
rotating at the precession rate $A$ (here $\varpi_d(0)$ is the value of $\varpi_d$ at $t = 0$).

The expression for the characteristic eccentricity becomes more complicated and depends on the value of $\varpi_d(0)$. The maximum possible eccentricity (for planetesimals starting with $h_p(0) = k_p(0) = 0$) is reached when $\varpi_d(0) = \varpi_d(0) = 0$ or $\pi$ (disk and binary periapses aligned or anti-aligned initially), depending on the signs of $A$, $B_d$, and $A - \dot{\varpi}_d$. Then the maximum eccentricity is given by

$$e_{\text{char}} = 2 \left( \left| \frac{B_b}{A} \right| + \left| \frac{B_d}{A - \dot{\varpi}_d} \right| \right).$$ (1.40)

Comparing this expression with Equation (1.25) we conclude that disk precession does not affect planetesimal eccentricity behavior as long as $|\dot{\varpi}_d| \ll |A|$.

However, in the opposite case of $|\dot{\varpi}_d| \gtrsim |A|$ the disk-driven forced part of the eccentricity vector is suppressed compared to the case of no precession. This is because rapid precession of the disk (compared to the rate of planetesimal orbital precession) effectively averages out the non-axisymmetric part of the disk potential, considerably reducing related eccentricity excitation. This has implications discussed in §1.8.2. At the same time the forced eccentricity contribution due to binary stays unchanged for planetesimals embedded in the precessing disk. We expect these asymptotic results to remain valid even in the case of non-uniform disk precession, both when it is much faster and much slower than $|A|$. However, all this discussion strictly applies only in the absence of gas drag.

1.8 Discussion

We can put our findings in the context of existing results on the purely gravitational dynamics (i.e. not accounting for gas drag) of planetesimals in binaries. Heppenheimer (1978) explored planetesimal dynamics under the gravity of the companion alone. Our results reduce to his in the limit of a zero-mass disk, i.e. when planetesimal dynamics is in the BB regime, see §1.5.2.
It was first shown analytically in R13 that the gravity of a massive disk can significantly suppress planetesimal eccentricity excitation in binaries. The reason lies in the fast precession of planetesimal orbits caused by the disk gravity, which effectively averages out $e_p$ forcing by the companion. This effect is present in our calculations as well and we reproduce the results of R13 in Case DB.

However, our work includes another important ingredient not considered previously in the framework of secular theory — gravitational forcing of planetesimal eccentricity by the disk itself, which should be present in addition to planetesimal precession if the disk is eccentric. While some numerical studies on this topic do exist (see §1.8.3) analytical understanding of their results has been hampered by the complexity of the problem.

In this work we have provided the first (to the best of our knowledge) calculation of the eccentric disk potential in application to planetesimal dynamics. Using this prescription we uncovered the existence of two entirely new regimes of planetesimal dynamics — Case BD (§1.5.3) and Case DD (§1.5.1) — in which eccentricity excitation by the disk exceeds that due to the secondary. The latter regime (DD) represents a very common situation in protoplanetary disks in binaries. As we have shown in §1.5 in many cases planetesimal excitation is in the DD regime throughout the whole disk.

Significance of this dynamical regime also lies in the fact that the disk drives planetesimal eccentricities to a value of order the local disk eccentricity, see Equation (1.33). Even though in the absence of any damping agents eccentricity of a particle starting on a circular orbit oscillates, see Equation (1.24), so that during some periods $e_p \ll e_{\text{char}}$, most of the time $e_p$ is of order $e_{\text{char}} \sim e_d$ in the regime DD. Thus, eccentricity of the disk gives rise to a lower limit on the characteristic planetesimal eccentricity (1.25), which is a very important finding.

In particular, it constrains the applicability of the axisymmetric disk approximation used in R13. Indeed, let us calculate $e_p$ using Equation (1.37), which is identical to the result of R13, for a system with $M_p = M_\odot$, $\nu = 0.3$, $e_b = 0.4$, $a_b = 20$ AU harboring an axisymmetric disk with $a_{\text{out}} = 5$ AU, $M_d = 10^{-2}M_\odot$, $p = 1$. At $a_p = 2$ AU we find $e_p \approx 10^{-2}$, which is
much less that it would be in a diskless case, $e_p^{BB} \approx 0.1$, see Equation (1.34). However, for this result to hold in a non-axisymmetric disk the disk eccentricity at 2 AU has to be less than $10^{-2}$. Whether such low $e_d$ is realistic is not clear at the moment (see §1.8.3).

Our current results have been derived assuming that the disk affects planetesimals only via its gravitational field. In practice planetesimals are also subject to gas drag, which has important consequences for their dynamics. First, gas drag lowers planetesimal velocities with respect to gas, which also lowers relative planetesimal velocities therefore positively affecting survival in mutual collisions. Second, it has long been known that gas drag introduces apsidal alignment of planetesimal orbits (Marzari & Scholl 2000), which considerably reduces relative collision velocities of near-equal bodies. However, planetesimals of different sizes would still collide at high speeds suppressing growth (Thébault et al. 2006, 2008). Third, gas drag damps the free part of eccentricity, see Beaugé et al. (2010). This should affect the time dependence of planetesimal eccentricity, which in our case is given by Equation (1.24). We address the effects of gas drag on planetesimal dynamics in binaries in Rafikov & Silsbee (2015a).

Because of the neglect of gas drag our current results are strictly valid only for relatively large objects, with sizes of several hundred km. For such planetesimals gas drag can be unimportant compared to purely gravitational forces during rather long time span, and may thus be neglected. Inclusion of gas drag does not negate our finding that disk gravity from an eccentric disk leads to high encounter velocities between planetesimals, even of kilometer size. Our results also clearly show that purely gravitational effects alone, in the absence of dissipative forces, can give rise to non-trivial behavior of $e_p$ (see e.g. §1.5.1, 1.5.3) not captured in previous analyses of the problem.

We also note that our assumed surface density profile (1.1)-(1.2) may not fully capture the distribution of $\Sigma$ in real disks. First, pressure forces drive differential precession in a hydrodynamical disk, which can be avoided only under rather special circumstances (Statler 2001). Second, these equations in their current form do not capture the possible presence of
the density waves in the disk driven by the companion perturbation. They can be accounted for by assuming the apsidal angle $\varpi_d$ of the fluid trajectories to vary with the distance in a particular fashion. For simplicity we did not consider such possibility in this work.

However, even if the expressions (1.1)-(1.2) are only approximate, this does not change our main conclusions about the key role of the disk gravity. Indeed, we find the values of $A_d$ and $B_d/e_g$, which determine the disk effect on planetesimal dynamics, to not depend sensitively on the power-law indices $p$ and $q$ over a range of reasonable values, see Figure 13. Thus, we do not expect our results to change dramatically if the behavior of $\Sigma(a_p)$ and $e_d(a_p)$ were to deviate from the pure power laws in $a_p$.

Finally, short-term variability of the disk surface density can induce fast-changing torques on planetesimals. These effects cannot be captured by our secular (time-averaged) approach. However, we do not expect them to act coherently on long timescales and therefore to be subdominant for the same reason that the short-period terms of the planetary perturbations play an insignificant role on long time intervals in classical celestial mechanics (Murray & Dermott 1999).

### 1.8.1 Implications for Planetesimal Growth

Planetesimal growth requires relative velocities of colliding bodies to be small, otherwise they get eroded or destroyed. We use our results to provide some insights on planetesimal accretion in binaries.

For bodies held together primarily by gravity the threshold collision velocity at which planetesimals can still survive is about the escape speed. Guided by this logic Moriwaki & Nakagawa (2004) and R13 use the simple criterion

$$e > e_{\text{crit}} = \frac{2v_{\text{esc}}}{v_k} \approx 6.1 \times 10^{-4} \frac{d}{10 \text{ km}} \times \left( \frac{M_\odot}{M_p} \frac{a_p}{1 \text{ AU}} \frac{\rho}{3 \text{ g cm}^{-3}} \right)^{1/2}$$

(1.41)
as the condition for planetesimal destruction in collisions. Here $v_{\text{esc}}$ is the escape speed from a planetesimal of a given radius $d$ and $\rho$ is the bulk density of planetesimal material.

In Figure 1.7 we display $e_{\text{crit}}(a_p)$ by the dashed line and compare it with the characteristic $e_p$ attained by planetesimals as a result of disk+secondary gravitational perturbation for different disk models. One can see that in all models where disk eccentricity $e_d$ is high, comparable to the free-particle diskless eccentricity $e_p^{BB}$ (Figure 1.7a,b), the disk does not help eliminate the fragmentation barrier. This is because $e_p$ cannot drop below $e_d$ and $e_d$ is high. The situation is clearly more helpful for planetesimal growth in lower-$e_d$ cases, see Figure 1.7c,d, even though it is still not as easy as in the case of axisymmetric disk studied in R13. On the other hand, it has been noted in Rafikov (2013b) that the catastrophic destruction condition (1.41) is likely too conservative and underestimates the ability of planetesimals to survive in mutual collisions. This issue is addressed in more detail in Rafikov & Silsbee (2015b).

Another potential problem that may arise in low-mass disks with $M_d \sim 10^{-3} M_\odot$ is the presence of secular resonance in the disk, see §1.6 and Figure 1.7. There $e_p$ becomes very large in a narrow range of $a_p$, making planetesimal collisions highly destructive. This phenomenon is non-local since high-$e_p$ objects can penetrate other disk regions and destroy planetesimals there as well.

However, this problem is unlikely to last for a long time as the small number of planetesimals from the vicinity of the secular resonance will be rapidly destroyed in collisions, leaving no more projectiles to destroy the remaining planetesimals in the rest of the disk. Also, our inference of high-$e_p$ at secular resonance is based on linear secular Equations (1.18)-(1.17), which were derived under the assumption of $e_p \ll 1$, clearly not fulfilled at the resonance. The actual $e_p$ in this part of the disk will be different from our predictions.
1.8.2 Lowering Planetesimal Excitation

Motivated by our results and their implications for planetesimal accretion we next discuss different scenarios (in order of their likely significance) in which relative velocities of planetesimals affected only by the gravity of gaseous disk and binary companion can be considerably lowered.

**Intrinsically Low $e_d$**

The major obstacle for planetesimal growth in high-$e_d$ disks has to do with our general result (§1.6) that $e_p$ is always above the smaller of $e_p^{BB}(a_p)$ and $e_p^{DD}(a_p) \sim e_d$. Thus, one of the most straightforward ways of lowering collision speeds is for the disk to have low $e_d$ either locally or globally for a long period of time. Our current understanding of eccentricity excitation in gaseous disks is based primarily on the results of numerical simulations, which are reviewed in §1.8.3. We describe possible ways of lowering $e_d$ there.

**Rapidly Precessing Disk**

When discussing the possibility of disk precession in §1.7 we noted that the disk-induced contribution to the forced eccentricity can be effectively suppressed if the disk precesses faster than the planetesimals, i.e. if $|\dot{\omega}_d| \gg |A|$. In this case $e_p$ can easily be below $e_p^{DD} \sim e_d$. The remaining excitation due to the binary will keep $e_p$ at the level of $e_p^{DB}$, which is low because of the fast planetesimal precession driven by the massive disk, see §1.5.4. Thus, fast disk precession effectively brings planetesimal dynamics to the situation described in R13 and can serve as a mechanism for lowering planetesimal excitation, as long as gas drag can be neglected (Rafikov & Silsbee 2015a). Whether the gaseous disk can precess at the rate exceeding $|A|$ at separations of several AU, where the giant planets are detected in close binaries, should thus be explored in more detail.
Figure 1.8: Illustration of planetesimal eccentricity behavior for a particular disk model with 
$p = 0.5, q = -1.5, e_0 = 0.05$, extending from 0.1 to 5 AU, in a binary with $a_b = 20$ AU,
$e_b = 0.2, M_p = M_\odot$, and $\nu = 1/2$. The meaning of the different curves is the same as
in Figure 1.7, with the addition of the black line corresponding to $e_d$. Because this model
has $p + q = -1$, the non-axisymmetric part of the disturbing function vanishes (if one
neglects edge effects) and $e_p$ is generally quite low, lower than the disk eccentricity $e_d$, and
compatible with planetesimal growth (for $\gtrsim 10$ km bodies) for $a_p \lesssim 1$ AU. We have shown
the disk eccentricity $e_d$ (black solid line) to illustrate that the planetesimal eccentricity $e_p$ is
much lower than $e_d$ in the inner disk (this is unlike the case of a disk with $p + q \neq -1$).
Globally Suppressed Eccentricity Excitation

Another way of making \( e_p^{\text{DD}} \) low is hinted to us by Equation (1.33), which shows that \( e_p^{\text{DD}} \) can be low even for high \( e_d \) if \( \psi_2 \) is very small. The same is true for \( e_p^{\text{BD}} \), see Equation (1.35). This is because the disk then produces zero contribution to the non-axisymmetric component of the disturbing function. Figure 1.3 shows that, ignoring the possible edge effects, this is possible e.g. if \( p + q = -1 \). Our fiducial disk model based on general ideas about the accretion disk physics and their eccentricity excitation has \( p = 1 \) and \( q = -1 \), which is not compatible with this condition. However, our present understanding of protoplanetary disks in binaries does not allow us to exclude disk models with \( p + q = -1 \). Interestingly, when \( p = 0 \) then the disk also has zero contribution to the axisymmetric component; a disk with \( p = 0 \) (i.e. uniform disk), \( q = -1 \) affects neither planetesimal precession nor eccentricity excitation by its gravity in the absence of edge effects.

In Figure 1.8 we show an example of one such model having \( p = 0.5 \) and \( q = -1.5 \), with rather high disk eccentricity at the outer disk edge, \( e_0 = 0.25 e_b = 0.05 \) (for \( e_b = 0.2 \)). One can clearly see that in this case \( e_p^{\text{DD}} \) is low and comparable to \( e_{\text{crit}} \) at \( \sim \text{AU} \) separations. This should facilitate planetesimal growth on these scales. The disk-induced excitation for this model is not exactly zero due to edge effects. This is more of an issue near the outer edge of the disk because as shown in Appendix A (and illustrated in Figure 1.2) \( p + q = -1 \) is closer to the line of convergence at the outer edge than at the inner edge\(^2\). This means that edge effects are more important in the outer disk for this set of power law indices.

It is also worth noting that some (though not all) of the lower planetesimal eccentricity in this Figure as compared to Figure 1.7 is due to lower assumed disk eccentricity \( e_d \) in the inner part of the disk. However, the drop in \( e_p \) as one moves away from the inner edge of the disk reflects the drop in the non-axisymmetric part of the disk disturbing function as the

\[^2\]The asymptotic behavior of Equation (1.75) shows that for \( p + q = -1 \), \( \psi_2 \propto \alpha_2 \) as \( \alpha_2 \to 0 \) and \( \psi_2 \propto \alpha_1^6 \) as \( \alpha_1 \to 0 \).
inner edge effect becomes less important and we see that $e_p$ drops well below the local value of $e_d$.

**Locally Suppressed Eccentricity Excitation**

Additionally, there are at least two ways in which $e_p$ can be reduced *locally*, within a narrow range of semi-major axes. First, even if the disk does not have $p + q = -1$ *globally*, as we assumed in making Figure 1.8, there could be parts of the disk in which this condition is fulfilled for a range of $a_p$, for example near the disk edges, where $\Sigma_p$ should be petering out to zero, or near dead zones or opacity transitions, where the material pileup is possible and a non-power law scaling of $\Sigma_p$ is likely. Our results do not directly apply to such situations since we assumed a purely power law behavior of $\Sigma_p(a_p)$ but based on them we can expect that it might be possible to have $\psi_2$ close to zero at radii, near which locally computed $p + q = -\partial \ln (e_d \Sigma) / \partial \ln a_p$ passes through $-1$ (edge effects mentioned in §1.8.2 may make situation even more complicated). At this location contributions of the inner and outer disks to $\psi_2$ should nearly cancel each other resulting in low $e_p^{\text{DD}}$. Of course, $e_p$ is lowered in this way only if the disk dominates eccentricity excitation, i.e. in the Case DD.

**Favorable Disk-Binary Orientation**

Second, so far we have always assumed planetesimal eccentricity to be given by the characteristic value $e_{\text{char}}$ defined by the Equation (1.25). This approach ignores the dependence of the actual maximum planetesimal eccentricity upon the relative disk-binary orientation, obvious from Equation (1.24). In particular, in §1.4.4 we noted that whenever the conditions (1.26) are fulfilled, the maximum eccentricity is much lower than $e_{\text{char}}$. Because of the different dependence of $B_d$ and $B_b$ on $a_p$ the first condition can be fulfilled only locally, within a narrow range of radii around $a_p = a_B$ given by Equation (1.32). Since $a_B$ lies within the disk only for relatively small $M_d$ (see Equation 1.32), we conclude that the first condition is fulfilled only for relatively light disks, $M_d \lesssim 10^{-2}M_\odot$. 

36
For most disk models considered in this work one finds ψ_2 > 0 (see Figure 1.3) and \( B_d > 0 \) (Equation 1.6), while \( B_b < 0 \) (Equation 1.10). Then the second condition in (1.26) implies \( \varpi_d \approx 0 \), i.e. that the binary and the disk apsidal lines need to be aligned for \( e_p \) to be suppressed at \( a_B \). For the more atypical cases with \( \psi_2 < 0 \) one finds that the disk-secondary anti-alignment (\( \varpi_d \approx \pi \)) is necessary to suppress \( e_p \) at \( a_B \).

The actual value of \( \varpi_d \) for disks inside binaries is not well understood and Okazaki et al. (2002) find numerically that both alignment and anti-alignment are possible for the disks stationary in the binary frame. Needless to say, if the disk is precessing, it is no longer possible for it to be aligned or anti-aligned with the binary companion for a long time and the conditions (1.26) are no longer relevant.

### 1.8.3 Comparison with Numerical Studies

There exist a number of numerical studies of planetesimal dynamics in binaries which treat structure of the gaseous disk by solving equations of hydrodynamics. However, with the exception of Kley & Nelson (2007) and Fragner et al. (2011), most of them account only for the effects of gas drag on planetesimal motion and neglect disk gravity.

The issue of the eccentricity that a gaseous disk develops under the perturbations by the companion has not been settled. Different numerical studies arrive at different conclusions, depending on the physics included in simulations and the numerical methods used. Some simulations find very high values of \( e_d \), of order 0.5 at the outer disk edge, that develop if the disk is very extended allowing the operation of an instability related to the 3 : 1 resonance studied by Lubow (1991). This mechanism of eccentricity excitation operates even if the companion is on circular orbit. Such a situation is unlikely to apply to the known binary systems, which have relatively massive (\( \nu \sim 0.4 \)) eccentric companions. Circumstellar disks in such systems should be truncated at rather small sizes, excluding the possibility of this instability.
In their SPH study of decretion disks in eccentric Be/X-ray binaries Okazaki et al. (2002) find $e_d \lesssim 0.1$ but the exact value and overall disk behavior (e.g. whether the disk is precessing) strongly depend on the resolution used. Paardekooper et al. (2008) employed a grid-based numerical scheme to simulate a circumstellar disk extending to $0.4a_b$ in a binary with the parameters of the $\gamma$ Cephei system. They find that the value of disk eccentricity very strongly depends on the details of the numerical scheme used, with $e_d(2\text{ AU})$ ranging from 0.2 to less than $10^{-2}$. Needless to say this difference should result in very different conclusions regarding the behavior of planetesimals.

Note that we use $a_{out} = 0.25a_b$ in this work, which is smaller that $a_{out}$ used by Paardekooper et al. (2008). A more compact disk is less affected by the binary and might develop smaller $e_d$. At the moment this is just a speculation since the exact value of $a_{out}$ should depend on a number of details such as disk viscosity, binary eccentricity, and so on, see Regály et al. (2011).

Marzari et al. (2009) find that disk eccentricity is lower when the self-gravity of the disk is properly incorporated in simulations. The same result — reduction of $e_d$ due to disk self-gravity — can be seen in circumbinary disks by comparing the study of Pelupessy & Portegies Zwart (2013), which includes disk self-gravity and finds a regular pattern of low $e_d$, and Marzari et al. (2013), which neglects disk self-gravity and finds very high disk eccentricity.

This observation is very relevant for our study since we find that massive disks give rise to lower planetesimal eccentricities if disk eccentricity $e_d$ can be reduced below the free-particle eccentricity $e_p^{BB}$, see §1.5. Lowering $e_d$ by the disk self-gravity would make massive disks even more attractive sites for planetesimal growth. Thus, in line with R13 we suggest that efficiency of planet formation may be a very strong function of the disk mass such that planets form only in binaries with massive disks. Although such systems are rare (Harris et al. 2012) there may be enough of them to explain a handful of known planet-hosting compact binaries.
1.9 Summary

In this work we explored secular dynamics of planetesimals embedded in an eccentric gaseous disk, with implications for planet formation in binaries. We derived, for the first time, the analytical expression for the disturbing function of a body subject to gravity of a massive, eccentric, confocal and coplanar disk, in the limit when both the disk and planetesimal eccentricities are small (Appendix A). This expression has been used in §1.4.4 to understand secular excitation of $e_p$ in presence of both the non-axisymmetric disk and the binary companion. Assuming initially circular orbits and neglecting any dissipation (such as due to gas drag) in this work, we found the general analytical solution for the evolution of planetesimal eccentricity — Equation (1.24) — which shows that $e_p$ oscillates from zero up to some maximum value.

Both the period and amplitude of oscillations depend on properties of the disk and the secondary. Depending on which agent — disk or secondary — dominates planetesimal precession and eccentricity excitation, we find four distinct regimes for the $e_p$ behavior. Two of them, in which gravity of eccentric disk dominates planetesimal eccentricity excitation, are novel results of this work. We have shown, in particular, that when the disk dominates both planetesimal precession and eccentricity excitation (so called Case DD, see §1.5.1) characteristic planetesimal eccentricity $e_p$ is of order the local disk eccentricity $e_d$. Thus, the value of $e_d$ sets a lower limit on $e_p$ and essentially determines the characteristic collision speeds of planetesimals. As a result, we generally find that eccentricity of the disk presents a serious obstacle for the growth of planetesimals with sizes of less than several tens of km.

We then discuss possible ways of lowering $e_p$, which would be favorable for planetesimal growth (§1.8.2). One of them is for the disk to be massive, typically $\gtrsim 10^{-2} M_\odot$, so that (1) its own self-gravity reduces disk eccentricity $e_d$ as has been suggested by some simulations and (2) disk gravity dominates planetesimal dynamics. Another possibility is for the disk to precess much faster than the precession rate of planetesimal orbits (§1.7). Some other ways of lowering $e_p$, both global and local (within a finite range of separations) are also described.
These possibilities may represent pathways to planetesimal growth in at least a subset of protoplanetary disks in binary systems.

Despite the neglect of dissipative effects such as gas drag (accounted for in Rafikov & Silsbee 2015a,b) the present study demonstrates the variety of planetesimal dynamical behaviors driven by the coupled gravitational perturbations of an eccentric disk and the binary. It thus represents an important step in building a complete picture of planetesimal dynamics in binaries.

The analytical description of the gravitational effects of the eccentric disk derived in this work (Appendix A) can be applied to a variety of other astrophysical problems: planetesimal dynamics in circumbinary disks (Silsbee & Rafikov 2015), dynamics of self-gravitating gaseous and stellar disks, and so on.

1.10 Appendix A: Disturbing Function Due to an Eccentric Disk

Here we present a calculation of the disturbing function due to an eccentric disk. We assume that the disk eccentricity and surface density are given by the power law ansatz (1.2) and apsidal angle is constant with radius. The latter assumption can be easily relaxed and analytical results obtained for $\varpi_d$ varying as a power law of the semi-major axis of a fluid element.

There are different ways in which such calculation can be approached. In particular, one can use the analogy with the Gauss averaging method (Murray & Dermott 1999), which treats the time-averaged potential of a point mass on an eccentric orbit as that produced by an elliptical wire along the orbit with the line density proportional to the time the planet spends at each point of its orbit. In the case of a gaseous disk, we can consider fluid in a narrow elliptical annulus between the two adjacent fluid trajectories. Because of the continuity equation the line density of this fluid along the annulus is also proportional to
the time fluid spends at a given location. Given that the density distributions are the same in two cases one can simply employ the expression for the disturbing function given by the Gauss method. For example, the secular contribution due to the outer disk becomes

\[
P_{\text{Gauss}}^{\text{out}} = n_p^2 a_p^2 \int_{a_p}^{a_{\text{out}}} \frac{2\pi a \Sigma_p(a)}{M_p} \left[ \frac{\alpha^2}{8} b_{3/2}^{(1)}(\alpha) e_p^2 - \frac{\alpha^2}{4} b_{3/2}^{(2)}(\alpha) e_p e_d(a) \cos (\varpi_p - \varpi_d) \right] da, \tag{1.42}
\]

where \( \alpha = a_p/a \). Similar expression can be written for the inner part of the disk as well. However, both \( \int_1^b b_{3/2}^{(1)}(\alpha) d\alpha \) and \( \int_1^b b_{3/2}^{(2)}(\alpha) d\alpha \) are non-convergent, as well as the sum of the inner and outer disk contributions in the vicinity of planetesimal orbit. This is a well-known problem with the Gauss expression for the secular disturbing function (Murray & Dermott 1999). For this reason we are unable to use Gauss’ method to calculate the disturbing function due to an eccentric disk for a planetesimal which is embedded in a disk with no gap.

Instead, we have resorted to a different approach previously used by Heppenheimer (1980) and Ward (1981) to compute the gravitational field of an axisymmetric disk with power law surface density profile. To use this approach for an elliptical disk we had to come up with a number of important modifications. The idea behind this method is to compute the disturbing function directly as

\[
R(S) = G \left\langle \int_S \frac{\Sigma(r_d, \phi_d) r_d dr_d d\phi_d}{\left(r_p^2 + r_d^2 - 2r_p r_d \cos \theta\right)^{1/2}} \right\rangle, \tag{1.43}
\]

where the integral is taken over the area of the disk \( S \), angle brackets \( \langle ... \rangle \) represent time averaging over planetesimal orbital motion, \( r_p \) is the (time-dependent) instantaneous radius of a planetesimal, \( \theta \) is the angle between vectors \( r_d \) and \( r_p \), see Figure 1.1. According to this figure \( \phi_d \) is the polar angle counted from the disk periastron, \( \phi_p \) is the angle of the planetesimal with respect to the planetesimal periastron, so that \( \theta = \phi_d + \varpi_d - \phi_p - \varpi_p \).

We divide the disk up into three regions as shown in Figure 1.9, so that \( S = S_c + S_0 - S_i \). Here \( S_c \) is the annulus bounded by circles with radii equal to the periastron of the outer
disk edge \( a_{\text{out}}[1 - e_d(a_{\text{out}})] \) and the periastron of the inner disk edge \( a_{\text{in}}[1 - e_d(a_{\text{in}})] \); \( S_o \) is the outer crescent region bounded the outer circle of \( S_c \) on the inside and the outermost elliptical trajectory on the outside; \( S_i \) is the inner crescent region bounded the inner circle of \( S_c \) on the outside and the innermost elliptical trajectory on the inside. The full disturbing function of an eccentric disk is given by

\[
R(S) = R(S_c) + R(S_o) - R(S_i)
\]  
\[ (1.44) \]

We now separately calculate the contributions due to different regions using an extension of the method employed by Heppenheimer (1980).

### 1.10.1 Contribution from the Annular Region \( S_c \)

We start by calculating the contribution from the annular region \( S_c \). In the following we define for brevity \( a_{\text{in}} = a_1 \), \( a_{\text{out}} = a_2 \), \( e_d(a_{\text{in}}) = e_1 \), \( e_d(a_{\text{out}}) = e_2 \), with \( r_{d,\text{in}} = a_1(1 - e_1) \) and \( r_{d,\text{out}} = a_2(1 - e_2) \) being the inner and outer radii of \( S_c \). We can write

\[
R(S_c) = \left\langle G \int_{r_{d,\text{in}}}^{r_{d,\text{out}}} \int_0^{2\pi} \frac{\Sigma(r_d, \phi_d)}{\sqrt{r_p^2 + r_d^2 - 2r_p r_d \cos \theta}} d\phi d \theta \right\rangle,
\]  
\[ (1.45) \]

where we have used the fact that \( d\phi_d = d\theta \).

As given in Statler (2001) and using Equation \[ (1.1) \], to first order in \( e_d \),

\[
\Sigma(r_d, \phi_d) = \Sigma_p(r_d) + e_d \left[ \zeta(r_d) \Sigma_p(r_d) (\cos \phi_d - 1) + r_d \cos \phi_d \frac{d\Sigma_p(r_d)}{dr_d} \right],
\]  
\[ (1.46) \]

where \( \zeta(r_d) \) was defined after Equation \[ (1.1) \]. Note that \( \Sigma_p \) is considered to be a function of the semi-major axis of a fluid element passing through a given point in the disk, as stated after Equation \[ (1.1) \]. For that reason \( \Sigma(r_d, 0) \neq \Sigma_p(r_d) \) but \( \Sigma(r_d, 0) = \Sigma_p(r_d/(1 - e_d)) \) (to
second order in $e_d$), i.e. at the semi-major axis $r_d/(1 - e_d)$ for which $r_d$ is the periastron distance.

Classical secular theory neglects terms in the disturbing function which are higher order than $e_p^2$ in planetesimal eccentricity, see \[1.4.1\]. Thus, in our subsequent calculations we will retain only terms proportional to $e_p^2$ and $e_p e_d$; terms of higher order in $e_d$ are neglected because, by assumption, $e_d \ll 1$. As we will see below, terms with no $\phi_d$ dependence lead to
corrections of order $e_p^2$. Therefore we can drop the terms with no $\phi_d$ dependence which are also proportional to $e_d$.

With this in mind, we write the contribution to the disturbing function from the annular component $S_c$ as

$$R(S_c) = I_1 + I_2,$$  \hfill (1.47)

$$I_1 = \left\langle G \int_{r_{d,\text{in}}}^{r_{d,\text{out}}} dr_d \Sigma_p(r_d) r_d \int_0^{2\pi} \frac{d\theta}{\sqrt{r_p^2 + r_d^2 - 2r_p r_d \cos \theta}} \right\rangle,$$  \hfill (1.48)

$$I_2 = \left\langle G \int_{r_{d,\text{in}}}^{r_{d,\text{out}}} dr_d \left[ \zeta(r_d) \Sigma_p(r_d) + r_d \frac{d\Sigma_p(r_d)}{dr_d} \right] e_d(r_d) r_d \int_0^{2\pi} \frac{\cos(\theta + \nu) d\theta}{\sqrt{r_p^2 + r_d^2 - 2r_p r_d \cos \theta}} \right\rangle.$$  \hfill (1.49)

Here we expressed $\phi_d = \theta + \nu$, where $\nu = \varpi_p - \varpi_d + \phi_p$, see Figure 1.1. We now evaluate these two contributions.

**Evaluation of $I_1$**

From the definition (1.11) of the Laplace coefficients we can write the inner integral over $\theta$ in Equation (1.48) as $(\pi/r_d) b_{1/2}^{(0)}(r_p/r_d)$ for $r_p < r_d$ (outer disk) and $(\pi/r_p) b_{1/2}^{(0)}(r_d/r_p)$ for $r_p > r_d$ (inner disk). Assuming surface density prescription (1.2) we can write

$$I_1 = \pi G \Sigma_0 \left\langle \int_{r_{d,\text{in}}}^{r_{d,\text{out}}} \left( a_{\text{out}}/r_d \right)^p \frac{r_d}{r_p} b_{1/2}^{(0)} \left( \frac{r_d}{r_p} \right) dr_d + \int_{r_p}^{r_{d,\text{out}}} \left( a_{\text{out}}/r_d \right)^p \frac{r_d}{r_p} b_{1/2}^{(0)} \left( \frac{r_d}{r_p} \right) dr_d \right\rangle.$$  \hfill (1.50)

We now define auxiliary function

$$I(x, y, z) \equiv \int_x^1 \alpha^y b_{1/2}^{(z)}(\alpha) d\alpha,$$  \hfill (1.51)

and a new constant factor

$$K = \pi G \Sigma_0 a_{\text{out}}^p a_p^{1-p}.$$  \hfill (1.52)
With these definitions we re-write expression (1.50) as

\[ I_1 = K \left( \left( \frac{r_p}{a_p} \right)^{1-p} [I(a_1/r_p, 1 - p, 0) + I(r_p/a_2, p - 2, 0)] \right). \] (1.53)

We note that \( a_1 \) and \( a_2 \) in these expressions approximate \( r_{d,in} = a_1(1 - e_1) \) and \( r_{d,out} = a_2(1 - e_2) \), correspondingly. However, the difference is a correction linear in disk eccentricity \( e_d \) and should be ignored for \( I_1 \).

We now proceed to the last, time averaging, step. For illustration we perform it first on the second integral in this expression, by expanding it in Taylor series in small quantity \( r_2 - \alpha_2 \), where \( r_2 = r_p/a_2 \), and \( \alpha_2 = a_p/a_2 \). We have

\[ I(r_p/a_2, p - 2, 0) = I(\alpha_2, p - 2, 0) - (r_2 - \alpha_2)\alpha_2^{p-2} b_{1/2}^{(0)}(\alpha_2) - \frac{(r_2 - \alpha_2)^2}{2} \frac{d}{d\alpha_2} \left[ \alpha_2^{p-2} b_{1/2}^{(0)}(\alpha_2) \right]. \] (1.54)

We may relate \( r_p \) and \( a_p \) using the eccentric anomaly \( E \) as \( r_p = a_p(1 - e_p \cos E) \). Then \( r_2 - \alpha_2 = -\alpha_2 e_p \cos E \) and

\[ \left( \frac{r_p}{a_p} \right)^{1-p} = 1 - (1 - p)e_p \cos E - \frac{p(1-p)}{2} e_p^2 \cos^2 E. \] (1.55)

Using these relations, the second integrand in (1.53) becomes (retaining only terms up to \( e_p^2 \))

\[
\begin{align*}
K \left< 1 - (1 - p)e_p \cos E - \frac{p(1-p)}{2} e_p^2 \cos^2 E \right> I(\alpha_2, p - 2, 0) \\
+ K \left< [1 - (1 - p)e_p \cos E] e_p \cos E \alpha_2^{p-1} b_{1/2}^{(0)}(\alpha_2) \right> \\
- \frac{K}{2} \left< \alpha_2^2 e_p^2 \cos^2 E \frac{\partial}{\partial \alpha_2} \left[ \alpha_2^{p-2} b_{1/2}^{(0)}(\alpha_2) \right] \right>. \tag{1.56}
\end{align*}
\]
Using $\langle \cos E \rangle = -e_p/2$ and $\langle \cos^2 E \rangle = 1/2$, Equation (1.56) reduces to

$$K \left[ (1 + e_p^2)(2 - p) \right] I(\alpha_2, p - 2, 0) + Ke_p^2 \left[ 2(p - 1)\alpha_2^2 b_1^{(0)}(\alpha_2) - \frac{d}{d\alpha_2} \left[ \alpha_2^2 b_1^{(0)}(\alpha_2) \right] \right].$$

(1.57)

We can apply the identical procedure to the first integral of Equation (1.53), resulting in

$$K \left[ 1 + e_p^2(2 - p) \right] I(\alpha_1, 1 - p, 0) + Ke_p^2 \left[ 2(2 - p)\alpha_1^2 b_1^{(0)}(\alpha_1) - \frac{\partial}{\partial\alpha_1} \left[ \alpha_1^3 b_1^{(0)}(\alpha_1) \right] \right].$$

(1.58)

where $\alpha_1 = a_1/a_p$. Note that the term second order in $e_p$ must be included in $r_1 - \alpha_1 = \alpha_1 e_p \cos E + \alpha_1 e_p^2 \cos^2 E$, where $r_1 = a_1/r_p$. The sum of (1.57) and (1.58) is equal to $I_1$ and represents the axisymmetric part of the disk disturbing function from the region $S_c$.

**Calculation of $I_2$**

In order to calculate $I_2$ — the non-axisymmetric component of $R(S_c)$ we use the prescription (1.2) for $\Sigma_p$ and $e_d$, and expand $\cos(\theta + v)$:

$$I_2 = -G(p + q) \left( \int_{r_{d,\text{in}}}^{r_{d,\text{out}}} dr_d \, r_d \Sigma_0 \left( \frac{a_{\text{out}}}{r_d} \right)^{p+q} \int_0^{2\pi} \frac{\cos \theta \cos v - \sin \theta \sin v}{\sqrt{r_d^2 + r_p^2 - 2r_pr_d \cos \theta}} d\theta \right).$$

(1.59)

In the inner integral over $\theta$ terms with $\sin \theta$ in the numerator vanish upon integration, while the terms with $\cos \theta$ result in Laplace coefficients $b_1^{(1)}$, see definition (1.11). Separately accounting for the contributions from the inner and outer disks when integrating over $r_d$ we obtain, analogous to Equation (1.53)

$$I_2 = -Ke_d(a_p)(p + q) \left( \cos v \left( \frac{r_p}{a_p} \right)^{1-p-q} \left[ I(r_p/a_2, p + q - 2, 1) + I(a_1/r_p, 1 - p - q, 1) \right] \right).$$

(1.60)
The final step of time-averaging is somewhat more challenging here because the cos $v$ term introduces additional time-dependence through $\phi_p$. It can be taken care of using the definition $v = (\varpi_p - \varpi_d) + \phi_p$ and the relation $\phi_p = E + e_p \sin E$ accurate to linear order in $e_p$. As before, we also expand integrals in (1.60) in a series in the small quantities $r_{1,2} - \alpha_{1,2}$. Since we are not interested in the terms $O(e_d e_p^2)$ and higher order (small factor $e_d$ is already present in Equation (1.60)), we only expand to first order in $e_p$. As a result of tedious but straightforward calculation we find

$$I_2 = -K e_d (a_p) e_p \cos (\varpi_p - \varpi_d) (p + q) \left[ \frac{(p + q - 3)}{2} I(\alpha_1, 1 - p - q, 1) - \frac{1}{2} \alpha_1^{2-p-q} b_{1/2}^{(1)}(\alpha_1) \right] + \frac{(p + q - 3)}{2} I(\alpha_2, p + q - 2, 1) + \frac{1}{2} \alpha_2^{p+q-1} b_{1/2}^{(1)}(\alpha_2) \right].$$ (1.61)

This completes our calculation of $R(S_c)$.

### 1.10.2 Contribution from the Inner Crescent $S_i$

We now calculate the disturbing function $R(S_i)$, given by Equation (1.43) with integration carried out over the inner crescent $S_i$. The width of the crescent is $O(e_d)$ meaning that we need to keep all variables only up to first order in $e_p$. The integrand, which led to an axisymmetric contribution in the case of $R(S_c)$ now leads to a non-axisymmetric contribution when integrated over this non-axisymmetric region of the disk.

Consider an ellipse with periastron distance $a_{p,1} = a_1(1 - e_1)$ and apoapstron distance $a_{a,1} = a_1(1 - e_1)$ bounding $S_i$ on the outside. Define the angle $\xi(r_d)$ as the angle between the periastron of this ellipse and the point of intersection of the ellipse and a circle of radius $r_d$, $a_{p,1} < r_d < a_{a,1}$. Linearizing the equation of an ellipse $r_d = a_1(1 - e_1^2) / [1 + e_1 \cos \xi(r_d)]$ in $e_1$ we get $r_d = a_1(1 - e_1 \cos \xi(r_d))$. This yields

$$\xi(r_d) = \arccos \frac{a_1 - r_d}{e_1 a_1},$$ (1.62)
where the arccos function is the inverse cosine function. We write explicitly

\[
R(S_i) = \left( G \Sigma_0 a^p_{\text{out}} \int_{a_{p,1}}^{a_{1,1}} \frac{r_d^{1-p} dr_d}{r_p} \int_{\xi - \Delta \varpi - \phi_p}^{\frac{2\pi - \xi - \Delta \varpi - \phi_p}{r_1}} \frac{d\theta}{\sqrt{1 + \alpha'^2 - 2\alpha' \cos \theta}} \right), \tag{1.63}
\]

where \(\alpha' = r_d/r_p \approx a_1/r_p\), and \(\Delta \varpi = \varpi_p - \varpi_d\). Then using the relation

\[
(1 + \alpha'^2 - 2\alpha' \cos \theta)^{-1/2} = \frac{1}{2} b_{1/2}(\alpha') + \sum_{j=1}^{\infty} b_{1/2}^{(j)}(\alpha') \cos (j\theta), \tag{1.64}
\]

the inner integral over \(\theta\) becomes

\[
\left[ \pi - \xi(r_d) \right] b_{1/2}^{(0)}(\alpha') - \sum_{j=1}^{\infty} \frac{2}{j} b_{1/2}^{(j)}(\alpha') \sin \left[ j\xi(r_d) \right] \cos (j(\Delta \varpi + \phi_p)). \tag{1.65}
\]

Then we may write

\[
R(S_i) = G \Sigma_0 a^p_{\text{out}} \left\langle \int_{a_{p,1}}^{a_{1,1}} \frac{r_d^{1-p} dr_d}{r_p} \left[ b_{1/2}^{(0)} \left( \frac{a_1}{r_p} \right) \left[ \pi - \xi(r_d) \right] \right] - \sum_{j=1}^{\infty} \frac{2}{j} b_{1/2}^{(j)} \left( \frac{a_1}{r_p} \right) \sin \left[ j\xi(r_d) \right] \cos (j(\Delta \varpi + \phi_p)) \right\rangle. \tag{1.66}
\]

We will use the following definite integrals

\[
\int_{a_{p,1}}^{a_{1,1}} \left[ \pi - \xi(r_d) \right] dr_d = \pi a_1 e_1, \quad \int_{a_{p,1}}^{a_{1,1}} \sin [\xi(r_d)] dr_d = \frac{\pi}{2} e_1 a_1, \quad \int_{a_{p,1}}^{a_{1,1}} \sin [j\xi(r_d)] dr_d = 0, \tag{1.67}
\]

for integer \(j > 1\). Then in (1.66) we may ignore the terms in the sum with \(j > 1\):

\[
R(S_i) = \pi G \Sigma_0 \left( \frac{a_{\text{out}}}{a_1} \right)^p e_1 a_1^2 \left\langle r_p^{-1} \left[ b_{1/2}^{(0)} \left( \frac{a_1}{r_p} \right) - b_{1/2}^{(1)} \left( \frac{a_1}{r_p} \right) \cos \Delta \varpi \cos \phi_p - \sin \Delta \varpi \sin \phi_p \right] \right\rangle, \tag{1.68}
\]

48
where $e_1$ and $a_1$ are the disk eccentricity and semi-major axis respectively, evaluated at the inner edge. Using $a_1/r_p = \alpha_1 + e_p \alpha_1 \cos E$ and the relation between $\phi_p$ and $E$ this becomes

$$R(S_i) = \pi G \Sigma_0 \left( \frac{a_{\text{out}}}{a_1} \right)^p a_1 e_1 \left\{ (\alpha_1 + e_p \alpha_1 \cos E) \right\} \left[ b_{1/2}(\alpha_1) + e_p \alpha_1 \cos E \frac{\partial b_{1/2}(\alpha_1)}{\partial \alpha_1} \right]$$

$$\times \left( \cos \Delta \omega \cos E - \sin \Delta \omega \sin E - e_p \cos \Delta \omega \sin^2 E - e_p \sin \Delta \omega \cos E \sin E \right) \right\}.$$  (1.69)

Expanding all products in this expression one gets a total of 20 terms. It is straightforward to angle-average them as before. Keeping only terms of order $O(e_1 e_p)$ and substituting $e_1 = e_0 \left( a_{\text{out}}/a_1 \right)^q$ we find that the disturbing function from the inner crescent is given by

$$R(S_i) = \frac{1}{2} Ke_d(a_p) e_p \cos (\omega_p - \omega_d) \alpha_1^{2-p-q} \left[ b_{1/2}(\alpha_1) - \alpha_1 \frac{\partial b_{1/2}(\alpha_1)}{\partial \alpha_1} \right].$$  (1.70)

### 1.10.3 Contribution from the Outer Crescent $S_o$

The derivation of $R(S_o)$ follows the same basic concept as that of $R(S_i)$ except that now $\alpha' = r_p/r_d$. As a result one finds the contribution of the outer crescent $S_o$ to be given by

$$R(S_o) = \frac{1}{2} Ke_d(a_p) e_p \cos (\omega_p - \omega_d) \alpha_2^{p+q-1} \left[ 2b_{1/2}(\alpha_2) + \alpha_2 \frac{\partial b_{1/2}(\alpha_2)}{\partial \alpha_2} \right].$$  (1.71)

### 1.10.4 Putting Everything Together.

Plugging Equations (1.57), (1.58), (1.61), (1.70), (1.71) into the expression (1.44) we find that the total eccentric disk-induced disturbing function, including the eccentricity independent term and terms proportional to $e_p^2$ and $e_de_p$, is given by

$$R = K \left[ \psi_0 + \psi_1 e_p^2 + \psi_2 e_d(a_p) e_p \cos (\omega_p - \omega_d) \right]$$  (1.72)
with

\[
\psi_0(\alpha_1, \alpha_2) = I(\alpha_1, 1 - p, 0) + I(\alpha_2, p - 2, 0) \tag{1.73}
\]

\[
\psi_1(\alpha_1, \alpha_2) = \frac{1}{4} \left[(1 - p)(2 - p)\psi_0(\alpha_1, \alpha_2) + 2(p - 1)\alpha_2^{p-1}b_{1/2}^{(0)}(\alpha_2) - \frac{d}{d\alpha_2}\left(\alpha_2^{p}b_{1/2}^{(0)}(\alpha_2)\right)\right] + 2(2 - p)\alpha_2^{2-p}b_{1/2}^{(0)}(\alpha_2) - \frac{\partial}{\partial\alpha_1}\left(\alpha_1^{3-p}b_{1/2}^{(0)}(\alpha_1)\right) \tag{1.74}
\]

\[
\psi_2(\alpha_1, \alpha_2) = -\frac{(p + q)(p + q - 3)}{2} [I(\alpha_1, 1 - p - q, 1) + I(\alpha_2, p + q - 2, 1)] + \frac{\alpha_1^{2-p-q}}{2} \left[(p + q - 1)b_{1/2}^{(1)}(\alpha_1) + \alpha_1 \frac{\partial b_{1/2}^{(1)}}{\partial\alpha_1}\right] + \frac{\alpha_2^{p+q-1}}{2} \left[(2 - p - q)b_{1/2}^{(1)}(\alpha_2) + \alpha_2 \frac{\partial b_{1/2}^{(1)}}{\partial\alpha_1}\right] \tag{1.75}
\]

This completes our calculation of the disturbing function due to an eccentric disk with properties given by Equation (1.2).

### 1.10.5 Asymptotic Behavior

Astrophysical disks typically span several orders of magnitude in radius. It is then plausible that far from the disk boundaries we can ignore the edge effects, i.e. the expression for the disturbing function does not depend on the \(a_{\text{in}}\) and \(a_{\text{out}}\) as \(a_{\text{in}} \to 0\) and \(a_{\text{out}} \to \infty\). This corresponds to the limit of \(\alpha_1, \alpha_2 \to 0\). Using Taylor expansion

\[
b_{1/2}^{(0)}(\alpha) = 2 + \frac{\alpha^2}{2}, \quad b_{1/2}^{(1)}(\alpha) = \alpha + \frac{3}{8}\alpha^3, \tag{1.76}
\]

for small \(\alpha\) in Equations (1.74)-(1.75), we determined that \(\psi_1\) is convergent and independent of \(\alpha_1, \alpha_2\) as \(\alpha_1, \alpha_2 \to 0\) go to zero for \(-1 < p < 4\). Similarly, \(\psi_2\) is convergent as \(\alpha_1, \alpha_2 \to 0\) for \(-2 < p + q < 5\). Convergence limits are illustrated in Figure 1.2.

Provided that the disturbing function is dominated by the local parts of the disk (i.e. the values of \(p\) and \(q\) fall within the white region in Figure 1.2) and the values of coefficients \(\psi_{1,2}\) are independent of \(\alpha_{1,2}\) when the disk edges are well separated from the planetesimal semi-
major axis \((\alpha_{1,2} \to 0)\), we can obtain simpler analytical expressions for these coefficients. Indeed, using the fact that 
\[ b^{(0)}_{1/2}(\alpha) = \frac{4}{\pi} K(\alpha), \quad b^{(1)}_{1/2}(\alpha) = \frac{4}{\pi \alpha} [K(\alpha) - E(\alpha)] \]
(here \(E\) and \(K\) are complete elliptic integrals) and series expansions (Gradshteyn & Ryzhik 1994)
\[
K(\alpha) = \frac{\pi}{2} \left( 1 + \sum_{n=1}^{\infty} A_n \alpha^{2n} \right), \quad E(\alpha) = \frac{\pi}{2} \left( 1 - \sum_{n=1}^{\infty} \frac{A_n}{2n-1} \alpha^{2n} \right), \quad A_n = \left[ \frac{(2n)!}{2^{2n}(n!)^2} \right]^2
\]
we can provide asymptotic expressions for \(\psi_{1,2}\) as follows:
\[
\psi_1(0,0) = -\frac{1}{2} + \frac{(1-p)(2-p)}{2} \sum_{n=1}^{\infty} \frac{(4n+1)A_n}{(2n+2-p)(2n+p-1)},
\]
\[
\psi_2(0,0) \to \frac{3}{2} - (p+q)(p+q-3) \sum_{n=2}^{\infty} \frac{2n(4n-1)A_n}{(2n-1)(2n+1-p-q)(2n-2+p+q)}.\]

The behavior of \(\psi_1(0,0)\) and \(\psi_2(0,0)\) as functions of \(p\) and \(p+q\) respectively are shown in Figure 1.3.

1.11 Appendix B: Details of the Numerical Verification

Here we describe the details of the numerical verification of our analytical results, see §1.4.5.

We directly integrated orbits of planetesimals affected by the gravity of an eccentric disk using the MERCURY package (Chambers 1999). All our integrations employed the Bulirsch-Stoer algorithm (Press et al. 1992). Accelerations due to the gravity of an eccentric disk \(g_d\) (used as an input for our integrations) were computed at different positions and for different disk parameters via direct numerical integration as
\[
g_d(\mathbf{r}) = -G \int_{\mathbf{s}} \frac{\Sigma(\mathbf{r}_d)}{|\mathbf{r} - \mathbf{r}_d|^3} d\mathbf{S}(\mathbf{r}_d),\]

51
where \(dS(r_d)\) is a surface element centered on \(r_d\). This two-dimensional integral was performed using standard integration by quadratures in SciPy. We used a small softening parameter in the integrand to better handle the singularity, and verified convergence to within a percent as we lowered the value of this parameter. The surface density of the eccentric disk was assumed to be given directly by Equation (1.1). In this calculation we did not make an assumption of \(e_d \ll 1\) and thus were not expanding Equation (1.1) in powers of \(e_p\) (as opposed to Equation (1.46)).

1.12 Appendix C: Precessing Disk

Here we explore secular evolution of planetesimals in the case of a disk precessing according to a simple linear prescription \(\varpi_d(t) = \varpi_{d0} + \dot{\varpi}_d t\). Plugging it into Lagrange Equations (1.18)-(1.17) one finds the following solution for the components of eccentricity vector \((k_p, h_p)\) with the initial conditions \(k_p(0) = 0, h_p(0) = 0\):

\[
k_p(t) = \frac{1}{A(A - \dot{\varpi}_d)} \left[ AB_d \left[ \cos(A t + \varpi_{d0}) - \cos(\dot{\varpi}_d t + \varpi_{d0}) \right] + B_b(A - \dot{\varpi}_d) \left[ \cos(A t) - 1 \right] \right],
\]

\[
h_p(t) = \frac{1}{A(A - \dot{\varpi}_d)} \left[ AB_d \left[ \sin(A t + \varpi_{d0}) - \sin(\dot{\varpi}_d t + \varpi_{d0}) \right] + B_b(A - \dot{\varpi}_d) \sin(A t) \right],
\]

which generalizes solution (1.19) to the case of non-zero precession. It can again be written as the sum of three distinct contributions as described in §1.7:

\[
e_p(t) = \left\{ \begin{array}{c}
k_p(t) \\
h_p(t)
\end{array} \right\} = -\frac{B_b}{A} \left\{ \begin{array}{c} 1 \\ 0 \end{array} \right\} - \frac{B_d}{A - \dot{\varpi}_d} \left\{ \begin{array}{c} \cos \varpi_d(t) \\ \sin \varpi_d(t) \end{array} \right\} + \left[ (AB_d)^2 + (B_b(A - \dot{\varpi}_d))^2 + 2AB_dB_b(A - \dot{\varpi}_d) \cos \varpi_{d0} \right]^{1/2} \left\{ \begin{array}{c} \cos(A t + \phi) \\ \sin(A t + \phi) \end{array} \right\} \]

52
where \( \phi \) is a phase defined analogous to (1.23) and is a function of \( \varpi_{d0} \), \( A \), \( B_d \) and \( B_b \). Note that in the case of a precessing disk, forced eccentricity due to the disk changes in time.
Figure 1.10: Behavior of the pre-factors for the axisymmetric ($\psi_1$) and non-axisymmetric ($\psi_2$) components of the disturbing function near disk edges. Different panels show for different disk models the dependence of $\psi_1$ (left) and $\psi_2$ (right) on $\alpha_2 = a_p/a_2$, for different values of $\alpha_1 = a_1/a_p$ (shown on panel), with $a_1$ and $a_2$ being the inner and outer semi-major axes of the disk. For the chosen values of $p$ and $q$, $\psi_1$ and $\psi_2$ are essentially constant except as $\alpha_1$ or $\alpha_2$ get close to unity. As a result, $\psi_1$ and $\psi_2$ are essentially constant far from disk edges in these models. This is not the case for the model with $p + q = -1$ in panel b, which is featured in §1.8.2 and Figure 1.8.
Bibliography


Heppenheimer, T. A. 1980, icarus, 41, 76


Murray C. D. & Dermott S. F. 1999, Solar System Dynamics

Ward, W. R. 1981, icarus, 47, 234

Chapter 2

Birth Locations of the Kepler Circumbinary Planets

2.1 Abstract

The Kepler mission has discovered about a dozen circumbinary planetary systems, all containing planets on \( \sim 1 \) AU orbits. We place bounds on the locations in the circumbinary protoplanetary disk, where these planets could have formed through collisional agglomeration starting from small (km-sized or less) planetesimals. We first present a model of secular planetesimal dynamics that accounts for the (1) perturbation due to the eccentric precessing binary, as well as the (2) gravity and (3) gas drag from a precessing eccentric disk. Their simultaneous action leads to rich dynamics, with (multiple) secular resonances emerging in the disk. We derive analytic results for size-dependent planetesimal eccentricity, and demonstrate the key role of the disk gravity for circumbinary dynamics. We then combine these results with a simple model for collisional outcomes and find that in systems like Kepler 16, planetesimal growth starting with 10-100 m planetesimals is possible outside a few AU. The exact location exterior to which this happens is sensitive to disk eccentricity, density and precession rate, as well as to the size of the first generation of planetesimals. Strong perturbations from the binary in the inner part of the disk, combined with a secular resonance at a few AU inhibit the growth of km-sized planetesimals within 2 – 4 AU of the binary. In
situ planetesimal growth in the Kepler circumbinary systems is possible only starting from large (few-km-sized) bodies in a low-mass disk with surface density $\lesssim 500 \text{ g cm}^{-2}$ at 1 AU.

### 2.2 Introduction

Recent discoveries of exoplanets in stellar binaries by radial velocity surveys and the *Kepler* mission stimulated significant interest in understanding the origin of such planetary systems. Planet-hosting stellar binaries come in two flavors, commonly denoted as S-type or P-type (Dvorak, 1982). They correspond to systems where the planet orbits one star with the other as an external perturber (S-type), or where the planet is in orbit around both stars (P-type).

In this paper, we focus on P-type or circumbinary systems. *Kepler* has so far revealed to us eight such systems, all containing sub-Jovian planets in orbits around main-sequence binaries. These systems have a range of parameters, but generally the stars have masses $\sim M_\odot$, and are on moderately eccentric orbits with semi-major axes $a_b \sim 0.1 - 0.2$ AU. The planetary orbits have semi-major axes smaller than 1.1 AU. The parameters of known *Kepler* circumbinary planets are summarized in Table 2.1.

There may also exist another population of circumbinary planets hinted at by transit timing of post-common envelope binaries. The most plausible of these are two planets around the NN Serpentis binary system (Marsh et al., 2014). It has been suggested (Völschow et al., 2014) that such planets are not primordial and formed from matter ejected during common envelope evolution. There are also a few directly imaged long period planetary mass circumbinary companions at projected separations $\gtrsim 100$ AU from the host star, e.g. ROXs 42Bb (Currie et al., 2014). We will not address the origin of such systems in this work.

Planet formation in a circumbinary disk faces challenges due to vigorous planetesimal excitation driven by the non-Newtonian potential of the binary. Large relative velocities between planetesimals are harmful to coagulation because collisions lead to planetesimal destruction rather than merging. It is generally believed that the circumbinary planets could
not have formed in situ via collisional agglomeration (Meschiari 2012; Paardekooper et al. 2012; Marzari et al. 2013), due to the high collision speeds of km sized planetesimals. That said, Meschiari (2014) proposes a model for in situ formation, in which large planetesimals form at the pressure maximum near the inner edge of the disk where the surface density of solids is expected to be increased. Bromley & Kenyon (2015) suggest that circumbinary planetesimals may settle onto special class of orbits where their growth would be promoted, and we will comment on this work further (§2.11.1).

<table>
<thead>
<tr>
<th>System</th>
<th>$M_p$</th>
<th>$R_p$</th>
<th>$M_s$</th>
<th>$R_s$</th>
<th>$a_b$</th>
<th>$e_b$</th>
<th>$a_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>K34</td>
<td>1.05</td>
<td>1.16</td>
<td>1.02</td>
<td>1.09</td>
<td>0.23</td>
<td>0.52</td>
<td>1.09</td>
</tr>
<tr>
<td>K16</td>
<td>0.69</td>
<td>0.65</td>
<td>0.20</td>
<td>0.23</td>
<td>0.22</td>
<td>0.16</td>
<td>0.70</td>
</tr>
<tr>
<td>K47</td>
<td>1.04</td>
<td>0.96</td>
<td>0.36</td>
<td>0.35</td>
<td>0.084</td>
<td>0.024</td>
<td>0.30, 0.993</td>
</tr>
<tr>
<td>K38</td>
<td>0.95</td>
<td>1.76</td>
<td>0.25</td>
<td>0.23</td>
<td>0.15</td>
<td>0.32</td>
<td>0.48</td>
</tr>
<tr>
<td>KIC 4862625</td>
<td>1.47</td>
<td>1.7</td>
<td>0.37</td>
<td>0.34</td>
<td>0.18</td>
<td>0.20</td>
<td>0.64</td>
</tr>
<tr>
<td>K 413</td>
<td>0.82</td>
<td>0.78</td>
<td>0.54</td>
<td>0.48</td>
<td>0.10</td>
<td>0.037</td>
<td>0.36</td>
</tr>
<tr>
<td>K35</td>
<td>0.89</td>
<td>1.03</td>
<td>0.81</td>
<td>0.79</td>
<td>0.18</td>
<td>0.14</td>
<td>0.60</td>
</tr>
<tr>
<td>KIC 9632895</td>
<td>0.93</td>
<td>0.83</td>
<td>0.19</td>
<td>0.2</td>
<td>0.18</td>
<td>0.05</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Recently Rafikov (2013) proposed that the gravitational effect of a massive axisymmetric protoplanetary disk may strongly suppress eccentricities of circumbinary planetesimals. That happens because disk gravity drives rapid relative precession of planetesimal and binary orbits, reducing the time-averaged non-axisymmetric component of the binary potential, which drives planetesimal eccentricity. However, simulations show that circumbinary disks do not remain axisymmetric and tend to develop eccentricity themselves (Pelupessy & Portegies Zwart 2013; Meschiari 2014). Silsbee & Rafikov (2015; hereafter SR15), developed a formalism to calculate the gravitational effect of an eccentric disk, and applied those results to planet formation in S-type systems. Rafikov & Silsbee (2015a; hereafter RS15a), additionally included the effects of gas drag, which damps free eccentricity and produces apsidal alignment of equal-size planetesimals. The dynamical results of RS15a were applied in
Rafikov and Silsbee (2015b; hereafter RS15b) to study collision outcomes in S-type planetary systems.

In this paper we extend these calculations to P-type systems. In the following discussion of planetesimal dynamics we account for the combined effects of (1) gas drag, (2) gravitational perturbations from the central binary and (3) gravitational perturbations from the eccentric protoplanetary disk, and derive expressions for the encounter velocities of circumbinary planetesimals. We then use these results to determine which sizes of planetesimals are able to grow at different locations in the disk through mutual collisions.

In Section 2.3, we describe our model system. In Section 2.4 we write down the disturbing function characterizing the gravitational perturbations from the binary and the disk, and present the equations governing planetesimal dynamics. In Sections 2.5 - 2.8 we describe planetesimal dynamics in several different regimes. Section 2.9 discusses collision outcomes calculated using the prescription described in Appendix C. Section 2.10 critically assesses our underlying assumptions, and explores the outcome of relaxing some of them. Section 2.11 compares our results with those in the literature. Finally, we summarize our main conclusions about circumbinary planet formation in Section 2.12.

2.3 General Setup

Our model system is a close stellar binary consisting of a primary with mass $M_p$ and a secondary with mass $M_s < M_p$ in orbit with semi-major axis $a_b$ and eccentricity $e_b$. Orientation of the binary is given by the apsidal angle $\varpi_b$ with respect to a fixed reference direction. Binary orientation (and $\varpi_b$) may in general vary in time as a result of binary precession; we consider this possibility in Section 2.7. For convenience, we define $M_b = M_s + M_p$, and $\mu = M_s/M_b$. Throughout this paper, unless otherwise noted, we provide numerical estimates for a fiducial system which has the binary parameters of Kepler 16: $M_p = 0.69M_\odot$, $M_s = 0.2M_\odot$, $a_b = 0.22$ AU, $e_b = 0.16$. 
Orbiting the barycenter of the binary is a gaseous protoplanetary disk of mass $M_d$. Simulations generally find disks around binaries with $\mu \sim 1/2$ to be tidally truncated on the inside, resulting in the inner cavity relatively devoid of gas (Artymowicz & Lubow 1994; Pelupessy & Portegies Zwart 2013). According to these calculations the inner edge of the disk is truncated at the inner radius $a_{in} = (1.7 - 3.3)a_b$, increasing as $e_b$ increases from 0 to 0.7. Most of the observed stellar orbits in circumbinary systems are on the low end of that eccentricity range (see Table 2.1, so we assume that the disk is truncated on the inside at $a_{in} = 2a_b$. This is somewhat lower than $a_{in} \approx 3a_b$ favored by Pelupessy & Portegies Zwart (2013).

Our disk model is analogous to that adopted in Silsbee & Rafikov (2015). We assume that circumbinary disk streamlines are ellipses with a common apsidal line and foci at the barycenter of the binary. The orientation of the disk is then defined by a single apsidal angle $\varpi_d$, which may be a function of time (see Section 2.6.1) if the disk precesses as a solid body (see a discussion of this assumption in §2.6). We adopt power law scalings for the surface density at periastron $\Sigma_d$ and eccentricity of gas streamlines $e_d$ given by

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

where $a$ is the semi-major axis of the elliptical fluid trajectory, and $a_0$ is a reference distance, which we take to be 1 AU. These dependencies lead to non-trivial surface density behavior described in Statler (2001); SR15.

For numerical estimates in this work we adopt the following set of disk parameters: $\Sigma_0 = 3,000 \text{ g cm}^{-2}$, $e_0 = 0.024$, and $\Sigma_d$ and $e_d$ slopes $p = 1.5$ and $q = 1$. These and Kepler-16 binary parameters are listed in Table 2.2 for convenience.

Our choice of $p$ corresponds to the MMSN model of Hayashi (1981). This value of $p$ also follows from the decretion disk model of Pringle (1991), assuming no mass to pass through the inner boundary of the disk (Rafikov 2013). Our fiducial values of $q$ and $e_0$ are chosen to
Table 2.2: Fiducial system parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_p$</td>
<td>$0.69 M_\odot$</td>
</tr>
<tr>
<td>$M_s$</td>
<td>$0.20 M_\odot$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.22</td>
</tr>
<tr>
<td>$a_b$</td>
<td>0.22 AU</td>
</tr>
<tr>
<td>$e_b$</td>
<td>0.16</td>
</tr>
<tr>
<td>$\Sigma_0$</td>
<td>3,000 g cm$^{-2}$</td>
</tr>
<tr>
<td>$e_0$</td>
<td>0.024</td>
</tr>
<tr>
<td>$p$</td>
<td>1.5</td>
</tr>
<tr>
<td>$q$</td>
<td>1</td>
</tr>
</tbody>
</table>

correspond to the behavior of the forced eccentricity of free particles orbiting in the potential of the binary with eccentricity $e_b$ (Moriwaki & Nakagawa 2004), see Equation (2.10). This estimate is just a reasonable zeroth order guess for $e_d(a)$, as eccentricity of the fluid disk is also affected by pressure, viscous forces and disk gravity.

Assuming the mass to be concentrated in the outer part of the disk (i.e. $p < 2$), we can relate $M_d$ to the outer radius of the disk $a_{\text{out}}$ and $\Sigma_0$:

$$
M_d = \frac{2\pi}{2 - p} \Sigma_0 a_0^p a_{\text{out}}^{2-p} 
\approx 0.037 M_\odot \frac{\Sigma_0}{3,000 \text{ g cm}^{-2}} \left( \frac{a_{\text{out}}}{75 \text{AU}} \right)^{0.5},
$$

where the numerical estimate has been performed for $p = 1.5$. The outer truncation radius of 75 AU is characteristic of the more massive disks around single stars (Andrews et al. 2009; Harris et al. 2012).

### 2.4 Equations of Planetesimal Dynamics

We now outline the mathematical framework that we use to describe planetesimal dynamics. A planetesimal orbit is characterized by the semi-major axis $a_p$ with respect to the binary barycenter, eccentricity $e_p$, and apsidal angle $\varpi_p$; its mean motion is $n_p = \sqrt{GM_b/a_p^3}$.
Planetesimal motion is affected by both conservative forces — gravity of the binary and the disk — and non-conservative gas drag. Analogous to Rafikov (2013), SR15, RS15a, we employ secular perturbation theory (Murray & Dermott, 1999; Moriwaki & Nakagawa, 2004) to determine the effect of the former on planetesimal eccentricity behavior. This approach uses the disturbing function, described next in §2.4.1 which accounts for the perturbations to planetesimal motion produced by the gravity of a massive protoplanetary disk and the non-Newtonian gravity of the binary. Our treatment of the effects of gas drag on planetesimal dynamics is described in §2.4.2. All these components are combined in §2.4.3, resulting in a set of general equations (2.18)-(2.19) describing secular planetesimal dynamics in circumbinary planetesimal disks.

### 2.4.1 Disturbing Function

Planetesimal disturbing function $R = R_d + R_b$ consists of contributions due to an eccentric disk $R_d$ and due to the non-Newtonian gravity of the binary $R_b$.

SR15 derived the disturbing function due to a disk with eccentricity and surface density given by Equation (2.1), and distance-independent apsidal angle $\varpi_d$ as

\[
R_d = n_p a_p^2 \left[ \frac{A_d}{2} e_p^2 + B_d e_p \cos (\varpi_p - \varpi_d) \right],
\]

\[
A_d = 2\pi \psi_1 \frac{G \Sigma_d(a_p)}{a_p n_p},
\]

\[
 \approx -1.6 \times 10^{-3} \text{yr}^{-1}
\]

\[
\times \frac{\Sigma_0}{3,000 \text{ g cm}^{-2}} \frac{\psi_1}{(-0.55)} \left( \frac{0.89 M_\odot}{M_b} \right)^{0.5} a_p^{-1}
\]

\[
B_d = \pi \psi_2 \frac{G \Sigma_d(a_p) e_d(a_p)}{a_p n_p}
\]

\[
 \approx 1.3 \times 10^{-5} \text{yr}^{-1}
\]

\[
\times \frac{\Sigma_0}{3,000 \text{ g cm}^{-2}} \frac{\psi_2}{1.85} \left( \frac{0.89 M_\odot}{M_b} \right)^{0.5} e_0 a_p^{-1}
\]
and \( a_{p,5} \equiv a_p/(5\text{AU}) \). Here \( \psi_1 \) and \( \psi_2 \) are coefficients of order unity (SR15), which are effectively constants far from the disk edges, for \( a_{\text{in}} \lesssim a_p \lesssim a_{\text{out}} \). For \( p = 1.5 \) and \( q = 1 \), SR15 show that \( \psi_1 \approx -0.55 \) and \( \psi_2 \approx 1.85 \), except near the edges of the disk. As shown in SR15, \( A_d \) and \( B_d \) are dominated by local disk properties for our choice of \( p \) and \( q \), so non-power-law behavior of surface density or eccentricity near the edges of the disk will not greatly affect the values of \( \psi_1 \) and \( \psi_2 \).

According to Moriwaki & Nakagawa (2004) the disturbing function due to the binary has the form similar to Equation (2.4):

\[
R_b = n_p a_p^2 \left[ \frac{A_b}{2} e_p^2 + B_b e_p \cos (\varpi_p - \varpi_b) \right],
\]

(2.7)

\[
A_b = \frac{3}{4} \mu (1 - \mu) \frac{n_b^2}{n_p} \left( \frac{a_b}{a_p} \right)^5
\]

(2.8)

\[
\approx 1.3 \times 10^{-4} \text{yr}^{-1} \frac{\mu (1 - \mu)}{0.17} \left( \frac{a_b}{0.22\text{AU}} \right)^2 \times \left( \frac{M_b}{0.89M_\odot} \right)^{0.5} a_p^{-3.5},
\]

\[
B_b = \frac{15}{16} \mu (1 - \mu) (1 - 2\mu) \frac{n_b^2}{n_p} \left( \frac{a_b}{a_p} \right)^6 e_b
\]

(2.9)

\[
\approx 6.5 \times 10^{-7} \text{yr}^{-1} \frac{f(\mu)}{0.096} \left( \frac{a_b}{0.22\text{AU}} \right)^3 \times \left( \frac{M_b}{0.89M_\odot} \right)^{0.5} e_b a_p^{-4.5}.
\]

Here \( n_b = \sqrt{GM_b/a_b^3} \) is the mean motion of the binary, and \( f(\mu) \equiv \mu (1 - \mu) (1 - 2\mu) \).

We use these disturbing functions in Sections 2.5 - 2.7 to calculate the orbital dynamics of planetesimals in several different dynamical regimes.

In the absence of gas drag and disk gravity, free particles in the binary potential attain forced eccentricity (Moriwaki & Nakagawa 2004)

\[
e_{\text{forced}} = \frac{B_b}{A_b} = \frac{5}{4} (1 - 2\mu) a_b e_b
\]

(2.10)

\[
\approx 0.024 \frac{1 - 2\mu}{0.56} e_b \frac{a_b}{0.16} \frac{1\text{AU}}{a_p},
\]
which provides a useful reference value, e.g. for our choice of $e_0$.

Introducing the planetesimal eccentricity vector $\mathbf{e}_p = (k_p, h_p) = e_p(\cos \varpi_p, \sin \varpi_p)$ we can then write down the full disturbing function as

$$R = n_p a_p^2 \left[ \frac{A}{2} (h_p^2 + k_p^2) + B_d k_p \cos \varpi_d ight.$$ 
$$
+ B_d h_p \sin \varpi_d + B_b k_p \cos \varpi_b + B_b h_p \sin \varpi_b \right],
$$

where $\varpi_d = \varpi_d(t)$, $\varpi_b = \varpi_b(t)$.

Here $A = A_d + A_b$ is a precession rate of planetesimal free eccentricity due to the axisymmetric components of both the disk and binary gravity. Its behavior as a function of distance from the binary is shown in Figure 2.1, together with separate curves for $A_d(a_p)$ and $A_b(a_p)$. The fact that $A$ goes through zero at some semi-major axis $a_A$ has very important implications for planetesimal dynamics, see §2.4.1.

The other terms in equation (2.11) that depend on the planetesimal orientation (i.e. $\varpi_p$) describe excitation of planetesimal eccentricity by the torques produced by the non-axisymmetric components of the disk and binary potentials. We discuss the relative role of different contributions to $R$ next.

**Transition Between Binary-Dominated and Disk-Dominated Regimes**

Because $A_b$ and $B_b$ fall off more rapidly with semi-major axis than their disk-related counterparts $A_d$ and $B_d$, see Equations (2.5)-(2.6) and (2.8)-(2.9), we find that outside a certain radius the disturbing function $R$ should be dominated by the disk gravity. This situation is analogous to the so-called DD regime discussed in SR15, in which both the axisymmetric and the non-axisymmetric components of the disturbing function are dominated by the disk.

In the opposite limit, very close to the binary, the gravitational perturbations are dominated by the binary system. This is analogous to the so-called BB regime of SR15 in which both the axisymmetric and the non-axisymmetric components of the disturbing function are
Figure 2.1: Planetesimal precession rate $A = A_b + A_d$ due to both binary and disk gravity (black curve) as a function of $a_p$, calculated assuming the fiducial system parameters in Table 2.2. Dotted and dashed curves represent $A_d(a_p)$ and $A_b(a_p)$. Several reference values corresponding to disk and binary precession rates considered in §2.6.1 and §2.7.1 are also shown (expressed in units of $A|_{6AU} < 0$ — planetesimal precession rate at 6 AU). The vertical dashed line marks $a_p = a_A$, see equation (2.12). For our fiducial binary + disk model (see Table 2.2) $A|_{6AU} = -1.2 \times 10^{-3} \text{yr}^{-1}$, $\dot{\omega}_b = 2.6 \times 10^{-3} \text{yr}^{-1}$.
dominated by the binary. At the same time, the disk still cannot be ignored because of the gas drag.

Equations (2.5) and (2.8) predict that planetesimal precession (i.e. $A$) switches from being binary- to disk-dominated at a characteristic semi-major axis $a_A$, where $|A_d| = |A_b|$: 

$$a_A = \left[ \frac{3\mu(1 - \mu) M_b a_b^2}{8\pi|\psi_1| \sum_0 a_0^p} \right]^{1/(4-p)} \approx 1.9\text{AU} \left[ \left( \frac{a_b}{0.22\text{AU}} \right)^2 \frac{M_b}{0.89M_\odot} \right]^{2/5} \times \frac{3000\text{g cm}^{-2}}{\Sigma_0} \left( \frac{\mu(1 - \mu)}{0.17} \right)$$  

$$a_B = \left[ \frac{15f(\mu) M_b a_b^3 e_b}{16\pi|\psi_2| \sum_0 a_0^{p+q} e_0} \right]^{1/(5-p-q)} \approx 1.5\text{AU} \left[ \frac{e_b}{0.16} \frac{0.024}{e_0} \frac{M_b}{0.89M_\odot} \left( \frac{a_b}{0.22\text{AU}} \right)^3 \right]^{2/5} \times \frac{3000\text{g cm}^{-2} f(\mu)}{\Sigma_0} \frac{0.096}{0.096}$$

where the numerical estimates have been performed for our fiducial system. This transition occurs via a secular resonance where $A = 0$, emerging because disk and binary drive planetesimal precession in opposite directions, see equations (2.5) and (2.8).

On the other hand, planetesimal eccentricity excitation switches from being binary- to disk-dominated at the different characteristic distance $a_B$, where $|B_d| = |B_b|$: 

$$a_B = \left[ \frac{15f(\mu) M_b a_b^3 e_b}{16\pi|\psi_2| \sum_0 a_0^{p+q} e_0} \right]^{1/(5-p-q)} \approx 1.5\text{AU} \left[ \frac{e_b}{0.16} \frac{0.024}{e_0} \frac{M_b}{0.89M_\odot} \left( \frac{a_b}{0.22\text{AU}} \right)^3 \right]^{2/5} \times \frac{3000\text{g cm}^{-2} f(\mu)}{\Sigma_0} \frac{0.096}{0.096}$$

For a single component of the disturbing function (axisymmetric or non-axisymmetric), the region of transition between binary-domination and disk-domination is quite narrow, since both $A_d/A_b$ and $B_d/B_b$ are rising fast with $a_p$ (as $a_p^{5/2}$).

If $a_A$ is widely separated from $a_B$, there will also be a region of the disk where excitation is dominated by the disk and precession by the binary, or vice-versa. In particular, in Section 2.7.2 we consider the limit of a low eccentricity disk ($e_0 \to 0$) in which $a_A$ can be substantially less than $a_B$, see Equations (2.12), and (2.13). Planetesimal dynamics in the
Figure 2.2: Three different dynamical regimes in the space of disk density and semi-major axis. Calculation is done for our fiducial system parameters ($M_p = 0.69 M_\odot$, $M_s = 0.2 M_\odot$, $a_b = 0.22$ AU, $e_b = 0.16$, $p = 1.5$, and $q = 1$), but with lowered disk eccentricity at 1 AU $e_0 = 2.4 \times 10^{-3}$, which is 10\% of the forced eccentricity [2.10] to broaden the DB regime. The vertical red line shows the location at which binary precession rate equals to the disk-driven planetesimal precession, the significance of which is discussed in §2.7.2.
intermediate region $a_A \lesssim a_p \lesssim a_B$ are then analogous to the DB regime of SR15. In the opposite limit of a high-eccentricity disk, $e_0 = e_d(1 \text{ AU}) \gtrsim 0.1$ (which is probably not very realistic), there would exist a region $a_B \lesssim a_p \lesssim a_A$, in which the eccentricity excitation is dominated by the disk, while precession is controlled mainly by the binary — analogous to the BD regime of SR15.

The locations of these regimes in the space of disk density and eccentricity are illustrated in Figure 2.2. The width of the DB regime depends on the degree to which the eccentricity of the disk falls below the free particle eccentricity in the binary potential. Figure 2.2 also includes the line where $|\dot{\varpi}_b| = |A_d|$. This gives roughly the radius outside of which excitation due to the binary is substantially reduced by its precession (see Section 2.7).

### 2.4.2 Gas Drag

Following RS15a, we consider gas drag to damp planetesimal eccentricity as

$$\frac{d\mathbf{e}_p}{dt} = -\frac{\mathbf{e}_p - \mathbf{e}_d}{\tau_d}, \tag{2.14}$$

where $\mathbf{e}_d = (k_d, h_d) = e_d(\cos \varpi_d, \sin \varpi_d)$ is the eccentricity vector of the local disk fluid element, and $\tau_d$ is the eccentricity damping time. This implies that gas drag drives planetesimal orbits towards full alignment with the gas trajectories on a characteristic timescale $\tau_d$.

This timescale is not independent of $\mathbf{e}_p$ in general. For planetesimals with radii $d_p$ on the order of km, a quadratic drag law $\mathbf{F} = -(C_D/2)\pi d_p^2 \rho_d v_r \mathbf{v}_r$ is appropriate (Weidenschilling 1977), where $C_D \approx 0.5$ is the drag coefficient, $\rho_d$ is the local gas density and $v_r$ is the relative planetesimal-gas speed. In this case $\tau_d \propto v_r^{-1}$, and RS15a had $\tau_d \propto e_r^{-1}$, where $\mathbf{e}_r = \mathbf{e}_p - \mathbf{e}_d$ is the relative eccentricity between the object and the gas.

In this work we have chosen to also account for the fact that gas orbits the binary more slowly than a planetesimal because of the radial pressure support in a gaseous disk. This gives rise to additional (predominantly azimuthal) irreducible relative velocity between the
gas and planetesimal $\Delta v_\phi = \eta v_K$, where $v_K$ is the Keplerian speed and the explicit expression for $\eta \ll 1$ is given by eq. (13) of RS15b. This velocity differential does not vanish even when fluid and planetesimal orbits coincide and $e_p = e_d$. To describe this effect, we introduce fiducial eccentricity

$$e_\phi = \frac{\Delta v_\phi}{2E(\sqrt{3}/2) v_K},$$

$$\approx 0.0038 \frac{M_\odot}{M_b} a_{p,5}^{1/2}.$$  

The numerical estimate uses the prescription for the scale height $h$ given in RS15b (their eq. (14)), adapted for a disk temperature\footnote{Stars in P-type binaries typically have lower masses and lower luminosities than the members of S-type binaries studied in SR15b.} of 200 K at 1 AU:

$$h = 0.028 \sqrt{\frac{M_\odot}{M_b}} \left( \frac{a_p}{\text{AU}} \right)^{1/4}. $$

We generalize the expression for the characteristic damping time $\tau_d$ from RS15a as follows:

$$\tau_d = \frac{2^{5/2} \pi^{3/2}}{3CDE(\sqrt{3}/2)} \rho_p d_p \frac{h}{\Sigma_d} \frac{1}{a_p} \left( e_r^2 + e_\phi^2 \right)^{-1/2} \approx 1.6 \times 10^5 \text{ yr} \frac{\rho_p}{3 \text{ g cm}^{-3}} \frac{d_p}{\text{km}} \times \frac{0.89 M_\odot}{M_b} \frac{3,000 \text{ g cm}^{-2}}{\Sigma_0} \frac{0.01}{\sqrt{e_r^2 + e_\phi^2}} a_{p,5}^{13/4}. $$

Here, $E(\sqrt{3}/2) \approx 1.21$ is a complete elliptic integral, $\rho_p$ is the planetesimal bulk density, and we have assumed $p = 1.5$ in the estimate.

Note that the numerical coefficient in equation (2.17) is different from the analogous expression in RS15a because they adopted a rough estimate $\rho_d = \Sigma_d/h$, whereas here we use $\rho_g = \Sigma_d/\sqrt{2\pi h}$, appropriate for an isothermal disk. We have picked the coefficient in the
definition of $e_\phi$ to give the correct damping time (in agreement with Adachi et al. (1976)) in the limit that $e_r \ll e_\phi$.

Equation (2.17) is only approximate for $e_\phi \sim e_r$. However, it does capture the reduction of $\tau_d$ for particles with low values of $e_r$ due to the irreducible velocity differential $\Delta v_\phi$ caused by pressure support. It is thus an improvement over the approximation used in SR15a. Low values of $e_r \lesssim e_\phi$ are more typical at several AU around P-type binaries than in S-type systems.

2.4.3 General Evolution Equations

We now combine the results of §2.4.1 and 2.4.2. We use the standard Lagrange equations (Murray & Dermott 1999; Rafikov 2013) to relate $dk_p/dt$ and $dh_p/dt$ to the disturbing function (2.11). Adding the contributions due to gas drag given by equation (2.14), we find, to lowest order in eccentricity, the following set of evolution equation for planetesimal eccentricity $e_p$:

\[
\frac{dk_p}{dt} = -Ah_p - B_d \sin \varpi_d(t) - B_b \sin \varpi_b(t) - \frac{k_p - k_d}{\tau_d}, \\
\frac{dh_p}{dt} = Ah_p + B_d \cos \varpi_d(t) + B_b \cos \varpi_b(t) - \frac{h_p - h_d}{\tau_d}.
\]

As before, $\varpi_d$ and $\varpi_d$ are in general functions of time.

These master equations provide a basis for subsequent analysis of planetesimal dynamics in Sections 2.5 - 2.7.
2.4.4 General Remarks on Planetesimal Dynamics

Before embarking on a detailed discussion of planetesimal dynamics in the following sections, we outline some general features of planetesimal eccentricity evolution described by equations (2.18)-(2.19).

First, in the case of quadratic drag these equations do not in general admit an analytical solution for arbitrary time dependence of $\varpi_d(t)$ and $\varpi_b(t)$. At the same time, there are several important limits where analytical treatment is possible, and these situations are covered in §2.5, 2.6, 2.7. These solutions allow us to gain important insights into how the binary or disk precession may affect planetesimal dynamics, which remain valid in more complicated setups (§2.8). In addition, some features of the general planetesimal dynamics with both $\varpi_d$ and $\varpi_b$ varying in time can be gleaned by considering a simpler case of linear gas drag, covered in Appendix 2.13 and discussed in §2.8.

Second, we find quite generally that any free eccentricity describing the initial conditions for planetesimal evolution damps away on a characteristic time $\sim \tau_d$. As a result, planetesimal eccentricity $e_p$ inevitably converges to its forced value, which is determined by many factors, see §2.5. This feature of the evolution has been previously pointed out in Beaugé et al. (2010) and RS15a, and implies that after the initial transient lasting for $\sim \tau_d$ planetesimals lose memory of their initial conditions.

In particular, collisions between planetesimals, which perturb them away from the equilibrium eccentricities, may be considered as a minor effect for the dynamics as long as they are infrequent enough, i.e. the mean time between them is longer than $\tau_d$. We provide a discussion of this approximation in Section 2.10.3. Convergence of $e_p$ to a certain fixed state greatly simplifies our analysis as we see in the following sections.

Circumbinary systems exhibit a wide range of planetesimal dynamical behavior throughout the disk. We now describe them under different assumptions about the disk and binary precession. We start by considering a simple case where we ignore the effects of binary precession. Although the latter is likely very important in reality, ignoring it at first allows us
to better illustrate certain aspects of planetesimal dynamics. Thus, we consider the case of non-precessing disk and binary in §\ref{sec:planetesimal_dynamics_no_precession}. We then include the possibility of the disk precession in §\ref{sec:planetesimal_dynamics_disk_precession}. Finally, in §\ref{sec:planetesimal_dynamics_binary_precession} and §\ref{sec:planetesimal_dynamics_both_precession} we consider a possibility of the binary precession.

### 2.5 Planetesimal Dynamics in the Absence of Disk and Binary Precession

If we ignore precession of both the disk and the binary, i.e. take $\varpi_d(t)$ and $\varpi_b(t)$ to be constant in time, we can easily solve equations (2.18)-(2.19). This approximation should be valid for small planetesimals, which have stopping times shorter than the two precession timescales, $\tau_d \lesssim \dot{\varpi}_d, \dot{\varpi}_b$. In this case the eccentricity vector $\mathbf{e}_p$ should rapidly adjust to the fixed values corresponding by the instantaneous values of $\varpi_d$ and $\varpi_b$.

Solutions for planetesimal dynamics in this limit accounting for both the disk and the binary gravity have been previously derived in RS15a, but for S-type binaries. They remain fully valid in the circumbinary case as well, as long as $A_b$ and $B_b$ are understood as being represented by our Equations (2.5) and (2.6). The forced eccentricity to which $\mathbf{e}_p$ converges after the initial period of damping the free eccentricity, is given by a sum of forced terms due to the binary $\mathbf{e}_{f,b}$ and to the disk $\mathbf{e}_{f,d}$:

\[
\mathbf{e}_p = \begin{cases} k_p \\ h_p \end{cases} = \mathbf{e}_{f,b} + \mathbf{e}_{f,d},
\]

\[
\mathbf{e}_{f,b} = \begin{pmatrix} \tau_d^2 B_b^2 \\ 1 + (A\tau_d)^2 \end{pmatrix}^{1/2} \begin{pmatrix} \cos (\varpi_b + \phi_b) \\ \sin (\varpi_b + \phi_b) \end{pmatrix},
\]

\[
\mathbf{e}_{f,d} = \begin{pmatrix} e_d^2 + \tau_d^2 B_d^2 \\ 1 + (A\tau_d)^2 \end{pmatrix}^{1/2} \begin{pmatrix} \cos (\varpi_d + \phi_d) \\ \sin (\varpi_d + \phi_d) \end{pmatrix},
\]
where $\phi_d$ and $\phi_b$ are phase angles given by

$$
\cos \phi_d = \frac{e_d - ABd\tau_d^2}{(e_d^2 + \tau_d^2 B_d^2)^{1/2}} \left[ 1 + (A\tau_d)^2 \right]^{1/2},
$$

and

$$
\cos \phi_b = \frac{-A\tau_d}{\sqrt{1 + (A\tau_d)^2}}.
$$

One can see that in general $e_p$ is misaligned with both the disk and the binary as it is parallel to neither $e_d$ nor $e_b = e_b(\cos \varpi_b, \sin \varpi_b)$. It is also clear that $e_p$ does not vary in time when $\varpi_d$ and $\varpi_b$ are fixed.

### 2.5.1 Relative Planetesimal-Gas Eccentricity

Solution (2.31) implies that the relative planetesimal-gas eccentricity $e_r = |e_p - e_d|$ is given by (RS15a)

$$
e_r = e_c \frac{A\tau_d}{\sqrt{1 + (A\tau_d)^2}},
$$

where the characteristic eccentricity $e_c$ is

$$
e_c = |A|^{-1} \left[ (Ae_d + B_d)^2 + B_b^2 + 2 \cos (\varpi_d - \varpi_b)B_b(Ae_d + B_d) \right]^{1/2}.
$$

We find it convenient to define a characteristic size $d_c$ for which $A\tau_d = 1$ when $\sqrt{e_r^2 + e_\phi^2}$ is replaced with $e_c$ in Equation (2.17), i.e. eccentricity damping time is of order the orbit precession time scale. Equation (2.17) implies that

$$
d_c = \frac{3CD\Sigma_d n_p a_p \rho_p |A| h e_c}{2^5/2\pi^{3/2}}
\quad = 1.3m \left( \frac{M_b}{0.89M_\odot} \right)^{1.5} \frac{3 \text{ g cm}^{-3}}{\rho_p} \frac{a_{p,5}^{-13/4}}{a_{p,5}}.
$$
Because the numerical estimate is made at 5 AU, far outside of $a_A$ and $a_B$ for typical disk parameters, we have assumed for simplicity that the contribution of the binary to the disturbing function is negligible, i.e. $A \approx A_d, B_b \approx 0$. In this regime (far from the star) both $e_c$ and $d_c$ are independent of disk mass $M_d$ (or $\Sigma_0$). This can be seen from Equations (2.26) and (2.27), and the fact that $A_d, B_d$ and $\Sigma_d$ are all proportional to $M_d$.

Using our newly defined $d_c$, we may rewrite equation (2.17) as $A \tau_d = \left( d_c/d_p \right) \left( e_c/\sqrt{e_r^2 + e_\phi^2} \right)$. Combining this result with Equation (2.32) we can solve for $A \tau_d$ and $e_r$:

$$A_d \tau_d = 2^{-1/2} \left[ \frac{1 - K + L}{(d_c/d_p)^2 + K} \right]^{1/2}, \quad (2.28)$$

$$e_r = e_c \left[ \frac{1 - K + L}{1 + K + 2(d_c/d_p)^2 + L} \right]^{1/2}, \quad (2.29)$$

where

$$K \equiv \left( e_\phi/e_c \right)^2(d_c/d_p)^2,$$

$$L \equiv \left[ (K + 1)^2 + 4(d_c/d_p)^2 \right]^{1/2}.$$ 

For most of the paper, we will be considering situations in which $e_c \gg e_\phi$. In this limit, $K \to 0, L \to 2d_c/d_p$, and planetesimal dynamics are determined purely by the values of $d_c$ and $e_c$, so that our results reduce to those of RS15a, with $\tau_d$ longer by a factor of $\sqrt{2\pi}$ (to compensate for the different definition of $\rho_d$).

### 2.5.2 Solution far from the Binary

A useful limit of planetesimal dynamics without disk or binary precession is obtained when we are justified in neglecting the binary perturbation. This regime is naturally realized far from the binary. This is equivalent to the DD regime of SR15 in which disk gravity dominates dynamics, so $|B_d| \gg |B_b|$ and $A \approx A_d$.  

76
It is easy to show in this case that except near the edge of the disk (where edge effects change the values of $\psi_1$ and $\psi_2$), $e_c$ becomes a constant multiple of the local disk eccentricity $e_d$. Indeed, if the precession of the disk is ignored, then Equation (2.26) gives us that $e_c = |B_d/A_d + e_d|$. Using Equations (2.5) and (2.6), this can be rewritten as

$$e_c = e_d \left| \frac{\psi_2}{2\psi_1} + 1 \right| \approx 0.65e_d, \quad (2.30)$$

as for $p = 3/2$, $q = 1$, we have $\psi_2 \approx 1.82$ and $\psi_1 \approx -0.55$. In this regime, planetesimal survival is easier at larger semi-major axes simply because larger $a_p$ means lower $e_d(a_p)$ for $q > 0$.

### 2.5.3 Radial Behavior of $e_r$

The behavior of the relative planetesimal-gas eccentricity given by equations (2.25) and (2.29) is illustrated in Figure 2.3 which shows $e_r$ as a function of orbital distance $a_p$ for different disk models. Unless otherwise noted, all panels assume the Kepler 16 binary parameters with our fiducial disk parameters as displayed in Panel A, and apsidal alignment of the binary and the disk, i.e. $\omega_d = \omega_b$.

The most pronounced feature seen in all panels is the secular resonance, where $A = 0$, see Equation (2.26). It inevitably appears in the non-precessing disks at $a_A$ because $A$ changes sign there, as a result of $A_b$ and $A_d$ having different signs, see Figure 2.1. The resonance location is different in panel B because of the lowered $\Sigma_0$ in this particular disk model, which pushes $a_A$ out to 4.8 AU, compared to $a_A \approx 1.9$ AU in all other panels. Note that although formally $e_c \to \infty$ at the resonance, $e_r$ stays finite there. Moreover, the increase of $e_r$ at the resonance is often not very pronounced: unless $d_p > d_c$, the resonance has little effect on relative eccentricities because the planetesimals are closely coupled to the gas (see Equation (2.29)). The existence of the secular resonance previously suggested in the drag-free environment by R13 was confirmed numerically by [Meschiari (2014)] in his simulations.
Figure 2.3: Radial dependence of $e_r$ (see Equation (2.25)) in a system with no binary or disk precession, but including gravity of both the binary and the disk, for 4 different disk models. Calculations assume the binary parameters of Kepler 16 with fiducial disk parameters unless otherwise labelled; $\omega_d = \omega_b$ is assumed. Characteristic $e_c$ is determined by Equation (2.26). Different colors correspond to different planetesimal sizes, as indicated in panel D.
disk gravity excites the eccentricities of \( d_p = 5 \) km planetesimals in a radially narrow disk region around \( a_p \approx 2.5 \) AU by a factor of \( \sim 10 \) compared to \( e_p \) at that radius in the absence of disk gravity.

We see that lowering the disk eccentricity (panel C) causes \( e_r \) to be dramatically lower outside of the secular resonance. This is because far from the star, disk gravity dominates, and we reach the limit of Section 2.5.2 where \( e_c \) is proportional to \( e_d \). On the other hand, reduction of \( e_0 \) has little effect near the star, where excitation is dominated by the binary. This is because very close to the star, terms \( A_d \) and \( B_d \) are irrelevant in Equation (2.26). At the same time, disk eccentricity still enters through \( e_d \), which is not negligible (see Equation (2.26)), so some variation of \( e_p \) is still present near the star as \( e_0 \) is varied.

Finally, panel D shows a shallower drop-off of planetesimal eccentricity at large \( a_p \) simply because we adopted a model with a different (lower) value of \( q \), and \( e_p \) is in the disk-dominated regime outside the secular resonance.

### 2.6 Dynamics with Disk-Dominated Excitation and Disk Precession

We will find later in this work (§2.9.1) that it is generally most promising to form planets outside of a few AU, relatively far from the binary. At large \( a_p \gtrsim a_A \) planetesimal eccentricity excitation is dominated solely by the disk gravity, even though binary gravity may still affect planetesimal precession, see §2.4.1.

For that reason we will now explore the limit in which the binary plays a negligible role in the planetesimal eccentricity excitation, i.e. \( |B_b| \lesssim |B_d| \). At the same time, the binary is still allowed to contribute significantly to planetesimal precession, i.e. \( A = A_b + A_d \) in general. Thus, the current limit is analogous to the DD and DB regimes of SR15, with the addition of gas drag and disk precession.
Here we also include a (realistic) possibility of the disk precession at some constant rate $\dot{\varpi}_d$. For the purposes of deriving an analytical solution, this rate must be independent of the semi-major axis, because the derivation of $R_d$ in SR15 assumes fluid trajectories to be apsidally aligned at all $a_p$. In practice this assumption is likely to break far from the binary. However, this is not a problem as the disk eccentricity is going to be very low there anyway (Pelupessy & Portegies Zwart 2013; Meschiari 2014).

Setting $\varpi_d = \dot{\varpi}_d t$ and $B_b \to 0$ in Equations (2.18)-(2.19) we find that, as shown in RS15a, these equations admit a periodic solution in the form

$$e_p = \left[ \frac{e_d^2 + \tau_d^2 B_d^2}{1 + (A - \dot{\varpi}_d)^2 \tau_d^2} \right]^{1/2} \begin{pmatrix} \cos (\dot{\varpi}_d t + \phi) \\ \sin (\dot{\varpi}_d t + \phi) \end{pmatrix}, \tag{2.31}$$

where the phase shift $\phi$ with respect to the disk apsidal line is still given by equation (2.23) but with $A$ replaced by $A - \dot{\varpi}_d$. In this solution, the forced eccentricity vector $e_p$ rotates at the rate $\dot{\varpi}_d$ and is fixed in the frame precessing with the disk.

Solution (2.31) implies that the relative planetesimal-gas eccentricity is given by

$$e_r = e_c \frac{(A - \dot{\varpi}_d) \tau_d}{\sqrt{1 + [(A - \dot{\varpi}_d) \tau_d]^2}}, \tag{2.32}$$

where the characteristic eccentricity $e_c$ is now

$$e_c = \left| \frac{B_d}{A - \dot{\varpi}_d} + e_d \right|. \tag{2.33}$$

Also, equations (2.27)-(2.29) still hold with the provision that $A$ is replaced with $A - \dot{\varpi}_d$ everywhere.
2.6.1 Effects of Disk Precession

We now explore the effect of disk precession on the behavior of planetesimal eccentricity. We examine the run of \( e_r \) in the disk for different disk models in Figure 2.4 assuming the binary parameters of Kepler 16. All disks have \( p = 1.5, q = 1, \) and \( \Sigma_0 = 3,000 \) g cm\(^{-2} \). Although we show the behavior of \( e_r \) starting at 1 AU, our disk-dominated excitation assumption is valid only for \( a_p \gtrsim a_B \) (the latter is shown by the green vertical line), which may be a problem for low \( e_0 \) (left panels in this figure) and has to be kept in mind. Because of that this figure would be accurate for all \( a_p \) only if the binary eccentricity were zero.

As a fiducial value of \( \dot{\omega}_d \) we take the planetesimal precession rate \( A|_{6\text{AU}} \) at \( a_p = 6 \) AU. This choice is almost entirely arbitrary, and is motivated only by the expectation of the reduced disk eccentricity beyond this radius. Precession of the binary is likely to be an insignificant driver of the disk precession outside of \( a_A \), where the disk gravity dominates. For our fiducial system parameters, \( A|_{6\text{AU}} = -1.2 \times 10^{-3} \text{yr}^{-1} \).

Figure 2.4 shows that \( e_r \) becomes independent of planetesimal size far from the binary, where \( e_c \) is low. This is due to apsidal alignment of planetesimals with the sizes shown in this Figure, as all of them are larger than the critical size \( d_c \) in the outer disk, simply because disk eccentricity is very low at large \( a_p \), see Equation (2.27).

Secular resonances

Equation (2.33) indicates that disk precession gives rise to two special locations in the disk. First, at the semi-major axis where

\[
A = \dot{\omega}_d, \tag{2.34}
\]

there is a secular resonance where \( e_c \to \infty \). This divergence happens because the relative precession between the planetesimal orbits and disk apsidal line vanishes, while the torque exerted by the non-axisymmetric component of the disk gravity is active. As a result,
eccentricity can grow without bound in the absence of gas drag. This resonance is an obvious generalization of the secular resonance discussed in §2.5.3.

Figure 2.1 illustrates the non-trivial behavior of $A$, as it is a combination of $A_d$, and $A_b$, which have opposite signs. We see that depending on the disk precession rate, there can be zero, one, or two secular resonances (2.34) associated with disk gravity.

If $\dot{\omega}_d = 0$ (non-precessing disk, yellow level in Figure 2.1), then there is only one resonance at the location where $A_d + A_b = 0$ (i.e. at 1.9 AU), and the situation is analogous to §2.5.3, see Figure 2.4A,B. As in Figure 2.3, the larger planetesimals are excited to higher eccentricities at the resonance because they are less damped by the gas drag (see Equation (2.32)).

Prograde precession gives rise to similar behavior, as shown in Figure 2.4C,D. This is not surprising since Figure 2.1 shows that prograde precession (green level) simply shifts the resonance location inwards, closer to the binary. This is indeed reflected in Figure 2.4C, D, where the resonance is now closer to the star.

As demonstrated in Figure 2.1, retrograde precession can either remove the resonance if it is very rapid, $|\dot{\omega}_d| > |\min(A)|$ (blue level), or give rise to two secular resonances as illustrated by the red level. Figure 2.4G,H illustrates the former possibility. It is obvious that $e_r$ does not exhibit sharp features in this case. The mild bump around 3 AU is due to the reduced $|A - \dot{\omega}_d|$ near the minimum of $A(a_p)$.

Finally, Figure 2.4E,F shows the case of slower retrograde precession, with two conspicuous secular resonances at $a_p \approx 2.2$ AU and $a_p = 6$ AU, in agreement with our expectations. It is clear that in this case $e_r$ can stay at a high level within an extended disk region between the two resonances, harming the prospects for planetesimal growth there.

Regions of low $e_r$

A second type of special location in the disk is possible when the disk precesses in a prograde sense, i.e. in the direction opposite to the precession of the planetesimal orbits. Equation (2.33) predicts that $e_c \to 0$ when $\dot{\omega}_d = A + B_d/e_d$. When this condition is fulfilled, relative
Figure 2.4: Run of relative planetesimal-gas eccentricity $e_r$ in the disk dominated excitation regime, calculated for eight different disk models. We have assumed the fiducial system parameters except where noted. All models on the right have $e_0 = 0.05$, while all models on the left have lowered disk eccentricity, $e_0 = 0.005$. Different panels correspond to different disk precession rates in units of $A|_{6AU}$ (also indicated in Figure 2.1) — the planetesimal precession rate at 6 AU: $\dot{\omega}_d = 0$ (A,B), $\dot{\omega}_d = -A|_{6AU}$ (C,D), $\dot{\omega}_d = A|_{6AU}$ (E,F), $\dot{\omega}_d = 2A|_{6AU}$ (G,H). Depending on the disk precession direction and rate curves of $e_r$ feature zero, one, or two secular resonances.
velocities of particles of any size with respect to gas (and, consequently, also w.r.t. each other) vanish at this location, naturally promoting growth. This cannot occur without disk precession, since \( B_d/e_d \) is always larger in magnitude than \( A \), except very close to the central binary where \( A \) has the same sign as \( B_d/e_d \).

Using equations (2.5) and (2.6), we see that \( e_c = 0 \) when \( \dot{\varpi}_d = A_b + A_d(1 + \psi_2/2\psi_1) \), and outside \( a_A \) the \( A_b \) term can be neglected. Since, by our assumption, \( \dot{\varpi}_d \) is constant with radius, but \( A_d < 0 \) is not, one finds that in general a prograde disk precession (\( \dot{\varpi}_d > 0 \)) will give rise to a radius in the disk where \( e_c = 0 \), since \( \psi_2/\psi_1 \approx -3 \).

This situation is clearly seen in Figure 2.4C,D, where disk precession in the direction opposite to \( A \) creates a region around \( a_p \approx 5 \) AU where \( e_r \to 0 \), independent of the disk eccentricity. This “valley of tranquility”, where relative planetesimal velocities are low, may represent a location where planetesimal growth is naturally promoted.

To summarize this analysis, disk precession can be either helpful or harmful to planet formation depending on its direction and magnitude, and on the location in the disk.

### 2.7 Planetary Dynamics Around a Precessing Binary

One important difference between the dynamical environments of P-type and S-type binary systems is that in the P-type systems, the precession rate of the binary itself can easily be comparable to or faster than the planetesimal precession rates for much of the disk (Rafikov 2013). It is therefore important to consider the precession of the central binary in calculating planetesimal eccentricities in situations where excitation due to the binary is important. To that effect, we discuss different mechanisms driving binary precession in Section 2.7.1 and Appendix 2.14. Based on that, we then explore the role of binary precession using a simple disk model in Section 2.7.2.
2.7.1 Binary Precession Rates

We consider four drivers of binary precession: (1) general relativistic precession, (2) the quadrupole due to tidal interaction between the two stars, (3) the quadrupole due to stellar rotation, and (4) the gravity of the circumbinary disk. In Appendix 2.14 we provide estimates of the magnitude of each contribution.

We find that precession due to the tidally induced quadrupole is dominant for massive stars that are close together, and disk-driven precession is dominant for less massive and more widely separated binaries. General relativistic precession and rotationally induced quadrupole precession are always subdominant to one or the other of tidal or disk precession for the Kepler systems.

Figure 2.5 shows our estimate of the binary precession rate as a function of $M_b$ and $a_b$. This assumes that $\mu = 1/3$, the apsidal motion constant $k_2 = 0.13$ (see Equations (2.45) and (2.46)) and the log of the surface gravity (shown in the work of Claret (2012) to be nearly constant over a wide mass range for 1 Myr old stars) is 3.66 in cgs units. We have taken these numbers from the pre-main-sequence stellar models of Claret (2012) for stars 1 Myr in age. We assume our fiducial disk model (Table 2.2) in calculating the contribution of disk gravity to binary precession.

We have additionally placed on this figure eight Kepler binary systems known to host circumbinary planets. The precession rates for these systems are only approximate, as they do not all have $\mu = 1/3$ assumed in our calculation, and there is a substantial change in stellar radii over the course of pre-main-sequence evolution. The contours correspond to locations in $a_b - M_b$ space where disk precession rate equals tidal precession rate, and where it exceeds the latter by a factor of 10. If the disk cavity were larger (i.e. $a_m$ were larger than $2a_b$), then the disk would drive substantially slower binary precession, see Appendix 2.14.

We see from Figure 2.5 that assuming precession to be disk-dominated is a good approximation for most of these systems, and particularly for Kepler 16, for which binary precession is dominated by the disk by a factor of hundreds over the tidal quadrupole pre-
Figure 2.5: Binary precession rate (yr$^{-1}$), as a function of $a_b$ and $M_b$. Thin white lines show the contours where disk precession equals tidal precession, and where it exceeds the latter by a factor of 10. Points correspond to known circumbinary systems listed in Table 2.1. We have assumed $\mu = 1/3$ in calculating the precession rate. See text for more details.
cession. Throughout the rest of the paper, we assume that binary precession is solely due to the disk.

Using Equations (2.5) and (2.47), with \( \psi_1 = -0.55 \), we find

\[
\frac{\dot{\varpi}_b}{A_d} = -0.15 \frac{a_p}{a_{in}}
\]  

(2.35)

for our disk model with \( p = 1.5 \). Equation (2.47) yields \( \dot{\varpi}_b \sim a_{in}^{-1} \) because we are assuming \( a_b/a_{in} = 1/2 \), and \( A_d \sim a_p^{-1} \) from Equation (2.5). Because \( A_d \) and \( A_b \) have different signs, \( |A| < |A_d| \). As a result, even inside of the radius where \( A_d = \dot{\varpi}_b \), the major contribution to the relative precession between the binary and the planetesimals may be coming from the binary precession — see Figure 2.1.

### 2.7.2 Axisymmetric Disk

As an application, here we consider planetesimal dynamics in the limiting case of an axisymmetric disk \( (e_0 = 0) \) with a central binary precessing at the rate \( \dot{\varpi}_b \). This section considers the same setup as in \cite{Rafikov2013}, but with the addition of gas drag.

Setting \( \varpi_b = \dot{\varpi}_b t \) and \( e_d \to 0 \), \( B_d \to 0 \) in Equations (2.18)-(2.19) we can easily solve them analytically. We find that the solution for \( e_p \) is identical to Equation (2.31) if we set \( e_d \to 0 \) and replace \( B_d \to B_b \), \( \dot{\varpi}_d \to \dot{\varpi}_b \) in the latter. In this solution, the forced eccentricity vector \( e_p \) rotates at the rate \( \dot{\varpi}_b \) and is fixed in the frame precessing with the binary.

The relative planetesimal-gas eccentricity is still given by Equation (2.32) but with \( \dot{\varpi}_d \) replaced by \( \dot{\varpi}_b \). And instead of Equation (2.33) we now have

\[
e_c = \left| \frac{B_b}{A - \dot{\varpi}_b} \right|.
\]  

(2.36)

Given that binary precession is prograde, \( \dot{\varpi}_b > 0 \), Figure 2.1 implies that there could be only one secular resonance associated with binary precession (illustrated by the orange level
calculated for our fiducial parameters of Kepler-16) and determined by the condition

$$A = \dot{\varpi}_b.$$  \hspace{1cm} (2.37)

It is located somewhat closer to the star than $a_A$. Given all that, it is clear that the plot of $e_r$ is going to be similar to Figure 2.4C,D. However the “valley of tranquility” is going to be absent, and the secular resonance would now correspond to the location where the condition (2.37) is satisfied.

In the case of very tight binaries ($a_b \lesssim 0.1$AU) binary precession is much faster than the planetesimal precession rate $A$ throughout the disk, thus strongly suppressing the excitation due to the binary. This effect was first noticed in (Rafikov, 2013). Even more modest precession rates, such as those calculated for Kepler-16 in this paper ($\dot{\varpi}_b = 2.6 \times 10^{-3}$ yr$^{-1}$) reduce the excitation effect of the binary by factors of a few. Looking at Figure 2.1 we see that in most of the disk outside $a_A$, the main contribution to the relative planetesimal-binary precession $A - \dot{\varpi}_b$ is due to the high value of $\dot{\varpi}_b$. Thus, binary precession effectively suppresses planetesimal eccentricity excitation due to its own non-Newtonian potential in the outer parts of the disk.

2.8 Dynamics With Both Binary and Disk Gravity and Both Binary and Disk Precession

To conclude our discussion of circumbinary planetesimal dynamics we provide a general description of the $e_p$ behavior in the most general situation when neither disk or binary gravity, nor disk or binary precession can be ignored, and quadratic gas drag damps planetesimal eccentricity. In this case the forced planetesimal eccentricity is no longer constant in time (even in some precessing frame) and its amplitude varies. As shown in Beaugé et al. (2010) and SR15, evolution of $e_p$ can be viewed as a superposition of the two precessions in the ec-
centricity space, resulting in a limit cycle behavior with $d_p$-dependent characteristics, which cannot be described analytically.

Despite this complication, important insights into the general problem can be gained by analyzing the general solution for the linear gas drag law described in Appendix 2.13. Examination of this solution shows that in general one should expect secular resonances of both types — given by Equations (2.34) and (2.37) — to exist in the disk. Given the discussion in §2.6.1 and 2.7.2 one expects up to three secular resonances to emerge. In particular, three resonances appear for slow retrograde disk precession, with two resonances being due to disk gravity (corresponding to the situation in Figure 2.4E,F) and one due to the binary precession (see Equation (2.37)). This makes planetesimal dynamics even more complex than before.

Nevertheless, the general features of the $e_r$ behavior outlined in §2.5.3 and 2.6.1 and shown Figures 2.3 and 2.4 remain in place: eccentricity reaches high values at resonances, with the larger increase of $e_r$ for bigger objects, less coupled to gas. At large separations disk gravity would still dominate and drive $e_p$ to size-independent behavior, see §2.5.2 and 2.6.1. Lower disk eccentricity $e_0$ would still result in lower planetesimal eccentricity, and so on.

This completes our discussion of secular planetesimal dynamics in circumbinary disks.

## 2.9 Collisional Outcomes and Growth of Planetesimals

To assess the prospects for planet formation in circumbinary systems, we must couple our understanding of the planetesimal dynamics outlined in previous sections with a prescription for the outcome of planetesimal collisions. We adopt that from RS15b, who based their calculation on the results of Stewart & Leinhardt (2009). In their framework the outcome depends on the masses of the planetesimals, an assumption about their internal strength, and their collision velocity, which is determined by the $e_p$ of each body involved in a collision.
For completeness, details of our collisional prescription are reproduced in Appendix C. We categorize collisions in three groups based on the mass of the largest surviving fragment.

- **Catastrophic** collisions leave a largest remnant containing no more than half the combined mass of the two colliding bodies.

- An **erosive** collision leaves a largest remnant smaller than the larger of the two incoming planetesimals.

- If the largest remnant is bigger than either of the two incoming bodies, we say that the collision leads to **growth**.

Figure 2.6 illustrates collisional outcomes in the space of sizes of colliding planetesimals $d_1$ and $d_2$ at two different locations in the circumbinary disk of the Kepler 16 system, using the disk-dominated excitation approximation without precession, and with fiducial disk parameters. White contours enclose regions leading to catastrophic disruption. Black contours enclose regions leading to erosion.

We see that collisions of bodies of exactly the same size lead to growth because collision velocities are small. Catastrophic disruption occurs only closer to the binary (panel A), since $e_c$ is not high enough further out (panel B) for catastrophic collisions to occur. We see that similar in size, but not exactly equal planetesimals tend to undergo catastrophic collisions, because their $e_p$ are different enough to result in significantly energetic collisions. However, even very unequal mass ratio collisions can lead to erosion.

The white star in each plot corresponds to the size $d_c$. We notice that this is near the region of catastrophic destruction in the plot. The lobes are biased towards $d_p > d_c$ because $d_c$ is smaller than the size of planetesimal with the lowest critical velocity for destruction, which is about 100 m, see Figure 2 of RS15b. Because the destruction region is near $d_c$, having a low value of $d_c$ means that planets can form starting from a population of smaller planetesimals. In fact, Figure 2.6 suggests that if the planetesimal population were to contain
only the objects with sizes $\gtrsim (30-50)d_c$, this population would have been completely immune to both catastrophic disruption and erosion.
In this paper, we consider catastrophic disruption to be the only obstacle to planetesimal growth. For a number of reasons discussed further in Section 2.10.1, we do not expect erosion to play a determining role in whether planetesimal growth occurs.

In the following sections we examine the prospects for collisional growth in each of the dynamical regimes discussed in Sections 2.5 - 2.7, which provide different prescriptions for $e_c$ and $d_c$. Using these, we calculate the smallest planetesimal size $d_{\text{min}}$ such that objects larger than $d_{\text{min}}$ do not suffer catastrophic disruption. This size is indicated in panel A of Figure 2.6. This is the smallest planetesimal size that we can start from and grow larger objects without ever encountering catastrophic disruption. Because of our neglect of erosion, overall planetesimal growth requires only the existence of objects larger than $d_{\text{min}}$ in the planetesimal populations.

If this size is under 10 m, we conclude that planetesimal growth via collisional agglomeration is easy under those environmental conditions. The choice of 10 m is somewhat arbitrary, but we note that changing from 10 to 100 m in our plots would make very little difference to our conclusions. If $d_{\text{min}} > 10$ m, $d_{\text{min}}$ provides an estimate of the minimum size of primordial planetesimals necessary to form planets in that environment. Unless otherwise specified, we are assuming the material properties appropriate for the “strong planetesimals” of Stewart & Leinhardt (2009).

### 2.9.1 Planetesimal Growth in the Disk-Dominated Excitation Regime

We first consider the collision outcomes in the outer part of the disk where excitation from the binary is unimportant compared with excitation from the disk, and use the results on planetesimal dynamics from Section 2.6.

Figure 2.7 shows the size $d_{\text{min}}$, above which the growth is unimpeded by catastrophic disruption, as a function of semi-major axis and disk eccentricity for different assumptions.
Figure 2.7: Minimum planetesimal size $d_{\text{min}}$ safe from catastrophic disruption in the disk dominated excitation regime (described in §2.6). The calculation was done for our fiducial disk around Kepler 16, see Table 2.2. Different panels correspond to different assumptions about the disk precession rate and direction. The whitened out region in the bottom left is where $|B_b| > |B_d|$ and we are not justified in using the disk dominated approximation. The dashed red line is drawn at the value of $e_0$ corresponding to the forced eccentricity of a free particle in the binary potential, see Equation (2.10).
about the disk precession. We have used our fiducial system parameters listed in Table 2.2. Collision outcomes have been calculated using Equations (2.33) and (2.27) for $e_c$ and $d_c$.

One can see that in general, the outer region of the disk is favorable for growth, even starting from small ($d_p \sim 10 \text{ m}$) planetesimals. This is true for two reasons. First, the local disk eccentricity is small at large $a_p$, which leads to low values of $e_c$. Second, far from the binary $d_c \propto a_p^{-13/4}$, and becomes quite low (around a meter) at around 5 AU. This means that all the planetesimals have $(A - \dot{\varpi}_d)\tau_d \gg 1$ and $e_r$ very near $e_c$.

**Effects of Disk Precession**

Looking at Figure 2.7, we see that in the absence of disk precession, the outer region of the disk is friendly to planetesimal growth beyond 3-4 AU (for all planetesimal sizes down to 10 m) depending on $e_0$. Starting with planetesimal sizes of a few km brings the inner edge of the growth-friendly region within 3 AU, even for disks with $e_0 = 0.05$ (twice the free-particle eccentricity).

In the middle panel the growth region expands because prograde disk precession ($-A|_{6\text{AU}}$ is positive) dramatically lowers $e_c$. In this case, the disk precesses at $\dot{\varpi}_d = A|_{6\text{AU}}$ ($A|_{6\text{AU}}$ is the planetesimal precession rate at 6 AU), and we find nearly the whole outer disk ($a_p > 3\text{AU}$) to be conducive to planetesimal coagulation even for highly eccentric disks. This is because of the valley of low $e_c$ in a prograde precessing disk, discussed in Section 2.6.1 and shown in Figure 2.4C.D.

On the other hand, retrograde disk precession may give rise to a second secular resonance, which is very damaging to planetesimal growth, see Figure 2.7C. This resonance makes conditions in the outer part of the disk, around 6 AU, much more hostile to planetesimal growth than in the absence of disk precession. In addition, comparing panels A and C, we see that the inner resonance in panel C is also moved slightly outwards because of the disk precession. As a result, planetesimals within a broad radial interval of the disk (1-8 AU) end up being strongly dynamically excited.
Effect of Planetesimal Strength

Next, we consider changing the material properties of the planetesimals, i.e. changing the $Q_{RD}^*$ term in Equation (2.48). Stewart & Leinhardt (2009) provide two prescriptions for $Q_{RD}^*$ depending on whether planetesimals are assumed to be solid rocks (strong planetesimals), or rubble piles (weak planetesimals). Figure 2.8 shows the same dynamical environment as Figure 2.7, but considers weak planetesimals. This does not make a big difference except for the sub-kilometer sized planetesimals, as there is not a big difference between weak and strong planetesimals in the gravity-dominated regime (for $d_p \gtrsim 1$ km). Comparing Figures, 2.7 and 2.8 we see differences of several AU in the extent of the region where 10 m sized planetesimals are vulnerable to catastrophic disruption, but insignificant differences in the extent of the region where km-sized planetesimals are subject to destruction.

Disk Mass

Another model parameter which we vary is the surface density at 1 AU. Unlike the case of tight S-type systems, there is little reason to believe that circumbinary disks contain less mass than their counterparts around single stars. In fact, if anything, there seems to be evidence for more massive disks in circumbinary systems (Harris et al., 2012).

In Figure 2.9 we consider both a denser and less dense disk than our fiducial system, with $\dot{\Sigma}_d = 0$, again using the dynamics discussed in Section 2.6. We have adjusted the scale on the y-axis of Figure 2.9 in both subplots so that the bottom of the plot is near $a_B$, where excitation switches to being dominated by the disk, see Equation (2.13). Because $a_B$ depends on disk mass, the scales on the y-axes are not the same in the two panels. We have whitened out the area in the plot where $|B_b| > |B_d|$, as we do not have an analytic solution for $e_c$ in that case.

The main effect of changing the disk mass is the variation in location of the secular resonance where $A_d + A_b = 0$: it moves out for lower $\Sigma_0$ (and $M_d$). For this reason lowering the disk mass is quite unfavorable for planet formation in the outer part of the disk. There
Figure 2.8: Same as Figure 2.7 except that we are using the collisional prescription for weak aggregates (Stewart & Leinhardt 2009) instead of that for solid rocks, see Appendix C.
Figure 2.9: Same as the top panel of Figure 2.7 but now for two different disk masses, resulting in different surface densities at 1 AU $\Sigma_0$, indicated on panels. We use the system parameters in Table 2.2 except for the disk density.
may however be a region favorable to growth interior to the resonance, a possibility we explore further in Section 2.9.3.

For the very massive disk with $\Sigma_0 = 3 \times 10^4 \text{ g cm}^{-2}$, we find that the dynamics look similar to the top panel of Figure 2.7 (again accounting for the change in scale on the $y$-axis). This is because in the disk dominated excitation regime, $d_c$ and $e_c$ are independent of disk mass, see discussion after Equation (2.27). The only difference we expect from increasing the disk mass is that the secular resonance moves inward. This is why the differences between Figures 2.7 and 2.9 become more striking as one moves inwards in semi-major axis.

### 2.9.2 Collisional Outcomes in an Axisymmetric Disk with Binary Precession

As found in Rafikov (2013), a massive axisymmetric disk is very helpful for reducing planetesimal collision velocities, both because of the enhanced planetesimal precession and because of the induced binary precession. Not surprisingly, when including gas drag, this result remains valid, as we show now.

In Figure 2.10 we show $d_{\text{min}}$ in the axisymmetric disk approximation as a function of $\Sigma_0$ and $a_p$. We are now using the dynamics discussed in Section 2.7, including the effects of non-zero binary precession, in particular Equations (2.35) and (2.36). We consider both strong (top) and weak (bottom) planetesimals.

Both panels exhibit similar structure, since the planetesimal dynamics are the same, independent of their material properties. We see the secular resonance given by Equation (2.37) running diagonally across both panels. Its appearance is different from Figures 2.7, 2.8 and 2.9 because those figures held disk mass constant.

Exterior to the resonance, where $A$ is disk-dominated, $d_c$ is independent of disk mass, see Equation (2.27). However a more massive axisymmetric disk lowers $e_c$ by increasing the rate of relative precession between the binary and planetesimal orbits, without adding to the excitation. We see that a high $\Sigma_0$ allows unimpeded coagulation as close as 2.5 AU even for
Figure 2.10: Size $d_{\text{min}}$, as a function of $\Sigma_0$ and $a_p$ for an axisymmetric disk with $p = 1.5$ in the Kepler 16 system. Planetesimal collisional velocities are calculated using the results of §2.7. The two panels correspond to the strong and weak planetesimals discussed in Stewart & Leinhardt (2009). Binary precession rate $\dot{\varpi}_b$ is given as a function of disk mass ($\Sigma_0$) by Equation (2.47).

Weak planetesimals. A low disk density is detrimental to planetesimal growth at large radii because the secular resonance moves out.

Interior to the secular resonance, we see that a massive disk is actually harmful to the survival of smaller planetesimals. Indeed, fixing e.g. $a_p = 1$ AU and increasing $\Sigma_0$ leads to higher $d_{\text{min}}$. This is because in the inner disk $A$ is dominated by the binary, $A \approx A_b$, so that
a more massive disk increases both the value of $d_c$ (because $\Sigma_d$ increases in Equation (2.27)), and $e_c$ (by moving the resonance inwards).

In the opposite limit of a low mass disk ($\Sigma_0 = 300 \, \text{g cm}^{-2}$ corresponding to a disk mass of a few Jupiter masses), in situ planetesimal growth at $a_p \approx 1\text{AU}$ is possible with km-sized initial planetesimals. A light disk ensures that the secular resonance is at several AU, and therefore not playing a role in the dynamics inside of an AU. It also means that $d_c$ is still substantially smaller than a kilometer, even at an AU separation. For example, for an axisymmetric disk in our fiducial system with $\Sigma_0 = 300 \, \text{g cm}^{-2}$, the critical size is just 8 m at 1 AU.

### 2.9.3 Planetesimal Growth Near the Star in the Limit of Short Stopping Times

Planetesimal dynamics are complicated near the star because both gas drag and binary gravity are important, and the disk and binary do not precess at the same rate, causing planetesimal eccentricities to be time dependent, see §2.8. Nevertheless, we still find analytic solutions for planetesimal orbits in two limits. One is the axisymmetric disk discussed in the previous section. The other is the limit of no disk and binary precession discussed in Section 2.5.

Planetesimals small enough to have stopping times short compared with the binary and disk precession times ($\max(\dot{\varpi}_b, \dot{\varpi}_d) \tau_d \ll 1$) should have their eccentricities approximately described by Equations (2.20)-(2.24) with $\varpi_b, \varpi_d$ given by the instantaneous orientation of the binary and the disk. We imagine $|{\dot{\varpi}}_d|$ to be smaller than $|{\dot{\varpi}}_b|$, so we are only requiring $\dot{\varpi}_b \tau_d \ll 1$, when assessing the validity of the approximation that stopping times are rapid.

Planetesimal eccentricity (2.20) obtained in Section 2.5 is a function of the mutual disk-binary apsidal orientation $\varpi_d - \varpi_b$. To be conservative in our estimate of the planetesimal destruction region, here we calculate the maximum value of $e_c$ as a function of $\varpi_d - \varpi_b$ in
Figure 2.11: Size $d_{\text{min}}$ in the inner part of the disk, adopting approximation of no disk and binary precession (\S 2.5), which is valid for small bodies. We calculate $e_c$ by maximizing Equation (2.26) over $\varpi_d - \varpi_b$. White areas correspond to locations where Equation (2.26) is no longer valid because $\dot{\varpi}_b \tau_d > 1$, for the largest size of planetesimal which is destroyed. System parameters are taken from Table 2.2 except for the disk density.
Equation (2.26), thus considering planetesimal dynamics in the least favorable part of the binary orbit.

We use these assumptions to generate Figure 2.11 which is similar to Figure 2.7. We display as white the region where \( \dot{\omega}_b \tau_d < 1 \) for planetesimals of size \( d_{\text{min}} \). This is where we expect the approximation of a slowly-precessing binary to break down.

We see a substantial difference of the outcomes depending on the density of the disk. Because the gravitational perturbations are dominated by the binary (i.e. \( A \) is independent of \( \Sigma \)), the critical size \( d_c \) is smaller in lighter disks, see Equation (2.27), leading to even km-sized bodies having more aligned orbits. Additionally, the secular resonance moves outwards for lower \( \Sigma_0 \), leading to lower \( e_c \) within an AU.

As a result, in the case of a low-density, low-eccentricity disk, we see that it is possible to have planetesimals greater than a few km in size grow undisturbed by catastrophic disruption. This is a simpler scenario for planetesimal growth than the one described in Meschiari (2014) which relies on a pressure maximum to trap small dust to enable planetesimal growth. One caveat of the in-situ growth scenario with a low-\( \Sigma_0 \) disk is that it may be difficult to grow Saturn-size circumbinary planets (such as Kepler 16b) because of the short supply of mass in such a disk.

For the denser disks (higher \( \Sigma_0 \)), catastrophic disruption is inevitable unless nature creates initial planetesimals on the order of tens of km in size (Johansen et al., 2012), see §2.11.2. These general conclusions are similar to those obtained in Section 2.9.2.

Although our analytic results are formally accurate only for \( d_p \lesssim \) several km, there is hardly a reason to believe that as they grow larger, planetesimals will become more vulnerable to destruction. Indeed, larger bodies are more resistant to higher velocity collisions. Also, we do not expect that they will have collision velocities dramatically higher than their lower-\( d_p \) predecessors. In fact, it seems much more likely that large bodies will have smaller collision velocities. This is because close to the binary, \( |A| \approx |A_b| \gg |\dot{\omega}_b| \), so \( d_c \) (for which \( |A| \tau_d \sim 1 \),
if we ignore $e_\phi$ in Equation (2.28)) is much smaller than the size of bodies for which $\tau_d \tilde{\omega}_b \sim 1$. Thus, larger bodies should be more resistant to destruction, see Figure 2.6.

### 2.10 Discussion

Here we will revisit some of the assumptions that went into the model described in the previous sections, and explore what happens when they are relaxed.

#### 2.10.1 Erosion

In the preceding sections, we have assumed that planetesimal growth is inevitable in the absence of catastrophic disruption events. This is not necessarily the case, as frequent collisions with smaller objects can still lead to mass loss, even though none of them are severe enough to destroy the planetesimal. This is the erosion regime defined in Section 2.9.

In RS15b (see their §4), we determined that in a collision between planetesimals of mass $m_1$ and $m_2$, the critical velocity for erosion is independent of $m_2$ in the limit that $m_2 \ll m_1$. Erosion can therefore be deleterious in regions where planetesimals are safe from catastrophic disruption because a low value of $d_c$ does not have as large of a protective influence.

If the characteristic eccentricity $e_c$ is high enough for collisions to be erosive in the limit of vanishing collision partner mass, then erosion might be expected to inhibit planetesimal growth. Looking at Figure 2.6 we see that erosion is apparently ubiquitous, even in a regime where strong planetesimals do not suffer catastrophic disruption. Therefore, if a substantial fraction of the mass that a planetesimal encounters is in small bodies, it might have difficulty growing. This outcome could be avoided if the majority of the mass of solid material in the disk is contained in objects with sizes larger than $d_{\text{min}}$, or if collision rates between large and small bodies are suppressed.
In practice, we do not expect erosion to be the major obstacle to planetesimal growth because small bodies are likely to be rapidly flushed out of the system due to gas drag in the slightly sub-Keplerian disk.

Indeed, using the results of [Adachi et al. (1976)], and ignoring the relative particle-gas eccentricity, we estimate an in-spiral timescale of

$$
\tau_m = \frac{a_p}{d_p} = \frac{256\sqrt{2}\pi}{507C_dE(\sqrt{3}/2)} \rho_p \frac{d_p}{\Sigma_d} \left( \frac{a_p}{h} \right)^3
$$

$$
\approx 6 \times 10^4 \text{yr} \times \frac{\rho_p}{3 \text{ g cm}^{-3}} \frac{d_p}{1.3 \text{m}} \frac{3,000 \text{ g cm}^{-2}}{\Sigma_0} \left( \frac{M_b}{0.89M_{\odot}} \right)^{1.5} \left( \frac{a_p}{0.5} \right)^{a/4}.
$$

In this estimate we have taken for $d_p$ the critical size of $d_c \approx 1.3 \text{ m}$ given by Equation (2.27) at 5 AU. Particles with sizes above $d_c$ are less likely to erode larger bodies due to their aligned orbits. This is a lower bound, as non-zero gas-particle eccentricity will only accelerate in-spiral.

Furthermore, as discussed in [Weidenschilling (1977)], for these small bodies moving slowly relative to the gas, the drag is actually not in a quadratic regime, and is stronger than predicted by our assumed quadratic drag law, making the in-spiral more rapid. The particles which in-spiral most rapidly are those for which $n_p \tau_d \approx 1$. Using our quadratic drag law, $n_p \tau_d = 1$ at 5 AU for particles with sizes of about a centimeter. The fact that we are doing this estimate at 5 AU, instead of the conventional 1 AU, leads to meter sized particles lasting substantially longer than the 100 yr timescale generally discussed in the literature [Weidenschilling, 1977].

The significance of the estimate (2.38) is that the disk cannot store most of its mass in bodies smaller than $d_c$ because $\tau_m$ is considerably shorter than the several Myr disk lifetime. As a result, the disk would rapidly lose all of its solid material by radial drift towards the central binary. By assuming most of the mass in the disk to be in planetesimals larger than several meters initially, we thus make erosion by smaller bodies to be a subdominant effect.
In addition, very small bodies with stopping times shorter than the dynamical time $n_p^{-1}$ orbit with sub-Keplerian velocities due to their strong coupling to gas, which experiences radial pressure support. These objects would impact larger bodies at speeds of several tens of m s$^{-1}$, even in the absence of any secular excitation. The same would happen in protoplanetary disks around single stars, which according to recent statistics of exoplanets, have no difficulty of forming planets. Thus, erosion by very small objects can be overcome in circumbinary systems in the same way as this happens around single stars.

2.10.2 Growth via “Lucky” Bodies

It is likely that planetesimal growth can occur even in presence of some catastrophic disruption, since the detrimental effect of these can be offset by other favorable collisions which enable growth. This balance of the mass loss and gain has been explored for the growth of centimeter-sized bodies by Windmark et al. (2012) and Garaud et al. (2013). These authors consider a distribution of encounter velocities and collision outcomes, rather than assuming all encounters to be at the mean velocity. By using such statistical approach they find over an order of magnitude change in the size of the largest particles that can grow in their coagulation simulations. Thus, statistical nature of coagulation may be an important factor of planetesimal growth, which we did not account for here.

In the km-sized planetesimal regime, bodies generally become both more resistant to collision, and experience lower velocity collisions with similar sized objects as they grow larger because their size moves further from $d_c$, see Equation (2.51). It therefore seems likely that including the full distribution of collision velocities will make an even larger impact in this scenario than it did for the growth of centimeter sized grains, since a km-sized body that grows via a few lucky collisions becomes harder to destroy. We leave the detailed exploration of the statistical growth of planetesimals in circumbinary systems to future work.
2.10.3 Limits of the Apsidally Aligned Regime

In the previous sections we have assumed that free eccentricity of planetesimals has been completely damped by gas drag. As the free eccentricity goes away on the timescale $\tau_d$, we need to demonstrate that $\tau_d$ is shorter than the characteristic planet formation timescale.

At a given semi-major axis in the disk, we can use Equation (2.17) to solve for the critical damping size $d_{\text{damp}}$ at which the free eccentricity damping time $\tau_d$ is equal to some characteristic time $\tau$ (not to be confused with the critical size $d_c$ for which the damping time is roughly the precession time $A^{-1}$). If we set $\tau$ equal to the expected lifetime of the circumbinary disks (Myrs), then all objects with $d_p > d_{\text{damp}}$ will not have a chance to settle to their quasi-equilibrium forced eccentricities during the disk evolution. This would violate our basic assumption of damped $e_{\text{free}}$ stated in §2.4.4 and would make planetesimal growth more complicated.

To demonstrate this last point, we repeat the calculation illustrated in Figure 2.10, however, now we do not assume planetesimal orbits to be aligned according to their forced eccentricity values. Instead, we set $e_{12}$ to be given by $e_c$ instead of by Equation (2.50). In the approximation where planetesimals decouple instantly from the gas with eccentricity equal to $e_d$, their eccentricity vector $e_p$ circulates around the forced eccentricity with a magnitude $|e_d - e_{\text{forced}}|$. As shown in RS15a, $|e_d - e_{\text{forced}}|$ is just $e_c$. Therefore a typical relative eccentricity $e_{12}$ between planetesimals in this population will be of order $e_c$. This is the approximation explored in Rafikov (2013).

Results of such calculation are shown in Figure 2.12. Comparing Figures 2.10 and 2.12 shows that the alignment assumption makes a substantial difference, with the zone of catastrophic disruption significantly extended in Figure 2.12 compared with Figure 2.10 for a given planetesimal size. Thus, lack of free eccentricity damping resulting in apsidal misalignment of planetesimals should have a deleterious effect on planetesimal growth (Bromley & Kenyon 2015).
Figure 2.12: $d_{\text{min}}$ for the same dynamical environment as Figure 2.10, but assuming planetesimal orbits to be unaligned so that $e_{12} = e_c$. Comparison of these two figures shows that the apsidal alignment is a very important effect, greatly facilitating planetesimal growth.
We expect protoplanetary disks to last for a few Myr (Haisch et al., 2001). This motivates us to set a characteristic timescale in the determination of the critical size $d_{\text{damp}}$ to be $\sim$ Myr. Figure 2.13 shows $d_{\text{damp}}$ for $\tau = 3$ Myr as a function of $a_p$ for several sets of disk parameters. This size is calculated numerically from Equations (2.1), (2.5), (2.8), (2.16), (2.27), (2.28), (2.33), (2.35), and (2.36).

We see that generally within 10 AU, the damping time is shorter than the disk lifetime for kilometer sized planetesimals, depending, of course, on the density of the disk. Although outside of 5 AU, we may be concerned about the alignment of 10 km planetesimals, at these radii, such large planetesimals are generally well above the mass threshold for collisional growth (see §2.7.2). This suggests that we are justified in considering the planetesimal orbits to be aligned. It is also worth noting that a higher value of $e_c$ leads to faster apsidal alignment, so the environments with high $e_c$, which present the highest danger to planetesimal coagulation are the same environments which are helped most by the apsidal alignment.

Collisions between planetesimals may also cause misalignment. However in the regions of the disk with long alignment timescales, planetesimals greater than $\sim 10$ m in size possess nearly apsidally aligned equilibrium orbits, so collisions between them are unlikely to lead to much misalignment.

2.10.4 Disk Density Structure

One may wonder how our results would change if the slopes of the $\Sigma_d(a_p)$ and $e_d(a_p)$ dependencies were varied. Observational evidence for the values of $p$ and $q$ is rather scant and is based predominantly in the sub-mm dust continuum observations on scales (tens of AU) much larger than the ones considered in this work\(^2\) (Andrews et al., 2009). However, the total masses of the circumbinary disks are found to be in the range of several percent of $M_\odot$ (Harris et al., 2012), in agreement with our estimate (2.2).\(^2\)

\(^2\)ALMA may be close to marginally resolving these scales in the future.
Figure 2.13: Maximum size of planetesimal in the disk dominated excitation regime (where $e_c$ is given by Equation (2.30)) such that $\tau_d$ is less than 3 Myr. We have assumed fiducial system parameters except where noted.
We have picked our fiducial value of $\Sigma_0$ to match an overall disk mass of a few percent of $M_\odot$, believing this to be better constrained than the surface density $\Sigma_0$ at 1 AU. If we change $p$ while keeping the overall mass of the disk constant, we see that $\Sigma_0$ would vary with $p$ as

$$\Sigma_0 \propto (2 - p) \left( \frac{a_{\text{out}}}{a_0} \right)^p.$$  \hspace{1cm} (2.39)

Equations (2.5) and (2.6) show that $\Sigma_d$ and $\psi_1$ are the only disk properties that determine $A_d$. The coefficients $\psi_1$ and $\psi_2$ in Equations (2.5) and (2.6) change by less than a factor of 2 between $p = 1.5$ and $p = 0.5$, so we see that at a given radius, changing $p$ at constant disk mass has about the same effect as changing $\Sigma_0$, as was done in Section 2.9.1. In other words, because $\psi_1$ and $\psi_2$ vary so weakly with $p$ and $q$, what matters most for the disk disturbing function is the local value of the surface density, not the distribution in the whole disk. And it was shown in SR15 that the edges of the disk are unimportant in determining the disturbing function for the values of $p$ and $q$ that we consider in this paper.

### 2.10.5 Non-Secular Terms in the Disturbing Function

In this study, we ignore the possibility of mean-motion resonances between the binary and the planetesimals. Previously, Meschiari (2014) found in simulations of the Kepler 16 system that planetesimals got trapped at the 5:1 resonance. Simulating the S-type $\gamma$ Cephei system, Leiva et al. (2013) find planetesimals getting trapped in first order resonances as high as 16:1. While it is not clear if these results would hold far from the binary where gravity of a massive eccentric disk dominates, the mean motion resonances may still play an important role in the dynamics inside of 1 AU.

It was also suggested in Paardekooper et al. (2012) that short-period terms in the disturbing function, varying with the binary period or the planetesimal orbital period will add additional eccentricity. These seem unlikely to mis-align aligned orbits: short period terms should affect planetesimals of all sizes equally, not changing their relative velocities,
as they act on a much shorter time-scale ($\sim n_p^{-1}$) than the gas drag damping time $\tau_d$ for the sizes we are concerned with. It was furthermore shown both analytically and in simulation by Bromley & Kenyon (2015) that in the Kepler 16 system these terms only induce absolute eccentricity variations of about a percent at 0.7 AU (relative variation should be much smaller).

2.11 Comparison with Previous Work on Planet Formation in Binaries

Here we compare our findings with existing results on planet formation in both circumbinary and circumstellar systems and put them in broader context.

2.11.1 Previous Work on Circumbinary Planet Formation

Despite the different setup and the range of physical effects taken into account in this work, many of our conclusions are similar to those reached in a number of previous studies of planet formation in circumbinary systems.

Paardekooper et al. (2012) and Meschiari (2012) numerically explored the interactions between swarms of planetesimals embedded in an axisymmetric gas disk around a central binary. They included the gravitational perturbations from the binary, and gas drag from the disk, but not the gravity of the disk, which as we know now (Rafikov 2013, SR15, RS15a) provides the most important gravitational effect outside of a few AU. They both found it difficult to explain in-situ accretion at $\lesssim 1$ AU separation without invoking initial planetesimals greater than 10 km in size.

Meschiari (2012) concluded that formation outside of 4 AU was possible starting from km-sized planetesimals, which is only slightly less optimistic than our conclusions for the axisymmetric disk of a similar mass. Our inclusion of disk gravity in this paper pushes the boundary of the coagulation region only a little bit inward because we find that the secular
resonance, emerging when we account for disk gravity, makes conditions in the disk around
2 AU unfavorable for planetesimal growth.

Meschiari (2014) did include disk gravity in some of his simulations and found prominent
secular resonance predicted in Rafikov (2013). However, he was primarily interested in in-
situ growth of circumbinary planets within 1 AU and did not explore planetesimal growth at
several AU in as much detail as we do in §2.9.1. At \( a_p \lesssim 1 \) AU Meschiari (2014) finds that km-
sized planetesimals can grow by accretion of small collisional debris. By carefully examining
planetesimal dynamics in §2.9.3 we find that growth starting with km-sized objects is possible
even with more standard collisional scenario not involving accretion of small particles, as long
as the disk density is not very high and the characteristic planetesimal size \( d_c \) is small.

Simulations of Marzari et al. (2013) also include the gravitational effect of a non-
axisymmetric disk. They calculated the disk structure in 2D geometry, and found complex
behavior for disk surface density and eccentricity with radius: their disk maintains substan-
tial \( e_d \) (several percent) out to nearly 10 AU. This eccentricity profile looks very different
(much higher) than found in simulations of Meschiari (2014), Pelupessy & Portegies Zwart
(2013), and may reflect transient behavior. As a result, Marzari et al. (2013) found it
difficult to grow even 25 km planetesimals out to 10 AU because of eccentricity excitation
due to disk gravity. This is at odds with our findings simply because we think that such
eccentricity profiles are unlikely and assume \( e_d \) to decay with radius. Another reason for
the discrepancy is the duration of their runs (\( \sim 10^4 \) yr), which is likely shorter that \( \tau_d \)
for the 5-25 km planetesimals they are considering. This leaves significant undamped free
eccentricity and results in large collisional velocities even among planetesimals of the same
size.

Bromley & Kenyon (2015) performed an analytic study of planetesimal dynamics, and
showed in particular that short-period terms in the disturbing function do not result in
increased collision velocities of planetesimals, in agreement with the argument of Rafikov
(2013). However, their prime focus was on pointing out the existence of a set of non-crossing,
aligned orbits of planetesimals, which they found explicitly neglecting gas drag. They argued that such orbits would allow planetesimals to collide with very low relative speeds (as low as in axisymmetric disks around single stars) and grow efficiently.

Based on our results, we interpret these trajectories as orbits with fully damped free eccentricity and only forced eccentricity remaining, which by itself requires some effective damping process (most likely gas drag) and works only for objects with $d_p \lesssim 10$ km at several AU, see §2.10.3. For such objects gas drag would in fact make forced $e_p$ size-dependent (§2.5, 2.6), leading to substantial collisional velocities for planetesimals of different sizes even when all free eccentricity is damped, as clearly demonstrated in our §2.9 (e.g. see Figure 2.6). This dynamical size segregation was previously broadly discussed in the context of planetesimal growth in S-type systems (e.g. Thébault et al. (2008), RS15a). Additionally, when one considers disk gravity, secular resonances arise, leading to orbit crossing even assuming apsidally aligned orbits, due to the rapid radial variation of $e_p$.

2.11.2 Large Initial Planetesimals

It is possible that some mechanism other than collisional agglomeration produces a population of large initial planetesimals resistant to collisional destruction by virtue of their size. It has been proposed (Goodman & Pindor, 2000; Youdin & Goodman, 2005) that coupling between solid material and the disk could produce an instability that would lead to over-dense rings of material which would collapse to form planetesimals. This so-called streaming instability has been explored numerically in Johansen et al. (2012), and was found to lead to rapid formation of planetesimals several hundreds of km in size.

The size distribution of minor bodies in our Solar System provides clues as to the size of the initial planetesimals, however studies of this have come to conflicting results. Morbidelli et al. (2009) find evidence for 100 -1000 km initial planetesimals in the current size distribution of asteroids. However, Weidenschilling (2011), using a different accretion model, is able to reproduce the current size distribution starting from $< 0.1$ km planetesimals. Addi-
tionally, Schlichting et al. (2013) studied the size distribution of objects in the Kuiper belt, and came to the conclusion that initial planetesimals on the order of a km in size provide the best fit.

Our present work shows that at least in circumbinary systems, rapid formation of large \((d_p \sim 10^2 \text{ km})\) planetesimals is not necessary and planetesimal growth is possible starting from km-sized (or even smaller, at large separations) bodies. Thus, at present, circumbinary exoplanets do not provide a strong argument in favor of streaming instability being the dominant planetesimal formation mechanism. It is also unclear whether the streaming instability would function in the perturbed circumbinary environment.

### 2.11.3 Differences Between P-type and S-type Binaries

This paper has a lot in common with RS15a,b. In this section we highlight some differences of the physics that goes into understanding dynamics in P-type systems.

In S-type systems, there is evidence (Müller & Kley, 2012) that massive disks end up apsidally aligned with the binary star. This alignment gives rise to a dynamically cold region in the protoplanetary disks of the S-type systems discovered in RS15a, which cannot exist in the circumbinary configuration. Circumbinary disks can in fact contain a dynamically cold region, but only if they are precessing in the prograde sense (see panels C and D of Figure 2.4).

The precession of the central binary cannot be ignored in P-type systems. This helps to lower the planetesimal eccentricity excitation due to the binary. This effect is absent in S-type systems.

Circumbinary disks can be much more extended than circumprimary ones, simply because they are not tidally truncated by the companion on the outside. For that reason planet formation can (and is likely to) occur at large radii (with subsequent inward migration of grown planets). We see from Equation (2.27) that everything else being equal, the critical size \(d_c\) at which collision velocities between similar sized planetesimals are highest scales as...
\( d_c \sim a_p^{-13/4}. \) The low value of \( d_c \) in the planet-forming region leads to the conclusion that all bodies km-sized or larger are very nearly apsidally aligned as \( d_p \gg d_c \) for all of them, see Equation (2.51). Additionally, the Keplerian velocity is smaller at higher separation, leading to further reduction of collision velocities. This permits planetesimal coagulation starting from relatively small initial sizes, 10 - 100 m, even in regions of higher \( e_c \) than is permitted in the S-type systems where massive planets reside at 1-2.5 AU.

Another consequence of the low density environment at large separation in circumbinary systems is that for a substantial fraction of the likely planet-forming part of the disk, there is some concern that alignment with the gas may not occur within the disk lifetime, see Section 2.10.3. This was not an issue for the S-type systems because their compact protoplanetary disks were likely much denser around 1-2 AU, effectively damping free eccentricity and aligning planetesimal orbits.

### 2.12 Summary

We studied planetesimal dynamics in circumbinary disks with the goal of understanding the conditions leading to planetesimal growth, which is a natural step towards formation of circumbinary planets such as Kepler-16. Our study simultaneously considers (1) gas drag (including non-trivial radial pressure support in the gaseous disk), (2) gravity from the eccentric precessing binary, and (3) gravity from the eccentric precessing disk. We found analytical solutions for planetesimal eccentricity behavior in many important limits.

We estimate the precession rate of the central binary and find binary precession to play a non-trivial role in the determination of planetesimal dynamics. We believe erosion by small bodies to not be a major hindrance to growth, as such objects should spiral in towards the central binary on timecales of tens of thousands of years. Based on our analytical results, we make the following conclusions:
• We find disk gravity to play the dominant role in planetesimal dynamics outside of several AU. Secular resonances (up to three for some choices of the disk and binary precession rates) significantly complicate planetesimal dynamics at \( a_p \sim \text{several AU} \).

• Apsidal precession of the central binary reduces the direct effect of binary gravity on planetesimal orbits. For many circumbinary systems discovered by \textit{Kepler} precession of the binary is dominated by the gravity of the protoplanetary disk.

• If planet formation is precluded only by catastrophic disruption, then we find that forming Kepler-16 and similar planets in situ requires that initial planetesimal sizes were large and disk masses were small. For example, even in the most favorable case of an axisymmetric disk with a mass of only a few Jupiter masses, we still require 1 km initial planetesimals to avoid catastrophic disruption at 0.7 AU. If the disk were eccentric or more massive, then even larger initial planetesimals would be required.

• Formation outside of \( \sim 3 \text{ AU} \) is much easier and more likely. Here a dense \((\Sigma_d(1\text{AU}) > 10^4 \text{ g cm}^{-2})\) disk (in agreement with sub-mm observations) is helpful as its gravity moves the secular resonance inwards of 2 AU, lowering planetesimal-planetesimal collision velocity. The dense disk also provides enough eccentricity damping to align \( \sim 10 \) km-sized planetesimals out to 10 AU during the disk lifetime.

• Disk precession can either facilitate or hinder plantesimal coagulation, depending on the direction and magnitude of precession. Slow retrograde precession results in emergence of a destructive secular resonance at several AU. On the other hand, prograde precession leads to a dynamically favorable location in the disk where the forced eccentricity is the same as the local gas eccentricity, relative planetesimal eccentricities are small, and there is no dynamical barrier to coagulation.

This work thus provides a dynamically-motivated picture of planetesimal growth towards building planetary cores in circumbinary protoplanetary disks.

This work is supported by NSF grant AST-1409524.
2.13 Appendix A: Planetesimal Dynamics with Linear Drag

In the case of linear gas drag (i.e. \( \tau_d \) independent of \( e_r \)) we can obtain the complete analytical solution for planetesimal dynamics including the precession of both the binary and the disk. With fixed \( \tau_d \) and \( \varpi_b(t) = \dot{\varpi}_b t, \varpi_d(t) = \dot{\varpi}_d t \) our general evolution equations (2.18)-(2.19) admit the following analytical solution:

\[
\begin{aligned}
&\begin{cases}
  k_p \\
  h_p
\end{cases} = e_{\text{free}} e^{-t/\tau_d} \begin{cases}
  \cos (At + \varpi_0) \\
  \sin (At + \varpi_0)
\end{cases} + \begin{cases}
  k_{f,d} \\
  h_{f,d}
\end{cases} + \begin{cases}
  k_{f,b} \\
  h_{f,b}
\end{cases}.
\end{aligned}
\]

(2.40)

Here \( e_{\text{free}} \) and \( \varpi_0 \) are the free eccentricity and periastron angle, \( k_{f,d} \) and \( h_{f,d} \) are the components of the forced eccentricity associated with the disk, and \( k_{f,b} \) and \( h_{f,b} \) are the components associated with the binary. These are given by

\[
\begin{aligned}
&\begin{cases}
  k_{f,d} \\
  h_{f,d}
\end{cases} = \left[ \frac{e_g^2 + \tau_d^2 B_d^2}{1 + \tau_d^2 (A - \varpi_d)^2} \right]^{1/2} \begin{cases}
  \cos (\varpi_d(t) + \phi_d) \\
  \sin (\varpi_d(t) + \phi_d)
\end{cases}, \quad \cos \phi_d = \frac{e_g - \tau_d^2 B_d (A - \varpi_d)}{(e_g^2 + \tau_d^2 B_d^2)^{1/2}[1 + \tau_d^2 (A - \varpi_d)^2]^{1/2}},
\end{aligned}
\]

(2.41)

and

\[
\begin{aligned}
&\begin{cases}
  k_{f,b} \\
  h_{f,b}
\end{cases} = \left[ \frac{B_b^2 + \tau_d^2}{1 + \tau_d^2 (A - \varpi_b)^2} \right]^{1/2} \begin{cases}
  \cos (\varpi_b(t) + \phi_b) \\
  \sin (\varpi_b(t) + \phi_b)
\end{cases}, \quad \cos \phi_b = \frac{-(A - \varpi_b) \tau_d}{[1 + \tau_d^2 (A - \varpi_b)^2]^{1/2}}.
\end{aligned}
\]

(2.42)

The relative particle-gas eccentricity is given by

\[
\begin{aligned}
&\begin{cases}
  k_r \\
  h_r
\end{cases} = \begin{cases}
  k_{f,b} \\
  h_{f,b}
\end{cases} - \tau_d \left[ \frac{B_d + e_g (A - \varpi_d)}{1 + \tau_d^2 (A - \varpi_d)^2} \right]^{1/2} \begin{cases}
  \cos (\varpi_d(t) - \varpi_r) \\
  \sin (\varpi_d(t) - \varpi_r)
\end{cases}, \quad \cos \phi_r = \frac{\tau_d (A - \varpi_d)}{[1 + \tau_d^2 (A - \varpi_d)^2]^{1/2}}.
\end{aligned}
\]

(2.43)

It is clear that \( e_r \) is in general a function of time.
2.14 Appendix B: Binary Precession Rate

Here we summarize results on the four major causes of binary precession which were discussed in Section 2.7. We write the precession rate as \( \dot{\varpi}_b = \dot{\varpi}_{GR} + \Sigma_{j=1}^2 (\dot{\varpi}_{T,j} + \dot{\varpi}_{R,j}) + \dot{\varpi}_{\text{disk}}, \) where \( \dot{\varpi}_{GR}, \dot{\varpi}_{T,j}, \dot{\varpi}_{R,j}, \dot{\varpi}_{\text{disk}} \) are the precession rates due to general relativity, tidal and rotational stellar quadrupoles, and disk gravity, respectively.

2.14.1 General Relativistic Precession

To first order in eccentricity, precession of Keplerian orbits due to general relativity is (Misner et al. 1973)

\[
\dot{\varpi}_{GR} = \frac{3(GM_b)^{1.5}}{c^2a_b^{2.5}} = 6.9 \times 10^{-6}\text{yr}^{-1} \left( \frac{M_b}{0.89M_\odot} \right)^{1.5} \left( \frac{0.22AU}{a_b} \right)^{2.5}.
\] (2.44)

2.14.2 Precession Due to Stellar Quadrupoles Induced by Tides and Rotation

Precession rates due to quadrupoles induced by tidal forces and from the rotational bulge are given in Sterne (1939) and Shakura (1985). They are given by

\[
\dot{\varpi}_{T,j} = 15k_{2,j}n_b \frac{M_r}{M_j} \left( \frac{R_j}{a_b} \right)^5 = 4.7 \times 10^{-6}\text{yr}^{-1} \frac{k_2}{0.13} \left( \frac{M_b}{0.89M_\odot} \right)^{0.5} \left( \frac{a_b}{0.22AU} \right)^{-6.5} \left( \frac{R_j}{2.03R_\odot} \right)^5,
\] (2.45)

\[
\dot{\varpi}_{R,j} = k_{2,j} \frac{M_b}{M_j} \omega_j^2 \left( \frac{R_j}{a} \right)^5 = 1.4 \times 10^{-6}\text{yr}^{-1} \frac{k_2}{0.13} \left( \frac{M_b}{0.89M_\odot} \right)^{0.5} \left( \frac{a_b}{0.22AU} \right)^{-6.5} \left( \frac{R_j}{2.03R_\odot} \right)^5,
\] (2.46)
Figure 2.14: Dependence of the dimensionless factor $\tilde{\phi}(a_b/a_{in}, \mu)$ in the disk-driven binary precession rate $\dot{\omega}_{\text{disk}}$ (see Equation (2.47)) on the relative size of the inner cavity $a_{in}/a_b$ and binary mass ratio $\mu$.

where $n_b$ is the Keplerian frequency, $r \neq j$, $\omega_j$ is the spin frequency of star $j$, $M_j$ is the mass of star $j$. $R_j$ is the radius of star $j$, and $k_2$ is the apsidal motion constant, both estimated for the primary in the Kepler 16 system from the pre-main-sequence stellar models of Claret (2012) assuming an age of 1 Myr and metallicity of $Z = 0.02$. For the numerical estimate in Equation (2.46), we have assumed the stars to be tidally locked, i.e. $\omega_j = n_b$. The numerical estimates in Equations (2.45) and (2.46) were done assuming the subscript “$j$” to refer to the primary star of the Kepler 16 system and the subscript “$r$” to refer to the secondary.
2.14.3 Precession Due to Disk Gravity

To calculate the binary precession due to disk gravity, we use Equations (20) and (A3) from Rafikov (2013), which assume that power law dependence of the disk surface density is sharply truncated at the inner radius \( a_{\text{in}} \). These state

\[
\dot{\varpi}_{\text{disk}} = \pi \tilde{\phi}(a_b/a_{\text{in}}, \mu) n_b \frac{\Sigma_0 a_b^3}{M_b a_{\text{in}}^{1+p}}
\]

\[
\approx 2.6 \times 10^{-3} \, \text{yr}^{-1} \left( \frac{0.89 M_c}{M_b} \right)^{0.5} \left( \frac{a_b}{0.22} \right)^{1/2-p} \frac{\Sigma_0}{3000 \, \text{g cm}^{-2}} \frac{\tilde{\phi}}{0.46}.
\]

(2.47)

Here \( \tilde{\phi}(a_b/a_{\text{in}}) \) is a function of the binary mass ratio \( \mu \) and ratio of binary semi-major axis to the radius of the central gap in the disk \( a_b/a_{\text{in}} \), and we have assumed \( a_b/a_{\text{in}} = 0.5 \). Figure 2.14 shows that \( \tilde{\phi} \) is a slowly varying function of \( a_b/a_{\text{in}} \), so roughly \( \dot{\varpi}_{\text{disk}} \sim (a_b/a_{\text{in}})^{1+p} \). For \( p = 1.5 \), assuming \( a_b/a_{\text{in}} \) to be 1/3 instead of 1/2 would lower \( \dot{\varpi}_{\text{b}} \) by almost a factor of 3.

2.15 Appendix C: Collisional Outcomes

Consider two bodies of mass \( m_1 \) and \( m_2 \), colliding at speed \( v_{\text{coll}} \). Stewart & Leinhardt (2009) give the mass of the largest remnant \( M_{\text{lr}} \) as

\[
\frac{M_{\text{lr}}}{M_{\text{tot}}} = -0.5 \left( Q_R/Q_{RD}^* - 1 \right) + 0.5
\]

(2.48)

Here, \( M_{\text{tot}} = m_1 + m_2 \) is the total mass of the colliding bodies, \( Q_R = 0.5 m_1 m_2 v_{\text{coll}}^2/M_{\text{tot}}^2 \) is the center of mass specific kinetic energy, and \( Q_{RD}^* \) is a quantity dependent on the material properties and \( M_{\text{tot}} \). As shown in Figure 2 of RS15b, the critical velocities for catastrophic disruption are on the order of several \( \text{m s}^{-1} \) for \( d_p = 100 \, \text{m planetesimals} \). The critical speed reflects both the strength of the body, and reaccumulation of fragments due to gravity. Material strength becomes subdominant to gravity for bodies larger than a few hundred meters.
We take collision velocity to be given by

\[ v_{\text{coll}} = \sqrt{(e_{12}v_K)^2 + \frac{2G(m_1 + m_2)}{d_1 + d_2}}, \]  

(2.49)

where the first term is due to the relative velocity at infinity, \( e_{12} = |\mathbf{e}_p(d_1) - \mathbf{e}_p(d_2)| \), and the second is due to the potential energy of the colliding objects. It should be noted (RS15a) that the actual velocity at infinity between two colliding bodies can be anywhere between \( 0.5e_{12}v_K \) and \( e_{12}v_K \). Thus Equation (2.49) may overestimate the true \( v_{\text{coll}} \) by up to a factor of 2. Because the collision velocity must be several times the escape velocity in order to be destructive, the potential energy term is relatively unimportant.

We can express the relative eccentricity of two planetesimals in terms of \( e_c, d_c \) and the planetesimal sizes \( d_1 \) and \( d_2 \) as (RS15a)

\[ e_{12} = e_c \frac{|(A - \hat{\omega}_d)(\tau_1 - \tau_2)|}{\sqrt{(1 + (A - \hat{\omega}_d)^2\tau_1^2)(1 + (A - \hat{\omega}_d)^2\tau_2^2)}}. \]  

(2.50)

Here, \( |(A - \hat{\omega}_d)\tau_i| \ (i = 1, 2) \) is given in terms of \( d_c \) and \( d_i \) by Equation (2.28).

In the commonly encountered limit of \( d_1 \) and \( d_2 \) both being much greater than \( d_c \), this simplifies to

\[ e_{12} \approx e_c \frac{|d_1 - d_2|d_c}{d_1d_2}. \]  

(2.51)

We see from this that for encounters between two large bodies, the relative eccentricity is never larger than \( e_cd_c/\min(d_1, d_2) \).
Bibliography

Adachi, I., Hayashi, C., & Nakazawa, K. 1976, Progress of Theoretical Physics, 56, 1756


Hayashi, C. 1981, Progress of Theoretical Physics Supplement, 70, 35


Shakura, N. I. 1985, Soviet Astronomy Letters, 11, 224


—. 2011, Icarus, 214, 671


Chapter 3

A Coagulation/Fragmentation Simulation for Planetesimals in Binary Systems

3.1 Abstract

In situ planet formation via core accretion is theoretically challenging in tight binary systems because of destructive high-velocity planetesimal collisions. Previous work has focused on detailed analysis of the collision velocities. In this paper, we study the effect of a distribution of collision velocities on the evolution of the size distribution of a planetesimal population. We numerically solve for the evolution of the distribution of planetesimal masses in a circumprimary disk about one component of a tight binary system. Our code accounts for both coagulation/fragmentation in situ, and the effects of a size-dependent radial inspiral, which preferentially removes small bodies from the disk. We determine the resilience of the coagulation process to destructive collisions, and compare this to previously used heuristics.

3.2 Introduction

Planets have been found in a wide variety of stellar systems. There are many unsettled questions about the mechanism of planet formation. In particular, tight eccentric planet-hosting binary systems such as γ Cephei present a challenge to theories of planet formation
involving core accretion through mutual collisions of small planetesimals. Gravitational perturbations from the eccentric companion are expected to drive the eccentricity of these planetesimals to high values, thus resulting in the possibility of collisional destruction rather than growth of planetesimals in mutual collisions. A simple calculation (Heppenheimer, 1978), including only the effect of gravity from the stellar companion on planetesimal orbits, yields planetesimal collision velocities of a few km s$^{-1}$ at the current location of the planet in the γ Cephei system. This is enough to destroy even planetesimal 100’s of kilometers in size. A number of ideas have been proposed to alleviate this problem.

It was proposed (Marzari & Scholl, 2000) that gas drag from the disk would apsidally align planetesimal orbits. While orbits would still have similar eccentricities as in Heppenheimer’s model, the relative eccentricities (which determine collision velocities) would be reduced. However, it was later recognized (Thébault et al., 2008) that only planetesimals of the same size are apsidally aligned, as the strength of the frictional coupling to the gas disk is determined by the planetesimal size. When this is taken into account, the beneficial effect of the gas drag is greatly reduced.

Rafikov (2013) noted that the gas disk couples to planetesimals not only through gas drag, but also gravitationally. He showed that a massive axisymmetric disk induces a rapid apsidal precession of planetesimal orbits, which mediates the eccentricity excitation from the companion. It is not at all clear however that the gas disk will be axisymmetric (Paardekooper et al., 2008; Regály et al., 2011; Marzari et al., 2009). For this reason Silsbee & Rafikov (2015) calculated the gravitational effect of an eccentric disk on planetesimal orbits. They found that collision outcomes depended heavily on the details of the disk structure. The disk could either increase or decrease collision velocities, depending on the radial dependence of the eccentricity and surface density.

from Stewart & Leinhardt (2009) to determine regions in the disk where planet formation could occur.

There remains however, some uncertainty in understanding which collisional environments allow planetesimal growth. Because of the size dependent orientation of planetesimals discovered in Thébault et al. (2008), planetesimals of similar size will collide at low velocity. Planetesimals of different sizes will collide at higher velocity. In an environment with a mixture of low and high velocity collisions, it is challenging, even qualitatively, to determine the evolution of the size distribution. It has been found that including the distribution of collision velocities can qualitatively change the outcome of the coagulation process (Windmark et al., 2012; Garaud et al., 2013). In particular, Windmark et al. (2012) found that using a Maxwellian collision velocity distribution instead of a delta function at the rms velocity of the Maxwellian would increase the size of the largest particle in their dust grain-growth simulations by a factor of $\sim 60$. If a particle becomes large enough, it is then less susceptible to fragmentation, so it can continue to grow larger.

The problem is complicated by the radial inspiral of particles due to interaction with the gas. The gas disk is slightly pressure-supported, and as a result moves $\sim 30$ m s$^{-1}$ slower than the Keplerian speed (Weidenschilling, 1977). As a result, solid bodies with stopping times longer than an orbital time feel a head-wind, which causes them to spiral into the star. The inspiral is more rapid in the case where there is some forced eccentricity relative to the gas (Adachi et al., 1976). It was suggested in Rafikov & Silsbee (2015b) that the eccentricity dependence of the inspiral rate could concentrate solid bodies in the regions of the disk with low particle-gas eccentricity. Inspiral also helps to alleviate the effect of erosion by small bodies by rapidly removing them from the system.

This paper uses the dynamics described in Rafikov & Silsbee (2015a) and Rafikov & Silsbee (2015b), to determine collision velocities and radial inspiral rates. Collisional outcomes are determined using the model from Stewart & Leinhardt (2009). With this physical model, we are able to numerically evolve the size distribution in a number of discrete annuli. This
allows us to determine the success of the coagulation process for a given environment, rather than just the outcomes of individual collisions. All the details of the code are contained in Appendix A. The main body of this chapter deals only with results from the code.

3.3 Model System

We consider a model system with the binary parameters of the \( \gamma \) Cephei system. The disk is identical to the disk described in Silsbee & Rafikov (2015). This is an eccentric disk with the innermost streamline having semi-major axis equal to \( a_{\text{in}} \) and the outermost having a semi-major axis of \( a_{\text{out}} \). The surface density has non-axisymmetric structure due to the disk eccentricity as described in Silsbee & Rafikov (2015) and Statler (2001). The apsidal angle of the disk with respect to the binary is denoted \( \varpi_d \) and is assumed to be fixed. The surface density at periastron and eccentricity of a gas streamline with semi-major axis \( a \) are assumed to be described by power laws with indices \( \mu \) and \( \nu \) respectively:

\[
\Sigma(r) = \Sigma_0 \left( \frac{a_0}{a} \right)^\mu \quad e(r) = e_0 \left( \frac{a_0}{a} \right)^\nu.
\]

In this work we will take \( a_0 = 1 \text{AU} \).

3.4 Results for Single-Annulus Coagulation

In this section we use the code in a single annulus to determine the prospects for coagulation ignoring the effects of radial inspiral. In Rafikov & Silsbee (2015b), we made predictions for locations in the disk in which planetesimal coagulation could occur, using a few heuristic measures — the absence of collisions leading to catastrophic disruption, or the absence of erosion by particles of similar masses. As discussed previously, the problem is a statistical one; our model permits a formally non-zero probability of coagulation under any set of environmental conditions because there is always a possibility of low-velocity collisions in
Figure 3.1: The evolution of the mass distribution as a function of time. Different panels correspond to environments with the different values of the critical eccentricity and critical mass as indicated on the panels. These were made for a location 2 AU from the primary star in the γ Cephei system. More details are provided in the text.
Figure 3.2: This shows the relative eccentricity and maximum collision velocities between particles as a function of their sizes. Each panel corresponds to a different level of disk eccentricity, which leads to the values of $e_c$ and $m_c$ labeled on the panels.
any environment, and a body that has a string of low-velocity collisions will eventually grow large enough to be resistant to higher velocity collisions.

Figure 3.1 shows how the distribution of mass in one annulus changes as a function of time. This was made at 2 AU in the \( \gamma \) Cephei system. We took \( \mu = 1, \nu = -1, \Sigma_0 = 15,000 \) g cm\(^{-2}\), a solid to gas ratio of 0.01, and \( \varpi_d = 0 \). \( \epsilon_0 \) was varied to yield the values of \( \epsilon_c \) labelled on the diagram. We used a bin spacing \( M_{i+1}/M_i = 1.15 \). The simulation started with all particles 10 centimeters in size. Different curves correspond to different times since the beginning of the simulation. The simulation for the upper left panel stopped when the largest body became larger than the maximum mass in the simulation.

Figure 3.2 shows the outcome of collisions as a function of the sizes of the collision partners, for the same 4 environments shown in Figure 3.1. This assumes that the collision speed is equal to \( e_{ij}v_K \), though in practice it can be as low as \( e_{ij}v_K/2 \) (see Equation (3.9)). The lines enclose regions of collisions which lead to erosion. None of the collisions lead to catastrophic disruption.

We observe, not surprisingly, that lower values of \( \epsilon_c \) lead to more growth. The only panel in which particles are able to grow to the maximum size in the simulation is the upper left panel, with \( \epsilon_c = 0.0004 \). In all other panels, growth stalls at some point. Growth would continue indefinitely in panel 1 were we to simulate larger bodies, as one can see from Figure 3.2 that there is very little erosion of the bodies in the largest mass-bin. One can see that the mass at which growth stalls depends on the value of \( \epsilon_c \), or equivalently on the width of the channel of non-erosive collisions in Figure 3.2.

What perhaps is a bit surprising is that no catastrophic disruption is necessary to halt growth. Even in the case with \( \epsilon_c = 0.0008 \), nearly twice the critical value necessary to stall growth, no collisions lead to catastrophic disruption. There are enough erosive collisions that none of the planetesimals experience a sufficient number of consecutive low-velocity collisions to grow large enough to be resistant to erosion.
3.4.1 Comparison with Simple Heuristics

In previous works (e.g., Rafikov & Silsbee 2015b), we used simple models to determine conditions under which growth would occur. The simplest of these is to assume that growth occurs if and only if there are no catastrophically disruptive collisions. It is also possible that erosion can prevent growth even in the absence of catastrophic disruption, so we proposed a model in which growth happens unless erosive collisions occur between planetesimals of “similar” sizes.

We tested the validity of these two models by running the following simulations. We initialized a population of planetesimals with size \( d_{\text{init}} \). For many values of \( e_c \) and \( d_c \), we ran multiple simulations to determine the critical value \( d_{\text{min}}(e_c, d_c) \) of \( d_{\text{init}} \) such that growth to 300 km will occur.

Figure 3.3 shows \( d_{\text{min}} \) as a function of \( d_c \) for 4 different values of \( e_c \) (blue line). The red line shows the minimum value of \( d_{\text{init}} \) such that no catastrophic disruption occurs for planetesimals of size \( d_{\text{init}} \) or larger. The green line shows the minimum value of \( d_{\text{init}} \) such that no body larger than \( d_{\text{init}} \) suffers an erosive collision with a partner with mass ratio greater than 1%. This was the same criterion used to denote the grey region in Figure 4 of Rafikov & Silsbee 2015b. On the right side of the plots, where no data is shown, \( d_c \) is high enough that growth can occur starting from any size. We conclude that depending on \( d_c \) and \( e_c \), that \( d_{\text{min}} \) is between 2 and 5 times larger than what would be predicted using the criterion of no catastrophic disruption. Note that this is a quite general result. It was shown in Rafikov & Silsbee 2015a that the distribution of collision velocities induced by secular perturbations from the companion star depend only on \( e_c \), \( d_c \) and the Keplerian speed.

For this simulation, we used a bin spacing \( M_{i+1}/M_i = 1.15 \), a minimum planetesimal size of 10 meters, a maximum planetesimal size of 300 km, a planetesimal inclination of \( 10^{-4} \), and a maximum running time of \( 10^6 \) years. We assumed a Keplerian speed of 26 km s\(^{-1}\), consistent with a location at 2 AU from \( \gamma \)-Cephei A.
Simulation results
No erosion
No catastrophic disruption

Figure 3.3: This shows $d_{\text{min}}$ as a function of $d_c$ for different values of $e_c$ (blue line). The green line is the estimate for $d_{\text{min}}$ assuming that growth can occur starting from the smallest size which does not suffer erosion by any particle greater than 1% of its mass. The red line is the estimate for $d_{\text{min}}$ assuming that growth can occur in the absence of catastrophic disruption.
3.5 Conclusions

We have written a coagulation fragmentation code for planetesimal coagulation in binary systems. It includes the size-dependent collisional outcomes calculated in Rafikov & Silsbee (2015b). It also considers radial inspiral of planetesimals. While it is currently configured to calculate the evolution of the size distribution in a circumprimary disk with an external perturber, it would be easy to alter it to work for a circumbinary disk as well.

While we have not yet used the full functionality of the code, we have derived one important result. We found that growth can occur in the face of some collisional erosion, but will not occur in environments in which there are collisions leading to catastrophic disruption.

3.6 Appendix A: Functionality of the Code

The code models the evolution of the size distribution of planetesimals in the disk in $N_{\text{ann}}$ discrete annuli. The planetesimal size distribution is calculated in $N_{\text{bins}}$ discrete logarithmically spaced mass bins. At any point in time, in a given annulus, all of the information about the size distribution is encoded in a single vector $\mathbf{N}$, such that $N_i$ is the number of planetesimals in mass bin $i$. The code performs two basic functions, described in the following sections. Independently in each of the annuli it models the evolution of the size distribution due to planetesimal-planetesimal collisions. It also models the size-dependent inward migration of planetesimals due to gas drag by passing particles inwards between annuli.

3.6.1 Planetesimal Collisions - Overview

In a given annulus, in each time step, the code calculates the mean collision rate $R_{ij}$ between planetesimals in bin $i$ and bin $j$ for all $i$ and $j$ in $[1, N_{\text{bins}}]$. This calculation is described in Section 3.6.4. The number of collisions $Z_{ij}$ in a given time step of length $\Delta t$ between bodies
in mass bins \( i \) and \( j \) is given by

\[
Z_{ij} = \text{Poisson}(R_{ij}\Delta t).
\]  

(3.2)

We define an operator \( F \) such that \( F_{ijk} \) is the number of fragments in mass bin \( i \) produced by a collision of a particle in bin \( j \) with a particle in bin \( k \) (see Section 3.6.2). The mass distribution \( \vec{N} \) is updated as

\[
N_i = N_i + \sum_{j=1}^{N_{\text{bins}}} \sum_{k=1}^{N_{\text{bins}}} C_{ijk}, \tag{3.3}
\]

where

\[
C_{ijk} = Z_{jk} \frac{1 + \delta_{jk}}{2} \left[ F_{ijk} - \delta_{ij} - \delta_{ik} \right].
\]  

(3.4)

This takes into account both addition of particles to bin \( i \) from collisional fragments (the \( F_{ijk} \) operator), as well as collisional destruction of particles in bin \( i \) (the \(-\delta_{ij}\) and \(-\delta_{ik}\) terms). The factor of \((1 + \delta_{jk})/2\) is to prevent over-counting when summing over both \( j \) and \( k \).

### 3.6.2 Collisional Outcomes

This section describes the physics which goes into the determination of \( F_{ijk} \). The outcome of a collision between a body in mass bin \( i \) and one in mass bin \( j \) is dependent on the collision velocity. High velocity collisions will result in many small fragments, and no large remnant bodies. Low velocity collisions will result in one body which contains nearly all of the pre-collision mass. Collision velocity is calculated in Section 3.6.3.

Given a collision velocity \( v_{\text{coll}} \), we then use the formula provided in Stewart & Leinhardt(2009) to determine the mass of the largest remnant \( M_{\text{lr}} \). We modify their formula to impose a minimum size of the largest remnant, even for arbitrarily high velocity collisions. We take
the mass of the largest remnant to be

\[
\frac{M_{\text{lr}}}{M_{\text{tot}}} = 1 - 0.5 \frac{Q_{R}}{Q_{RD}}. \quad (3.5)
\]

\[M_{\text{tot}} = m_{i} + m_{j}\] is the total mass of the two incoming bodies. \(Q_{R}\) is specific center of mass energy of the collision given by

\[Q_{R} = 0.5 \frac{m_{i} m_{j} v_{\text{coll}}^{2}}{M_{\text{tot}}}. \quad (3.6)\]

\(Q_{RD}\) is the critical value of \(Q_{R}\) which leads to catastrophic disruption (i.e. \(M_{\text{lr}} < 0.5M_{\text{tot}}\)). This is a function of the sizes and composition of the planetesimals, and \(v_{\text{coll}}\) \([\text{Stewart \& Leinhardt} 2009]\). When \(M_{\text{lr}}\) is calculated, it will in general be between \(M_{i}\) and \(M_{i+1}\) for some \(i\). If \(M_{\text{lr}} < 0\), then we put all of the mass into fragments as described below. We then add a fraction \(P_{i}\) to bin \(i\) and \(P_{i+1}\) to bin \(i+1\) given by

\[
P_{i} = \frac{M_{\text{lr}} - M_{i}}{M_{i+1} - M_{i}}, \quad P_{i+1} = \frac{M_{i+1} - M_{\text{lr}}}{M_{i+1} - M_{i}}. \quad (3.7)
\]

We assume that the rest of the fragments follow a size distribution with power law index \(t\). We define an upper cutoff mass \(M_{\text{cutoff}} = \max(\min(M_{\text{lr}}, M_{\text{tot}} - M_{\text{lr}}), bM_{\text{tot}})\). Here \(b\) is an adjustable parameter which determines the cutoff mass for catastrophic collisions. We take it to be \(10^{-4}\). Using this power-law distribution, mass \(M_{\text{lost}}\) is lost from the simulation because it is in bodies with masses less than \(M_{1}\). \(M_{\text{lost}}\) is given by

\[
M_{\text{lost}} = (M_{\text{tot}} - M_{\text{lr}}) \min \left(1, \frac{\int_{0}^{M_{\text{min}}} m^{t+1} dm}{\int_{0}^{M_{\text{cutoff}}} m^{t+1} dm} \right), \quad (3.8)
\]

where \(M_{\text{min}} = M_{0}/r_{m}, \ r_{m} = M_{i+1}/M_{i}\). The remaining fragment mass \(M_{\text{frag}} = M_{\text{tot}} - M_{\text{lr}} - M_{\text{lost}}\) is assigned to mass bins such that \(N_{\text{frag},i} \propto M_{i}^{-t}\), where \(N_{\text{frag},i}\) is the number of fragments in mass bin \(i\). This relation holds for \(0 < i \leq i_{\text{cutoff}}\), where \(i_{\text{cutoff}}\) is the index of the largest bin with mass less than \(M_{\text{cutoff}}\). \(M_{\text{cutoff}}\) will typically lie between two mass bins. In this case we choose \(i_{\text{cutoff}}\) randomly such that \(\langle M(i_{\text{cutoff}}) \rangle = M_{\text{cutoff}}\). Similarly, at the end
of a time step if \( X \) particles are expected to be added to a mass bin, then \( \lfloor X \rfloor \) are added with probability \( \lfloor X \rfloor - X \), and \( \lceil X \rceil \) are added with probability \( X - \lfloor X \rfloor \).

### 3.6.3 Determination of Collisional Velocity

Rafikov & Silsbee (2015a) determine the relative eccentricity between planetesimals as a function of their sizes. They take into account secular perturbations from the binary companion, and the eccentric gas disk, as well as drag forces from the disk. Due to interaction with the gas, planets follow different orbits depending on their size. Small planetesimals have eccentricity similar to that of the gas streamlines. The orbits of large planetesimals have the forced eccentricity determined by the gravitational potential of the binary and the disk. The collision velocities can be expressed in terms of a critical eccentricity \( e_c \) and a critical size \( d_c \), corresponding to a critical mass \( m_c = 4\pi \rho_p d_c^3 / 3 \).

\( m_c \) is the mass around which particles transition from alignment with gas streamlines to having the forced eccentricity. It is the mass at which the stopping time is comparable to the period of precession of free eccentricity. \( e_c \) is the relative eccentricity between the two limiting cases discussed in the previous paragraph. It is the maximum relative eccentricity between particles of any size. These results are derived and discussed quantitatively in Rafikov & Silsbee (2015a) (which forms Chapter 1 of this thesis).

Rafikov & Silsbee (2015a) give the relative eccentricity between planetesimals as a function of their masses in their Equation (64). Given the relative eccentricity \( e_{ij} \) between two planetesimals, there is a distribution of relative encounter velocities \( v_{ij} \) between \( v_{\text{min}} = 0.5e_{ij}v_K \) and \( v_{\text{max}} = e_{ij}v_K \), where \( v_K \) is the local Keplerian speed. This distribution is given in Equation (62) of Rafikov & Silsbee (2015a) as

\[
\frac{df_{12}}{dv_{12}} = \frac{v_{\text{max}}^{-2}v_{12}^2}{E(\sqrt{3}/2)[(v_{\text{max}}^2 - v_{12}^2)(v_{12}^2 - v_{\text{min}}^2)]}.
\] (3.9)
Equation (3.9) gives the distribution of encounter velocities between the two planetesimals at infinity in a planar model with no random motions. We also consider velocity due to the random motion of planetesimals. We assume that the velocity at infinity $v_{\text{inf}}$ is given by

$$v_{\text{inf}} = \sqrt{v_{ij}^2 + (I_{ij}v_K)^2}$$

(3.10)

$I_{ij}$ is the relative inclination between the two particles which we approximate as

$$I_{ij} = \sqrt{I_i^2 + I_j^2}$$

(3.11)

In this approximate treatment, we have ignored random motions in the plane of the disk. We take the particle inclination $I_0$ to be fixed and independent of particle size throughout the simulation. There is some additional energy in the actual collision due to the gravitational attraction between the two planetesimals. Therefore, assuming the colliding partners are spherical, they collide with velocity $v_{\text{coll}}$ given by

$$v_{\text{coll}} = \sqrt{v_{\text{inf}}^2 + 2G(m_1 + m_2)/(d_1 + d_2)}$$

(3.12)

A random number between 0.5 and 1 (in units of $v_{\text{min}}$) is drawn each time step from the distribution in Equation (3.9). Ideally this would be done independently for each set of collision partners, but to save time we only draw one number for each time step and use it to determine the collision velocity for all collisions in that time step.

### 3.6.4 Collision Rate

The collision rate depends on the relative inclinations of the particles. If the random eccentricities and inclinations are small compared with the forced eccentricity ($i_r < e_{\text{forced}}/2$), we use Equation (63) from Rafikov & Silsbee (2015a) (corrected so that $e_{\text{max}}$ is replaced with $v_{\text{max}}/a_p$). This gives a probability $P_{ij}$ per unit time of a given planetesimal from bin $i$
colliding with a given one from bin \( j \) of

\[ P_{ij} = \frac{\sigma_{ij} E(\sqrt{3}/2)v_{\text{max}}}{2\pi^3 a_p^2 \Delta a_{ij}}. \] (3.13)

\( \sigma_{ij} \) is the collision cross-section including gravitational focusing:

\[ \sigma_{ij} = \pi(d_i + d_j)^2 \left( 1 + \frac{2G(m_1 + m_2)}{(d_1 + d_2)v_{\text{inf}}^2} \right). \] (3.14)

This simple prescription needs to be modified in the case of collisions within the same mass bin, where the focusing factor can vary significantly depending on the location within the mass bin of the two colliding partners. That is two bodies with masses right at the center of the mass bin will have a higher focusing factor than one body at the bottom end of the mass bin, and one body at the top end, as in the latter case the approach velocity will be higher. This is discussed in Appendix C.

### 3.6.5 Radial Inspiral

Using Equations (11-14) of Rafikov & Silsbee (2015b) we obtain the radial drift velocity of particles as a function of their size and orbital radius. Particles drift inwards due to a combination of the sub-Keplerian rotation of the disk, and the particle-gas eccentricity. Let \( \Delta r_i \) be the radial width of annulus \( i \), and let the drift rate of a particle in annulus \( i \) and mass bin \( j \) be \( v_{i,j} \). We then let the change in the number of particles in annulus \( i \) and mass bin \( j \) be given by

\[ \Delta N_{i,j} = -\frac{N_{i,j} v_{i,j} \Delta t_{\text{insp}}}{\Delta r_i} + \frac{N_{i+1,j} v_{i+1,j} \Delta t_{\text{insp}}}{\Delta r_{i+1}}, \] (3.15)

### 3.6.6 Time Step Determination

In this section we describe how the time-steps for the coagulation process and the radial inspiral are determined.
Coagulation/Fragmentation Time Step

The time step for the coagulation/fragmentation process must be short enough that the distribution does not change appreciably during the length of a time step. In principle, this means we would like

$$\frac{|N_i(t) - N_i(t + dt)|}{N_i(t)} < \epsilon$$

for all $i$, where $\epsilon$ is some small number. In practice, this is complicated by the fact that mass is added to bins in discrete quantities, and that some bins have a very small number of particles. In a bin with a small number of particles, even adding one additional particle will make a large fractional change to the amount of mass in that bin. Additionally, collisions are a random process, so a fluctuation can change the mass in a bin by more than expected. To address these concerns, we instead pick the largest time step $\tau_{\text{coag}}$ such that for all $i$

$$\left\langle \frac{|N_i(t) - N_i(t + \tau_{\text{coag}})|}{N_i(t)} \right\rangle < \epsilon_1 \quad \text{or} \quad \left\langle \frac{|m_i(N_i(t) - N_i(t + \tau_{\text{coag}}))|}{\sum_{i=1}^{\text{Nbins}} m_i N_i} \right\rangle < \epsilon_2,$$

where the angle brackets refer to averaging over the distribution of encounter velocity in Equation (3.9). To save computational time, we approximate this average by using the median value of $v_{ij}$. Here, as before, a time step is acceptable if every bin changes by a small fractional amount less than $\epsilon_1$, but it is also acceptable if the change in mass is less than $\epsilon_2$ times the total mass in the system. In our calculations, we typically take $\epsilon_1 = 0.05$, and $\epsilon_2 = 10^{-6}$. $\tau_{\text{coag}}$ is re-calculated at the beginning of each time step, separately for each annulus, and the minimum value among all the annuli is used for each time step.
Radial Inspiral Time Step

The code uses a different time step for the radial drift. Radial drift is re-calculated every interval of time $\tau_{\text{drift}}$ such that

$$\tau_{\text{drift}} \cdot \text{max} \left( \frac{v_{\text{drift}}}{\Delta r_i} \right) = \epsilon_{\text{drift}}$$

(3.18)

While we see from Equation (3.15) that $\epsilon_{\text{drift}}$ must not be larger than 1, it is apparent from the results of Section 3.7.3 that there is little advantage to picking $\epsilon_{\text{drift}}$ less than 1. $\tau_{\text{drift}}$ is calculated once at the beginning of the simulation, as $v_{\text{drift}}$ and $\Delta r_i$ do not change with time.

3.7 Appendix B: Analytic Tests of the Code

In this appendix we describe the tests we did of the code against analytic expressions in certain limiting cases. Unfortunately we are not aware of an analytic solution to a problem that incorporates coagulation, fragmentation, and radial inspiral. However, we test these three functionalities independently and find that the code does a reasonable job of reproducing known solutions to the accuracy warranted given the imperfect knowledge of the physical processes involved.

3.7.1 Tests Against Known Solutions of the Coagulation Equation

In this section we describe the results of simulations of the coagulation process without fragmentation, and with simplified collision rates and initial conditions that have analytic solutions we can compare with. There are three known analytic solutions to the coagulation equation.

$$\frac{dN_k}{dt} = \frac{1}{2} \sum_{i+j=k} A_{ij} N_i N_j - N_k \sum_{i=1}^{\infty} A_{ik} N_i$$

(3.19)
These solutions are for kernels $A_{ij} = 1$ (Smoluchowski [1916]), $A_{ij} = i + j$ and $A_{ij} = ij$ (Trubnikov [1971]), with the initial condition of $N$ particles of a single mass. Our code reproduces these solutions with small deviations as discussed below.

The top panel of Figure 3.4 shows the number of particles per unit mass as a function of mass for three different times for the kernel $A_{ij} = 1$. In this figure we have used a mass bin spacing $M_{i+1}/M_i = 1.15$. Initially there were $N = 10^{12}$ particles all of mass 1. Because the distribution evolves more rapidly if there are more particles, we express time in terms of $\eta = Nt$. In the limit of large $N$, the shape of the distribution depends only on $\eta$, rather than on $N$ and $t$ independently.

The large scatter in the points at small masses is due to the logarithmic spacing of the mass bins and the initial condition that all the mass is in bodies with mass 1. Since we are assuming perfect sticking in collisions for this exercise, mass bins containing integer masses will be over-represented, leading to an uneven distribution at low masses. This effect is not important in any of our planetesimal coagulation simulations both because we include fragmentation, and because do not start with all the mass in one mass bin. Other than this discrepancy at low mass, there is nearly perfect agreement between our code and the analytic solution.

The middle panel of Figure 3.4 shows the number of particles per unit mass as a function of mass for three different values of $\eta$ for the kernel $A_{ij} = i + j$. The distribution evolves more rapidly than in the case of the constant kernel because of the higher collision rates. In this case, in addition to the scatter at low masses, we see that the numerical solution is slightly ahead of the analytic solution at the high-mass end of the distribution.

The bottom panel of Figure 3.4 shows the number of particles per unit mass as a function of mass for three different values of $\eta$ for the kernel $A_{ij} = ij$. Here the system evolves in a more complicated manner than the previous two cases (Trubnikov [1971]). For $\eta < 1$, orderly growth from smaller to larger particles progresses as in the previous cases. However, for $\eta > 1$, one body becomes much larger than the others, and eventually consumes all the mass.
Figure 3.4: Comparison of simulation and analytic results starting with all particles of mass 1. $\eta = Nt$ is a time coordinate, re-scaled by the number of particles initially in the system. Each panel covers a different coagulation kernel.
in the system. Our simulation is not well-equipped to handle this runaway growth phase, so we restrict ourselves to reproducing the analytic solutions for $\eta < 1$.

The discrepancy between the numerical and analytic solutions at $\eta = 0.99$ looks somewhat alarming. This is because the distribution function is evolving quite rapidly at this time. We also plotted the analytic solution for $\eta = 0.997$, and we see that it matches the numerical solution well, implying that this difference only corresponds to a lag of 0.7% in the numerical solution.

### 3.7.2 Fragmentation Cascades

We are not aware of fragmentation cascades with time-dependent behavior that are analytically solvable as in the coagulation case, but there are some steady-state models that lead to analytic predictions. In this section we use these models to verify that the portion of our code that produces the fragment distribution from a collision is operating correctly.

O’Brien & Greenberg (2003) provide a model of collisional fragmentation that leads to a steady state size distribution given by a power-law. Their model has one free parameter $s$, which parametrizes the strength of the particle:

$$Q_{\text{RD}}^* = Q_0 d^s.$$  \hspace{1cm} (3.20)

They show that the size distribution in the steady-state collisional cascade is given by

$$N_s(d) = Bd^{-p},$$  \hspace{1cm} (3.21)

where $N_s$ is the number of fragments of size $d$, and $B$ is an overall scaling constant. The power-law index of the size the steady-state collisional cascade is given by

$$p = \frac{21 + s}{6 + s}. $$  \hspace{1cm} (3.22)
O’Brien & Greenberg (2003) have assumed that the collisional cross-section between two bodies is proportional to the square of the size of the target body (i.e. the geometric cross-section). If we relax this assumption, as was done in Tanaka et al. (1996), and instead let the collision cross-section be proportional to $d^\alpha$, one can show that

$$p = \frac{15 + 3\alpha + s}{6 + s}. \quad (3.23)$$

Equation (3.23) was derived assuming a steady-state mass distribution with no upper or lower cut-off. Since we are only able to simulate a finite range in masses, we must have a way of supplying mass to the system to achieve a steady state, otherwise all mass will eventually be contained in particles smaller than the lower cutoff size for the simulation. We do this by fixing the counts in the top 40% of the mass bins, and letting the collisional cascade provide the mass to the mass bins corresponding to smaller particles. By doing this we continuously re-supply mass to the large mass bins.

Figure 3.5 shows a comparison of the steady-state result of the simulation against the expected power law distribution from Equation (3.23) for three different values of $s$ and $\alpha$. This was made with bin spacing $M_{i+1}/M_i = 1.15$. In each case, the top panel shows the number of bodies in each bin as a function of bin mass. The blue line shows the number in the simulation, and the green line shows the expected number based on Equation (3.23). The bottom panels show the logarithmic slope $d\ln N_i/d\ln(M_i)$ as a function of mass, calculated as

$$d\ln N_{\text{bin}}/d\ln(m) = \frac{\ln N_{i+1} - \ln N_{i-1}}{\ln M_{i+1} - \ln M_{i-1}}. \quad (3.24)$$

Panels A and B show the case where where $s = 0$ and $\alpha = 2$ (Dohnanyi 1969). In this case Equation (3.23) yields a slope in the size distribution of -7/2. Assuming constant density so that $M \sim d^3$, we can show that

$$dN/dm = B'M^{-q}, \quad (3.25)$$
where \( q = (p + 2)/3 \), and \( B' \) is the appropriate normalization. The number of particles contained in the bin with bin mass \( m \) scales as \( N_{\text{bin}}(m) \sim m^{-(p-1)/3} \sim m^{-\frac{3+\alpha}{3+\alpha}} \), due to the logarithmic spacing of the mass bins.

Panels C and D are the same as A and B except that \( s = -1.5 \) and \( \alpha = 0 \), leading to a slope \( d \ln(N_{\text{bin}})/d \ln(m) = -2/3 \). The slope is reproduced faithfully except right near the lower end of the mass distribution. This time, there’s a regular oscillation that doesn’t go away with decreasing time step. This is due to the finite spacing of mass bins, combined with the fact that if \( s \neq 0 \), the critical size ratio for a collision to be destructive to the larger particle is size-dependent.

In panels E and F, we let \( s = 0 \) as in Figure 3.5, but let \( \alpha = 1 \). This leads to the expectation that \( d \ln(N_{\text{bin}})/d \ln(m) = -2/3 \).

In these three plots, we see that our simulation has nearly perfect agreement with the expected power-laws, except for understandable deviations at specific points. Objects at the small end of the mass distribution, have nothing smaller than themselves in the simulation, so they are destroyed less frequently than the model would predict, hence the steepening of the power-law. Then the over-abundance of small masses creates an under-abundance of objects of slightly larger masses, creating the wave-like pattern seen. This effect is reduced in panels C and D because we tuned the \( Q_0 \) in Equation (3.20) so that objects at the lower end of the simulated mass range would not be destroyed by smaller particles, even if they were in the simulation. The wiggle in this plot is due to the finite spacing of the mass bins.

### 3.7.3 Mass Concentration Due to Radial Inspiral

In this section we describe our test of the portion of the code that handles the size-dependent radial inspiral of particles. Because it is not easy to derive an analytic solution for the evolution of the surface density with a complicated radial dependence of the inspiral rate,
Figure 3.5: Comparison of theory from O’Brien & Greenberg (2003) (green) with the results from our code. The upper panel shows the number of particles per mass bin as a function of mass. The bottom panel shows the logarithmic slope of this relation. The right-most 40% of the mass bins are held fixed, so as not to introduce error due to the finite extent of our simulation in mass space. This is for $\alpha = 2$ and $s = 0$. 
we made this test for the simple inspiral rate given by

$$v_{ij} = a_i^2/d_j \text{AU yr}^{-1},$$

(3.26)

where $a_i$ measured in AU yr$^{-1}$, and $d_j$ in centimeters. As stated in Section 3.6.5, $v_{i,j}$ is the inspiral velocity of a particle in annulus $i$ and mass bin $j$, here measured in AU yr$^{-1}$. Our initial mass distribution was given by

$$dN(a_i, t = 0) = \left(4 - \frac{a_i}{\text{AU}}\right) da$$

(3.27)

This leads to an evolution of the number density

$$dN(a_i, t) = \begin{cases} s^{-2} (4 - a/s) da_i & a_i < 1/(0.25 + d_i^{-1}t) \\ 0 & a_i > 1/(0.25 + d_i^{-1}t) \end{cases}$$

(3.28)

where $s = 1 - d_i^{-1}at$. Panels A, C, and E correspond to the cases with 16, 51, and 191 radial annuli between 1 and 4 AU respectively, and $\epsilon_{\text{drift}} = 1.0$. The corresponding bottom panels B, D, and F are for for $\epsilon_{\text{drift}} = 0.3$. We see that generally the solutions stay within several percent of the analytic solutions. The exceptions to this are where the analytic solution is zero. The simulation does not do a good job of resolving the sharp outer edge of the distribution because particles are effectively re-distributed through each annulus every time-step, leading to substantial numerical diffusion. The time step makes little difference to the results, but the agreement is a strong function of the number of annuli used.

We are ignoring interactions between particles in neighboring radial annuli. This is unlikely to be a significant issue unless there is a secular resonance. If a resonance is present in the disk, then particles in a narrow region of semi-major axis will be excited to high eccentricity, and may collide with and destroy particles at very different semi-major axes. We have not taken this into account in our treatment.
Figure 3.6: Comparison of the evolution of the surface density profile due to radial inspiral in the simulation (dotted lines) vs. the analytic result (solid lines) for the inspiral rate given in Equation (3.26).
3.8 Appendix C: Collisions Between Bodies in the Same Size-Bin

In some cases, the random velocities in mass bin \( i \) are small compared with both the escape speed for those bodies and with the velocity due to the relative eccentricity between the most massive bodies in bin \( i \) and the least massive. When this happens, then Equation (3.14) significantly overpredicts the collision rate between bodies in mass bin \( i \) with other bodies in mass bin \( i \). The gravitational focussing factor for these collisions is calculated for bodies of exactly the same size, whereas in reality bodies in the same mass bin do not all have exactly the same mass, and therefore approach each other at greater speed. To address this, we approximately re-calculated the collision rate between two bodies in the same mass bin by averaging over the masses of the two collision partners.

In general the collision rate \( R \) between one body and a population of collision partners can be expressed as a product of the effective cross-section \( \sigma \), the approach velocity \( v \) and the space number density \( n \) of collision partners.

\[
R = \sigma vn. \tag{3.29}
\]

Because of gravitational focussing, the cross-section \( \sigma \) depends on the velocity, and can be expressed as

\[
\sigma = A \left( 1 + \left[ \frac{v_0}{v} \right]^2 \right), \tag{3.30}
\]

where \( A \) is the geometric cross-section, \( v_0 \) the escape velocity when the two objects are touching, and \( v \) their relative velocity at infinity. Let \( v_{\text{max}} \) be the relative non-random velocity between the heaviest body in the mass bin and the lightest. Let \( x_1 = (m_1 - m_{\text{bottom}})/(m_{\text{top}} - m_{\text{bottom}}) \) be the position of body 1 in the mass bin, and define \( x_2 \) analogously for the second body. \( m_{\text{top}} \) and \( m_{\text{bottom}} \) are the masses of the heaviest and lightest bodies in the bin. The relative non-random velocity between particles 1 and 2 is equal to \( v_{\text{max}} |x_1 - x_2| \), and the
total relative velocity $v_{12}$ is given by
\[ v_{12} = \sqrt{v_{\text{rand}}^2 + ((x_1 - x_2)v_{\text{max}})^2} = v_{\text{rand}}\sqrt{1 + \alpha^2(x_1 - x_2)^2}, \] (3.31)
where $\alpha = v_{\text{max}}/v_{\text{rand}}$. Combining equations (3.29), (3.30) and (3.31), we can express the rate $R_{12}$
\[ R_{12} = Anv_{\text{rand}} \left[ \sqrt{1 + \alpha^2(x_1 - x_2)^2 + \frac{\beta^2}{\sqrt{1 + \alpha^2(x_1 - x_2)^2}}} \right], \] (3.32)
where $\beta = v_0/v_{\text{rand}}$. To calculate an average $\langle R \rangle$ over all possible parirs of colliding particles in mass bin $i$, we average over $x_1$ and $x_2$:
\[ \langle R \rangle = \int_{x_1=0}^{1} \int_{x_2=0}^{1} R_{ij} dx_1 dx_2 \]
\[ = \frac{Anv_{\text{rand}}}{3\alpha^2} \left( \alpha^2 \sqrt{1 + \alpha^2} + 2(1 - \sqrt{1 + \alpha^2})(1 + 3\beta^2) + 3\alpha(1 + 2\beta^2) \right) \sinh^{-1}(\alpha). \] (3.33)
Similarly, we can calculate the expected collision velocity as
\[ \langle v_{\text{coll}} \rangle = \frac{\int_{x_1=0}^{1} \int_{x_2=0}^{1} R_{ij} v_{\text{rand}} \sqrt{1 + \alpha^2(x_1 - x_2)^2} dx_1 dx_2}{\int_{x_1=0}^{1} \int_{x_2=0}^{1} R_{ij} dx_1 dx_2} \]
\[ = \frac{3\alpha^2(1 + \alpha^2/6 + \beta^2)v_{\text{rand}}}{\alpha^2 \sqrt{1 + \alpha^2} + 2(1 - \sqrt{1 + \alpha^2})(1 + 3\beta^2) + 3(\alpha + 2\alpha\beta^2) \sinh^{-1}(\alpha)}. \] (3.34)
Bibliography

Adachi, I., Hayashi, C., & Nakazawa, K. 1976, Progress of Theoretical Physics, 56, 1756


Smoluchowski, M. V. 1916, Zeitschrift fur Physik, 17, 557


Tanaka, H., Inaba, S., & Nakazawa, K. 1996, icarus, 123, 450


Trubnikov, B. A. 1971, Soviet Physics Doklady, 16, 124


Chapter 4

Modeling the Nearly Isotropic Comet Population in Anticipation of LSST Observations

4.1 Abstract

We run simulations to determine the expected distribution of orbital elements of nearly isotropic comets (NICs) in the outer solar system, assuming that these comets originate in the Oort cloud at thousands of AU and are perturbed into the planetary region by the Galactic tide. We show that the Large Synoptic Survey Telescope (LSST) should detect and characterize the orbits of hundreds to thousands of NICs with perihelion distance outside 5 AU. Observing NICs in the outer solar system is our only way of directly detecting comets from the inner Oort cloud, as these comets are dynamically excluded from the inner solar system by the giant planets. Thus the distribution of orbital elements constrains the spatial distribution of comets in the Oort cloud and the environment in which the solar system formed. Additionally, comet orbits can be characterized more precisely when they are seen far from the Sun as they have not been affected by non-gravitational forces.
4.2 Introduction

The distribution of cometary orbital elements in the Oort cloud depends on the dynamical evolution of the solar system’s planetesimal disk and the environment in which the solar system formed. Unfortunately, the vast majority of Oort cloud comets are unobservable, usually being seen only when they are perturbed onto orbits with perihelion $\lesssim 5$ AU. Furthermore, the orbits of visible comets have generally been modified by poorly understood non-gravitational forces (Yeomans et al., 2004). For both reasons it is difficult to infer the properties of the Oort cloud from the statistical distribution of comet orbits.

This paper describes the expected distribution of orbital elements of nearly isotropic comets (NICs). We define these to be comets that have been perturbed into the planetary region from the Oort cloud. Theoretical models of the distribution of NICs have been constructed by others e.g., Wiegert & Tremaine (1999), Levison et al. (2001) and Fouchard et al. (2013, 2014a,b); the novel feature of the present study is that we focus on much larger heliocentric distances (up to 45 AU) in anticipation of deep wide-angle sky surveys that are currently under development. We use a simple physical model that assumes a static spherical Oort cloud with comets uniformly distributed on a surface of constant energy in phase space. Models of the formation of the Oort cloud (e.g., Dones et al., 2004) suggest that this approximation is reasonable except perhaps at the smallest semi-major axis we examine, $a = 5,000$ AU, where the cloud is somewhat flattened. Furthermore, we assume that these comets evolve solely under the influence of the Galactic tide and perturbations from the giant planets. We do not consider the stochastic effects of passing stars, as they have effects qualitatively similar to the effects of the Galactic tide (Heisler et al., 1987, Collins & Sari, 2010); see Fouchard et al. (2013, 2014a,b) for a detailed comparison of the influence of these two agents on the Oort cloud. We follow the evolution of these comets through N-body simulations for up to 4.5 Gyr and use the results of these orbit integrations to construct a simulated comet catalog.
This topic is of special interest now as the Large Synoptic Survey Telescope (LSST) will likely see many of these NICs in the outer solar system. LSST will survey 20,000 square degrees of sky (48% of the sphere) about 2,000 times over 10 years, down to an $r$-band magnitude of 24.5 (LSST Science Collaboration et al., 2009). LSST has a flux limit 3.2 magnitudes fainter than, and more than three times the area of, the current leader in finding faint distant solar system objects — the Palomar Distant Solar System Survey (Schwamb et al., 2010). It is expected to find tens of thousands of trans-Neptunian objects (LSST Science Collaboration et al., 2009); however we are not aware of predictions made specifically for objects originating in the Oort cloud.

Francis (2005) studied the long-period ($P > 200$ years) comet population using the Lincoln Near-Earth Asteroid Research (LINEAR) survey (Stokes et al., 2000). Most observed long-period comets likely originated in the Oort cloud. He found a sample of 51 long-period comets which were either discovered by LINEAR or would have been, had they not previously been discovered by another group. Fifteen of these had perihelion distances beyond 5 AU, but none beyond 10 AU. He used this sample to estimate properties of the Oort cloud. He found a “suggestive” discrepancy between the distribution of cometary perihelion distances in the observed sample and in theoretical models (Tsujii, 1992; Wiegert & Tremaine, 1999), but cautioned that the difference could be the result of a poor understanding of the rate at which comets brighten as they approach the Sun due to cometary activity. LSST will address this question by observing many comets at large heliocentric radii where they are inactive (see Section 4.3.2).

Hills (1981) proposed that the apparent inner edge of the Oort cloud at around 10,000 AU is not due to a lack of comets at smaller semi-major axes, but rather because the perihelion distances of those comets evolve slowly, so they are ejected or evolve to even smaller semi-major axes due to perturbations from the outer planets before they become visible from Earth. In contrast, comets with semi-major axis $a \gtrsim 10,000$ AU have their perihelion distance changed by more than the radius of Saturn’s orbit in one orbital period,
so they are able to jump across the dynamical barrier of the outer planets, and be seen in the inner solar system (Hills 1981). This barrier is not 100% leak-proof, but as is demonstrated later in the paper, one expects the number density of comets with initial semi-major axes of 10,000 AU to decline by over two orders of magnitude interior to 10 AU. LSST should detect NICs at distances > 10-15 AU and so will enable us to estimate the population of this inner Oort cloud directly, because we will be able to see NICs outside the region of phase space from which they are excluded by the giant planets. The properties of this cloud may contain information about the density and mass distribution in the Sun’s birth cluster (Brasser et al., 2012).

Observing NICs far from the Sun also probes in unique ways the parts of the Oort cloud that do send comets near Earth. Non-gravitational forces due to outgassing when the comet comes near the Sun are the primary source of error in determining the original orbits of these comets (Yeomans et al., 2004). It is somewhat uncertain at what radius outgassing begins, but a reasonable estimate would be around 5 AU (see discussion in Section 4.3.2). Therefore, astrometric observations of comets beyond ∼ 10 AU should allow much more precise determination of their original orbits (see discussion at the end of Section 4.3.1).

### 4.3 Simulation Description

We divide phase space into three regions, based on the perihelion distance of the cometary orbit. We define the “visibility region” as containing orbits with perihelion distance \( q \) in the range \( 0 \text{ AU} < q < 45 \text{ AU} \). A “buffer region” includes orbits with \( 45 \text{ AU} < q < 60 \text{ AU} \). All other orbits are defined to be in the “Oort region”.

We simulated orbits with the Rebound package, developed by Rein & Liu (2012). We used their IAS15 integrator, a 15th order adaptive-timestep integrator that is sufficiently accurate to be symplectic to double precision (Rein & Spiegel 2015).
Our goal is to model the steady-state distribution of NICs with perihelia within 45 AU of the Sun that are produced from an Oort cloud with orbital elements uniformly distributed on an energy surface in phase space (so \(dN \sim \sin I dI de^2\), where \(I\) and \(e\) are the cometary inclination and eccentricity). This approximation assumes that perturbations from the Galactic tide, passing stars or molecular clouds over the last four Gyr have been sufficient to isotropize comets both in position space (seen from the Sun) and velocity space at a fixed position. This has been shown to be roughly true for comets with semi-major axes greater than 2,000 AU (Duncan et al., 1987).

To initialize the simulation, we generated comets at random from this phase-space distribution for four discrete values of initial semi-major axis, \(a_i = 5,000, 10,000, 20,000,\) and 50,000 AU, with perihelion distances in the range 60 AU to \(60 + \Delta\) AU. Then, using an analytic approximation to the torque from the Galactic tide (see Appendix A), we determined an upper bound on the time \(\tau_{\text{entry}}\) (as a function of \(a_i\)) such that no comet from outside \((60 + \Delta)\) AU could enter the buffer or visibility regions within the next \(\tau_{\text{entry}}\) years. We chose \(\Delta = 5\), but it is straightforward to see that the results do not depend on this choice. We picked \(\tau_{\text{entry}} = 10^7, 2.5 \cdot 10^6, 6.25 \cdot 10^5\) and \(10^5\) years, for \(a_i = 5,000, 10,000, 20,000\) and 50,000 AU respectively. These numbers satisfy our upper bound.

We then evolved the comets under the influence of the Sun, the Galactic tide, and the four outer planets for \(\tau_{\text{entry}}\). After \(\tau_{\text{entry}}\) had elapsed, we removed any comet with perihelion greater than 60 AU from the simulation. The remaining comets were allowed to evolve under the influence of the Galactic tide and gravity from the four giant planets and the Sun. At fixed intervals \(\tau_{\text{sample}}\) (taken to be 10 years), we recorded the position and velocity of any comet that was within 45 AU of the Sun in a catalog. This procedure gives us the same expected comet count and distribution of orbital elements as if we had allowed the system to evolve to steady state, and then catalogued the comets visible within 45 AU at an instant in time, and multiplied the flux by \(\tau_{\text{entry}}/\tau_{\text{sample}}\).
Comets are removed from the simulation if they collide with a planet, come within 0.1 AU of the Sun, move outside 200,000 AU, or are perturbed back into the Oort region ($q > 60$ AU). \footnote{Because the boundary between the buffer and Oort region at 60 AU corresponds to a perihelion distance twice the semi-major axis of Neptune’s orbit, we expect the planets to have a negligible effect on the orbits of comets in the Oort region. Therefore, it is reasonable to assume that a comet with an orbit aligned such that the Galactic tide pulls it from the buffer region into the Oort region will not return for a long time.}

### 4.3.1 Orbital Elements

The treatment of orbital elements for highly eccentric orbits that pass through the orbits of massive planets is somewhat subtle. A comet that is having a close encounter with one of the giant planets will undergo large short-term perturbations to its orbital elements that do not reflect changes to its orbit that will last longer than the duration of the encounter. Short-term perturbations from such encounters are more serious for comets with large semi-major axes because the energy of the comet in the planetary potential well can equal or exceed the total orbital binding energy of the comet. At distance $r_p$ from a planet with mass $M_p$, the fractional change in energy due to the potential energy of the planet is

$$\frac{\Delta E}{E} = \frac{2a_c M_p}{r_p M_\odot} = 0.19 \frac{a_c}{100 \text{ AU}} \frac{1 \text{ AU}}{r_p} \frac{M_p}{M_{\text{Jupiter}}},$$

(4.1)

where $a_c$ is the semi-major axis of the comet prior to the close encounter. We stress that $a_c$ and $a_i$ are not the same quantity: $a_c$ is the current semi-major axis of the comet, whereas $a_i$ is the semi-major axis of the comet when the simulation was initialized.

To lessen the short-term planetary perturbations to cometary orbital elements, we adopt the following procedure. For comets with $a_c < 100$ AU, we simply report heliocentric orbital elements. These comets have large enough binding energy that they would have to pass close to a planet (within 2.0 AU for Jupiter) to obtain enough extra kinetic energy for $a_c$ to vary by more than 10% during the close passage (see Equation (4.1)). In order to prevent very close planetary approaches from contaminating our results, we discarded any observation in
which the comet is currently close enough to a planet that the specific potential energy due to the planet is more than 10% of the specific binding energy in an orbit around the Sun with semi-major axis of 100 AU. This occurs for only 0.004% of all catalog entries, or 0.3% of all catalog entries with $R < 10$ AU.

Comets in the visibility region with $a_c > 100$ AU often have potential energies due to the planets which are comparable to their binding energies. For this reason, we report the barycentric orbital elements of the comet the last time it was in the range [90 AU, 110 AU]. These elements are well-behaved, since they are calculated far outside the orbits of the giant planets where it is appropriate to represent the solar system as having all its mass located at the barycenter.

The ease with which LSST can determine orbital elements for slowly moving nearly unbound objects is also of interest to this study. To address this question, we searched the JPL small body database\footnote{http://ssd.jpl.nasa.gov/?horizons} for objects with semi-major axis greater than 300 AU and perihelion distance greater than 10 AU. It listed seven objects with a data-arc longer than 5 years. The estimated errors in $x = 1/a$ ranged from $1.5 \cdot 10^{-6}$ AU$^{-1}$ to $1.5 \cdot 10^{-5}$ AU$^{-1}$ for these objects. It therefore seems reasonable to expect orbits to generally be determined to at least this level of accuracy purely from 10 years of LSST data.

### 4.3.2 Disrupted Comets

There is a substantial body of evidence suggesting that comets “fade” over time (e.g., Fernández [1981]; Wiegert & Tremaine [1999]). A number of processes have been proposed to explain this phenomenon (Weissman [1980]): a comet could run out of volatile material, it could have its surface covered with a crust that prevents volatiles from escaping, or it could physically be broken apart by outgassing or tidal stress.

Fernández (2005) gives 3 AU as a likely cut-off to comet activity based on the sublimation temperature of water, but cautions that many comets experience some activity outside 3 AU.
due to the sublimation of more volatile elements. Comet 67P/Churyumov-Gerasimenko first showed signs of activity when it was 4.3 AU from the Sun (Snodgrass et al., 2013). Comet Hale-Bopp showed substantial activity on its approach to perihelion as far out as 7.2 AU (Weaver et al., 1997), and at 27.2 AU after perihelion passage (Kramer et al., 2014). When we calculate numbers of visible NICs, we restrict ourselves to NICs further than 5 AU from the Sun. For this reason, we do not consider the effect of comet activity on magnitude, and just calculate the magnitude from the size and albedo of the bare nucleus, see Section 4.6.2. We believe that our assumption that there is negligible activity beyond 5 AU is reasonable, though not certain, given existing observations. In any event this assumption gives us a conservative estimate of the number of comets that a survey like LSST will discover.

Because comet activity does not affect brightness in our model, we are only sensitive to physical disruption of comets, not loss of volatiles. For this reason, throughout this paper, we refer to “disruption”, rather than “fading”.

In the results that follow, we remove a comet after it has made 10 apparitions in the catalog with radius $R < 3$ AU (corresponding to a total exposure to the Sun at $R < 3$ AU of about 100 years, since $\tau_{\text{sample}} = 10$ years). Comets are also removed if they ever travel within 0.1 AU of the Sun (even if they do not appear so close in the catalogue) or if they suffer a collision with one of the gas giants.

4.4 Comet Lifetimes in the Simulation Region

Figure 4.1 shows the fraction of NICs in our simulations with $a_c > 34.2$ AU (period $> 200$ years) appearing in the region with $R < 45$ AU for more than $t$ years, as a function of $t$. Different curves correspond to different values of the initial semi major axis $a_i$. In this plot we terminate each orbit integration after 4.5 Gyr. The error bars are derived from the re-sampling procedure described in Section 4.5.1. In this and all subsequent plots, only comets
with periods greater than 200 years (corresponding to semi-major axes greater than 34.2 AU) are counted.

Yabushita (1979) argued using a simple random walk model that the number of NICs surviving more than $N_{peri}$ perihelion passages should scale as $P(>N_{peri}) \propto N_{peri}^{-1/2}$. Assuming for the sake of argument that NICs spend a fixed amount of time in the visibility region per perihelion passage, then the number of NICs having a given number of catalog entries is proportional to the number of NICs surviving for more than a given number of perihelion passages. We should therefore recover the same power-law as Yabushita (1979) if his model is a good approximation to the full physics captured by the simulation. This plot largely confirms the predictions of Yabushita (1979), but the exponent of the power-law seems to be slightly steeper than his value of $-1/2$.

Deviations from power-law behavior at short times occur because the visibility region is larger than the region of influence of the planets, so there is some delay before NICs that have entered the visibility region interact with the planets. Comets with smaller values of $a_i$ experience less torque due to the Galactic tide, so the delay is larger. They also have more binding energy that must be overcome prior to ejection. This explains the trend seen in Figure 4.1 that NICs with smaller $a_i$ take longer to be ejected.

The fact that some of our simulated particles survive for longer than 1 Gyr leads to concern about the physical validity of our assumption that there is a static Oort cloud. Likely, many of the NICs that will be observed with LSST exited the Oort cloud more than 1 Gyr in the past, when it may have had different physical properties.

Even if the properties of the Oort cloud have not changed over 4.5 Gyr, long-lived comets may bias our simulations, as the following argument shows. Suppose, as seems reasonable from Figure 4.1 that the true distribution of time $t$ that a comet spends in the visibility
region is given by a power law, i.e.

\[
\frac{dp}{dt} = \begin{cases} 
\frac{(\alpha-1)t^{-\alpha}}{t_{\text{min}}^{1-\alpha}} & t > t_{\text{min}} \\
0 & t < t_{\text{min}}
\end{cases} \tag{4.2}
\]

for some \( \alpha \) in the interval \( 1 < \alpha < 2 \). Then, if we terminate our integrations at some time \( t_{\text{cutoff}} \), the average time that a comet spends in the visibility region during our simulation is

\[
\langle t \rangle = \frac{1}{2 - \alpha} \left[ t_{\text{cutoff}} \left( \frac{t_{\text{min}}}{t_{\text{cutoff}}} \right)^{\alpha-1} + (1 - \alpha)t_{\text{min}} \right] \tag{4.3}
\]

In the limit that \( t_{\text{cutoff}} \gg t_{\text{min}} \), this becomes

\[
\langle t \rangle = \frac{t_{\text{cutoff}}}{2 - \alpha} \left( \frac{t_{\text{min}}}{t_{\text{cutoff}}} \right)^{\alpha-1} \tag{4.4}
\]

In our simulations, \((8, 1, 0, 0)\) particles with \( a_i = (5,000, 10,000, 20,000, 50,000) \) AU survive with \( q < 45 \) AU for the duration of the integration (4.5 Gyr). For \( a_i = 20,000 \) AU and 50,000 AU respectively, the longest lived particles survived for 1.1 and 1.4 Gyr. Now, consider the case of the simulation with \( a_i = 20,000 \) AU. We drew particles with lifetimes from the distribution in Equation (4.2). By chance, we drew no particles with lifetime greater than \( t_{\text{max}} = 1.1 \) Gyr. We will assume that we drew randomly from the distribution in Equation (4.2) subject to the constraint that we draw no particles with lifetimes greater than \( t_{\text{max}} \) (this is not quite true, given that we did draw one particle with lifetime exactly \( t_{\text{max}} \)). We can renormalize the distribution in Equation (4.2) subject to the constraint that no comet have lifetime longer than \( t_{\text{max}} \), and calculate the mean. This is given by (assuming \( t_{\text{max}} \gg t_{\text{min}} \))

\[
\langle t \rangle = \frac{\alpha - 1}{2 - \alpha} \left[ t_{\text{max}} \left( \frac{t_{\text{min}}}{t_{\text{max}}} \right)^{\alpha-1} \right] \tag{4.5}
\]

Taking the ratio of Equations (4.4) to (4.5), and assuming that a comet makes a fixed number of appearances in the catalog per unit time spent in the visibility region, we find we have
likely underestimated our total flux by something close to a factor of $(4.5/1.1)^{2-\alpha}/(\alpha - 1)$. This expression is 4.0 if we take $\alpha = 1.5$, as suggested by Yabushita (1979), but declines to unity for $\alpha = 2.0$. In principle, this bias can be reduced by simulating more comets but the computational resources necessary to reduce the bias substantially are prohibitive.

In the following sections we present simulation data for different values of $t_{\text{cutoff}}$. $t_{\text{cutoff}}$ is defined relative to the time when the comet first enters within 45 AU except when $t_{\text{cutoff}} = 4.5$ Gyr, in which case it is defined relative to the start of the simulation. We believe that a value of a few Gyr is most observationally relevant, and most of the following discussion is based on such cutoff times, however given the limitations discussed above, we show results for shorter cutoff times as well.

### 4.5 Concentration of NICs Due to the Giant Planets

In this section we describe how the distribution of NICs is affected by the planets.

#### 4.5.1 Expected Distribution of $R$ and $q$ with no Planets

We first calculate the expected distribution of orbital radius $R$ and perihelion distance $q$ in the absence of perturbations from the planets, constructing what we call the zero-planet model. We assume a uniform distribution of cometary orbital elements in phase space at four fixed energies. These results provide a natural normalization to the plots in the following subsection, which show the distributions of $R$ and $q$ in our simulated catalog.

**Distribution in Radius and Perihelion at a Snapshot in Time**

Let there be $N_0$ comets distributed uniformly on the energy surface corresponding to semi-major axis $a$. There is no need to distinguish between the initial semi-major axis $a_i$ and the current semi-major axis $a_c$ here, since the semi-major axis is not changed in the absence of planetary perturbations. Approximating the orbits as parabolic in the visibility region, we
Figure 4.1: Distribution of the number of NICs having $R < 45$ AU for longer than $t$ years. The curves show four different values of the initial semi-major axis $a_i$, normalized by the number of comets in the simulation which ever enter within $R = 45$ AU. We terminate the integrations after 4.5 Gyr. The black lines show two power laws with exponents $-0.5$ and $-1$. Only comets with $a_c > 34.2$ AU (corresponding to long-period comets — comets with period $> 200$ years) were counted.
find that the radial velocity at radius \( R \) of an orbit with perihelion distance \( q \) is

\[
v_R(q, R) = \sqrt{\frac{2GM_\odot(1 - q/R)}{R}}. \tag{4.6}\]

Since we assume a uniform distribution of comets on the energy surface, the probability density of the squared eccentricity \( e^2 \) is

\[
N(e^2)de^2 = N_0de^2. \tag{4.7}\]

Then, since \( q = a(1 - e) \), the number of comets per unit perihelion distance \( N(q) \) is given by

\[
N(q) dq = 2N_0 \left( 1 - \frac{q}{a} \right) \frac{dq}{a} \sim \frac{2N_0}{a} dq, \tag{4.8}\]

where the last equality holds because we are interested in comets with \( q \ll a \). A comet on a near-parabolic orbit with perihelion \( q \) spends a fraction of its time \( f(R, q)dR \) in the radial interval between \( R \) and \( R + dR \), where

\[
f(R, q)dR = \frac{2dR}{Pv_R(q, R)}, \tag{4.9}\]

and \( P \) is the period of the orbit. Then, using Equations (4.8) and (4.9), we can solve for the number of comets in a radial interval \( N(R)dR \):

\[
N(R) = \int_{q=0}^{R} N(q)f(R, q)dq = \frac{2\sqrt{2}N_0R^{3/2}}{\pi a^{5/2}}. \tag{4.10}\]

The number of comets with perihelion in the range \( q \) to \( q + dq \) expected to be present out to a maximum value of \( R \) is given by

\[
N(q|R_{\text{max}})dq = N(q)dq \int_{q}^{R_{\text{max}}} f(R', q)dR'. \tag{4.11}\]
This evaluates to
\[ N(q|R_{\text{max}})dq = \frac{2\sqrt{2}}{3\pi a^{5/2}} \left( \frac{R_{\text{max}}}{q} - 1 \right)^{1/2} \left( 2 + \frac{R_{\text{max}}}{q} \right) q^{3/2} dq. \] (4.12)

As mentioned in Section 4.3, because of the way we have set up our simulation, and the fact that we sample every \( \tau_{\text{sample}} \) years, we would expect our catalog to contain a number of comets with orbital elements \( \psi \) equal to
\[ N_{\text{cat}}(\psi|R_{\text{max}}) = \frac{\tau_{\text{entry}}}{\tau_{\text{sample}}} N(\psi|R_{\text{max}}), \] (4.13)
if we had not included any planets in the simulation.

**Concentration Factors**

In this subsection we compare our catalog to the one which would be produced had we not included planets in our simulations. In Figure 4.2, we compare our simulated comet catalog with the zero-planet model (Equations (4.10) and (4.13)). We have plotted a “concentration factor” — the ratio of comets appearing in our catalog within a given radial interval to the number calculated from Equations (4.10) and (4.13), assuming the same density of comets in the Oort cloud as was used to initialize our simulations. As stated previously, only comets with periods greater than 200 years are counted. Each panel corresponds to a different value of the initial cometary semi-major axis \( a_i \). Different colored lines correspond to different values of \( \tau_{\text{cutoff}} \) as shown in the legend.

These data represent the results from following \( N_{\text{sim}} \) comets where \( N_{\text{sim}} = 10,000, 16,000, 40,000, \) and \( 420,000 \) for \( a_i = 5,000, 10,000, \) and \( 20,000 \) AU, and \( 50,000 \) AU respectively. Note that fewer than 10% of these ever enter the region \( R < 45 \) AU (955, 1548, 3949, and 29215 respectively). Most comets do not evolve to \( q < 60 \) AU within the entry period and are therefore discarded at the end of the entry period. Some of those that do come within
$q < 60$ AU never reach $q < 45$ AU (if the angular momentum is nearly perpendicular to the torque).

We would like to know the true distribution of orbital elements in the limit that we simulate a very large number of comets. To estimate our random error, we employed re-sampling. For each point on the curve, we drew $N_{\text{sim}}$ comets with replacement from our $N_{\text{sim}}$ simulated comets. The points are the mean of the re-sampled distribution, and the error bars correspond to the 16th and 84th percentiles of the re-sampled distribution (if the distribution were normally distributed, these would be 1-sigma error bars). The error bars on nearby points are highly correlated in most of our plots because the same comet contributes to several bins in the course of its evolution, hence the low point-to-point scatter relative to the error bars.

Although these data reflect the orbits of thousands of comets, the statistical errors are large in many cases, since often the majority of the contribution to a particular bin comes from only one or two long-lived comets, (see the discussion in §4.4).

Figure 4.3 is similar to Figure 4.2 except that instead of plotting the number of appearances in our catalog in bins of heliocentric radius $R$, we plot appearances in bins of perihelion $q$. The distribution of $q$ provides more information about the NIC orbits: a sharp feature in the perihelion distribution is smoothed out when one looks at the distribution of heliocentric radius, since the same orbit can be observed at a range of values of $R$.

In Figures 4.2 and 4.3 and in most of the subsequent plots, we have horizontally offset the blue, green, yellow and red curves by $-4.5\%$, $-1.5\%$, $1.5\%$, and $4.5\%$ respectively in order to make the curves distinguishable in regions where they overlap.

The following qualitative features of Figures 4.2 and 4.3 are straightforward to explain:

- The flux of comets with $a_i \lesssim 20,000$ AU shows a sharp drop-off (in both $N(q)$ and $N(R)$) interior to 10 AU. This is because comets originating at small semi-major axes are subjected to weak Galactic tides and change their perihelia slowly. The majority of kicks given to a comet with $q < 10$ AU are large enough to either unbind a comet
Figure 4.2: Number of comets in the simulated catalog at radius $R$, normalized by the expected number assuming no planets (Equations (4.10), (4.13)). Different curves correspond to different values of the cutoff time $\tau_{\text{cutoff}}$. Errors were determined via bootstrapping (see Section 4.5.1 for details). Points and error bars from curves corresponding to different cutoff times have been horizontally displaced slightly for clarity.
Figure 4.3: Number of comets in the catalog at perihelion $q$, normalized by the expected number assuming no planets (Equations (4.12), (4.13)). Different curves correspond to different values of the cutoff time $\tau_{\text{cutoff}}$. 

Concentration factor $a_i = 5,000$ AU

Concentration factor $a_i = 10,000$ AU

Concentration factor $a_i = 20,000$ AU

Concentration factor $a_i = 50,000$ AU

Legend:
- Blue line: $10^7$ years
- Orange line: $10^9$ years
- Green line: $10^8$ years
- Red line: $4.5 \times 10^9$ years
infalling from outside a few thousand AU, or to reduce \(a_i\) to the point that the timescale to change the perihelion distance is much longer than an orbital period. We therefore expect to see a jump in the number of comets appearing inside 5 AU at values of \(a_i\) exceeding that at which a comet can go from perihelion \(\gtrsim 15\) AU (largely unaffected by Jupiter and Saturn) to perihelion \(\lesssim 5\) AU in one orbit. This occurs at approximately 30,000 AU. Therefore, in order for a comet starting from inside \(\sim 30,000\) AU to appear in the inner solar system, it needs to have a lucky orientation with respect to Jupiter and Saturn\(^3\). This lucky orientation can either yield multiple small energy kicks on subsequent perihelion passages, or yield a kick that increases the semi-major axis so that the comet receives a larger torque from the Galaxy (Kaib & Quinn, 2009).

Thus we conclude that comets with \(a_i \lesssim 20,000\) AU are mostly ejected by interactions with the outer planets before they reach small heliocentric radii. Hills (1981) arrives at a similar result considering the effects of passing stars discretely rather than as a smooth Galactic tidal potential. Collins & Sari (2010) provide a discussion of when it is appropriate to treat the influence of the Galaxy as being due to a smooth tidal field, and when it should be modeled as discrete stellar encounters.

The lower concentration factors for comets with smaller values of \(a_i\) do not mean that we will see fewer NICs for a given number of Oort cloud comets at that energy. This is because the fraction of the total comets that are in the visibility region at a given time in the zero-planet model scales with \(a^{-2.5}\) (see Equation (4.10)).

- The difference between the green curves and the orange and red curves grows with increasing \(q\). This is because the kicks from the planets are smaller at large \(q\), so the comets survive longer. The exception to this is the curve for \(\tau_{\text{cutoff}} = 100\) Myr and \(a_i = 5,000\) AU. In this case, comets have mostly had insufficient time to be torqued to small values of \(q\). It should be noted however, that the time to reach a given perihelion distance is not completely determined by the initial semi-major axis because a comet

\(^3\)Or a kick from a star that passes unusually close to the Sun, a rare event not included in our model.
could be scattered to larger $a_i$ by an early encounter with Neptune, and subsequently evolve more rapidly.

There is little difference between the results for $t_{\text{cutoff}} = 1 \text{ Gyr}$, and $t_{\text{cutoff}} = 4.5 \text{ Gyr}$ for comets with $a_i \geq 20,000 \text{ AU}$. As discussed previously, this is likely an artifact of our simulations including insufficiently many comets with these values of $a_i$ to capture the tail of the lifetime distribution (see Section 4.4).

- The concentration factors for large cutoff time in Figure 4.3 approach unity as $q$ approaches 45 AU, however the concentration factors in Figure 4.2 are still on the order of 10 at 45 AU. This is because even a concentration of comets with $q \ll 45 \text{ AU}$ affects the distribution of comets at $R = 45 \text{ AU}$.

- We do not expect $N(q)$ to drop exactly to unity as soon as $q$ is larger than the extent of the planetary perturbations, because NICs which have interacted with the planets may be systematically carried away from the planetary region by the tide at a different rate than they were carried in (due to a change in orientation or semi-major axis).

### 4.6 Distribution of Orbital Elements for Visible Comets

The above analysis shows the degree to which the giant planets concentrate NICs in the outer solar system and exclude them from inside the orbit of Jupiter. In this section, we use an estimate of the size distribution of NICs, the relationship between magnitude, size and heliocentric distance, and the concentration effect due to interactions with the planets to calculate the number of NICs expected to be seen in an all-sky survey as a function of the limiting magnitude $m_{\text{lim}}$. 
4.6.1 Size Distribution

Comets have so far been observed primarily within a few AU of the Sun, where their brightness is influenced strongly by their activity. At the larger distances that we focus on here, comets are believed to be generally inactive (see discussion in Section 4.3.2), so their brightness is determined solely by their size, distance, and albedo. Let $H$ be the apparent magnitude of a comet 1 AU from the Sun and 1 AU from the observer, seen from zero phase angle. Based on a sample of long-period comets from about $H = 5$ to $H = 9$, Sosa & Fernández (2011) derive a relation for active comets between the radius $r$ (in kilometers) and $H$:

\[
\ln r = \alpha + \beta H, \tag{4.14}
\]

where $\alpha = 2.072$ and $\beta = -0.2993$. We caution that our use of this formula requires substantial extrapolation: the largest comet used to determine the formula has a radius of 1.8 km, more than an order of magnitude smaller than the smallest comets detectable at 30 AU in a survey with the limiting magnitude of LSST (see Section 4.6.2). Sosa & Fernández (2011) note that the relation in Equation (4.14) predicts a radius (13 km) for comet Hale-Bopp that is somewhat below other estimates (mostly falling in the 20–35 km range). This discrepancy suggests that Equation (4.14) may underpredict the radii of large comets, in which case our estimates of the observable comet population will be conservative. Note that by using Equation (4.14) we are assuming that long-period comets are mostly the same population as NICs (or at least have the same size distribution).

Hughes (2001) finds that the number $N_{\text{peri}}$ of long-period comets with brightness $H < 6.5$ passing through perihelion in the inner solar system per year per AU of perihelion distance is given by

\[
\frac{dN_{\text{peri}}}{dH} = c_0 e^{\gamma H}, \tag{4.15}
\]
with \( c_0 = 2.047 \cdot 10^{-3} \) and \( \gamma = 0.827 \). We can then transform variables to \( r \) using Equation (4.14). We find that

\[
dN_{\text{peri}}/dr = -\frac{c_0}{\beta} e^{-\frac{\gamma \alpha}{\beta} r^{\frac{\gamma}{\beta} - 1}} = 2.09 \cdot r^{-3.76}. \tag{4.16}
\]

This distribution holds down to \( r(H = 6.5) = 1.1 \text{ km} \), however, for simplicity we extrapolate it down to 0.9 km — the smallest comet visible at 5 AU in our model (see next section). This size distribution leads to a weak divergence in total mass at the large end of the spectrum. Nevertheless, we assume that the power-law behavior holds up to several tens of kilometers. The size distribution in Equation (4.16) is steeper than the relation \( (dN_{\text{peri}}/dr \sim r^{-2.79}) \) estimated in Hughes (2001) because he uses a different relation between \( H \) and \( r \). It is also substantially steeper than the relation \( (dN_{\text{peri}}/dr \sim r^{-2.92}) \) found in Snodgrass et al. (2011) for the Jupiter-family comets. If the size distribution is shallower than we have estimated at large radii, then our estimates of the observable comet population will be conservative.

### 4.6.2 Visibility Model

In this section we describe our model for determining how likely a given simulated comet is to be visible. We assume that comets have an \( r \)-band geometric albedo \( A_g \) of 0.04 as suggested in Lamy et al. (2004). We find that the magnitude \( m \) of an inactive comet is

\[
m = -27.08 - 2.5 \log \left( \frac{r^2 A_g (\text{AU})^2}{R^4} \right)
= 24.28 - 2.5 \log (A_g/0.04) - 5 \log (r/1\text{km}) + 10 \log (R/5\text{AU}), \tag{4.17}
\]

where we have used \(-27.08\) as the apparent \( r \)-band magnitude of the Sun. Equation (4.17) is only valid for comets far from the Sun, since we have assumed that \( R_{\text{Sun,comet}} = R_{\text{Earth,comet}} \), that the phase angle is zero, and most importantly, that we are seeing the bare nucleus of
an inactive comet. \cite{Lamy2004} state that magnitude drops off at about a rate of 0.04 magnitudes/degree of phase angle, meaning that error due to this nonzero phase angle is limited to at most 0.23 magnitudes for a comet at 10 AU. Similarly, at 10 AU, the largest possible error arising from the approximation that the Sun-comet distance is equal to the Earth-comet distance also corresponds to a magnitude error of $\Delta m = 0.23$.

### 4.6.3 Weighting of Observations

Using Equation \eqref{rmin}, we can solve for $r_{\text{min}}(R)$, the radius in kilometers of the smallest comet visible at distance $R$ in a survey with limiting magnitude $m_{\text{lim}}$, assuming $A_g = 0.04$:

$$r_{\text{min}}(R) = 0.903 \cdot 10^{0.2(24.5-m_{\text{lim}})} \left( \frac{R}{5 \text{ AU}} \right)^2.$$ \hspace{1cm} (4.18)

The comet size is not explicitly tracked in our simulations. We assume that the sizes of comets in our simulation are drawn from the distribution in Equation \eqref{size_distribution}, with a lower cutoff radius of 0.903 km — the smallest comet visible at 5 AU in our model. To account for the fact that not all comets are visible at all orbital radii, we weight a simulated comet appearance at radius $R$ by the fraction of comets that would be visible at the observed value of $R$ given the assumed size distribution in the simulation. An appearance at high $R$ will receive a low weight, since most comets would not be visible so far away. We assign weight 1 to observations at $R = 5$ AU, as all comets in our assumed size distribution would be visible at 5 AU. Then at general $R$, we assign weight

$$W(R) = \int_{r_{\text{min}}(R)}^{\infty} r^{-3.76} dr \int_{r_{\text{min}}(5 \text{ AU})}^{\infty} r^{-3.76} dr = \left( \frac{5 \text{ AU}}{R} \right)^{5.52}.$$ \hspace{1cm} (4.19)

We quote numbers of comets with a given set of orbital elements per $10^{11}$ comets larger than one kilometer in a spherical distribution at the assumed initial semi-major axis. To
achieve this normalization, we multiply our counts by

$$\frac{10^{11} f_{60-65}(a_i) \tau_{\text{sample}}}{N_{\text{init}}} \int_{0.903}^\infty \frac{N(r)dr}{\tau_{\text{entry}} \int_1^\infty N(r)dr},$$

\[(4.20)\]

where \(f_{60-65}(a_i)\) is the fraction of the phase space of orbits with semi-major axis \(a_i\) that consists of orbits with perihelia between 60 and 65 AU, given by

$$f_{60-65}(a_i) = \frac{(a_i - 60 \text{ AU})^2 - (a_i - 65 \text{ AU})^2}{a_i^2},$$

\[(4.21)\]

and \(N_{\text{init}}\) is the number of comets we initialize between 60 and 65 AU.

### 4.6.4 Distribution of Visible Comets

In this section we present results from our simulations showing how many NICs are visible over the whole sky at a given snapshot of time as a function of \(R\) and \(q\), for observations taken between \(R = 5\) AU and \(R = 45\) AU. In all cases, we assume a limiting \(r\)-band magnitude \(m_{\text{lim}}\) of 24.5, equivalent to the one-exposure limit for LSST (LSST Science Collaboration et al., 2009). The number of distant NICs expected to be discovered by LSST differs from the results presented here for two reasons. First, LSST is expected to operate for 10 years, so it should see more than just the comets visible in a snapshot, particularly in the case of the closer comets where \(R/v\) is less than 10 years. Second, LSST will only survey 48% of the sky, so will only see about half of the comets that would be visible in an equivalent all-sky survey. Comets move slowly enough that trailing losses will be insignificant given the 30 second exposure time. Using Equation (8) from Ivezic et al. (2008), we estimate a comet at 10 AU will have a limiting magnitude only 0.06 magnitudes brighter due to trailing losses.

Figures 4.4 and 4.5 show the number \(N\) of NICs expected to be visible outside 5 AU per unit of \(\ln R\) and \(\ln q\) respectively, per \(10^{11}\) comets with \(r\) greater than 1 km at the labeled initial semi-major axis in the Oort cloud. The shapes of the curves are substantially different for different values of \(a_i\), particularly in the region between 5 and 10 AU, where the statistics
Figure 4.4: Number of NICs expected to be seen per logarithmic interval in $R$ at a snapshot of time in an all-sky survey with limiting $r$-band magnitude $m_{\text{lim}} = 24.5$. This assumes there are $10^{11}$ comets with $r > 1$ km at the value of initial semi-major axis $a_i$ specified in each panel. Errors were determined via bootstrapping (see Section 4.5.1 for details). Different curves correspond to different values of $t_{\text{cutoff}}$. 
Figure 4.5: Number of NICs expected to be seen per logarithmic interval in $q$ at a snapshot of time. This assumes there are $10^{11}$ comets with $r > 1$ km at the value of $a_i$ specified in each panel. Different curves correspond to different values of $t_{\text{cutoff}}$. 

$10^7$ years, $10^9$ years, $10^8$ years, $4.5 \cdot 10^9$ years
are the best. In the \( a_i = 5,000 \) AU case, the expected count \textit{increases} by a factor of 5 from \( R = 5 \) AU to \( R = 10 \) AU for \( t_{\text{cutoff}} \geq 1 \) Gyr. In the \( a_i = 50,000 \) AU case it \textit{decreases} by a factor of \( \approx 5 \). Observations of comets in this regime will therefore allow us to observationally constrain the distribution of \( a_i \). The peak of \( RdN/dR \) moves smoothly from around 15 AU for \( a_i = 5,000 \) AU to less than 5 AU for \( a_i = 50,000 \) AU.

As shown in Appendix B, we expect \( RdN/dR \) and \( qdN/dq \) to decline as \( R^{-3.02} \) and \( q^{-3.02} \) respectively in the zero-planet model. Deviations from this behavior are due to variation in the concentration factor as shown in Figures 4.2 and 4.3.

We also examined what happened if we broke up the sample into two groups depending on the current semi-major axis of the comet. Figure 4.6 is identical to Figure 4.5 except that we have only considered those comets that have semi-major axes greater than 300 AU. The error bars are smaller, because the comets with the most appearances tend to diffuse to smaller values of \( a_c \), leaving a population with less spread in number of appearances. For this reason, this subset of comets, although smaller in number, has more power to discriminate between Oort cloud models. In Figure 4.7 we plot \( qdN/dq \) for only the comets in Figure 4.5 but not Figure 4.6, i.e., those comets whose orbits have \( a_c < 300 \) AU. We see that \( qdN/dq \) declines more sharply with \( q \) than in the whole sample of long-period comets. This is because it is difficult for a comet to attain \( a_c < 300 \) AU at large perihelion, because the kicks are too small. We also note that the overall numbers are larger by a factor of a few for comets with \( a_c < 300 \) AU.

\subsection*{4.6.5 Distribution in Semi-major Axis}

Figure 4.8 shows the semi-major axis distribution (number of appearances per unit logarithmic interval in semi-major axis) for all the NICs in a given perihelion bin (denoted by the color of the curve) and initial semi-major axis (panel). The error bars are generally larger for the points at small semi-major axis, implying that the statistics in these bins are dominated by a few comets.
Figure 4.6: Number of NICs with $a > 300$ AU expected to be seen per logarithmic interval in $q$ at a snapshot of time. This assumes there are $10^{11}$ comets with $r > 1$ km at the value of $a_i$ specified in each panel.
Figure 4.7: Number of NICs with $a < 300$ AU expected to be seen per logarithmic interval in $q$ at a snapshot of time. This assumes there are $10^{11}$ NICs with $r > 1$ km at the value of $a_i$ specified in each panel.
Figure 4.8: Number of NIC appearances per logarithmic interval in semi-major axis for different values of perihelion distance (different color curves) and different values of the initial semi-major axis (different panels). Each panel assumes that the Oort cloud contains $10^{11}$ comets with $r > 1$ km at the specified value of $a_{i}$. 
Figure 4.9: Number of NIC appearances per linear interval in inverse semi-major axis $x = 1/a$ for different values of perihelion distance (different colors) and different values of $a_i$ (different panels). Each panel assumes that the Oort cloud contains $10^{11}$ comets with $r > 1$ km at the specified value of $a_i$. 

183
We find it more illuminating to make this plot in the coordinate $x = 1/a$, which is proportional to the energy. Doing so allows us to test the predictions of random walk models such as those in [Yabushita (1979)] and [Everhart (1976)]. In their simplest form, one can imagine a comet starting with energy $E = -\epsilon$. Every perihelion passage, it takes a step of size $\epsilon$ towards higher or lower energy. It is removed if $E = 0$ (it becomes unbound), or if $E < E_{sp}$, where $E_{sp}$ is the critical energy level to be re-classified as a short-period comet. We ignore the possibility that a short period orbit could be perturbed back to the long-period regime. Comets are injected near $x = 0$. In the limit that $-E_{sp}/\epsilon \gg 1$, one can show that a steady-state distribution of comet energies normalized by the rate of perihelion passage is given by a linear equation of the form

$$N(E) = k(E - E_{sp}), \quad (4.22)$$

where $k$ is a constant depending on the comet injection rate and the size of the kick.

We see some support for Equation (4.22) in Figure 4.9. We have only considered comets with $a_c > 1,000$ AU ($x < 0.001$ AU$^{-1}$). This is a small enough range that we would expect the curves to be nearly flat if Equation (4.22) were correct (since $E_{sp}$ is much more negative than the energies shown in Figure 4.9). The curves are generally flat for perihelion distances $5$ AU $< q < 30$ AU. Inside 5 AU, the error bars are too large to say much. Outside 30 AU, there is a definite trend favoring energies closer to 0, as would be expected since the diffusion rate at these perihelion distances is so slow that the steady-state energy distribution cannot be achieved.

We see evidence of a weak “Oort spike” when $a_i = 50,000$ AU - an excess of comets with $x < 10^{-4}$ AU$^{-1}$ interpreted as the result of the initial conditions of cometary orbits. This is seen only for comets with $q < 5$ AU, since for larger values of $q$, the spike is drowned out by the large numbers of comets which remain at more negative energies for many orbits. As a quantitative measure of the observed Oort spike, [Fernández & Sosa (2012)] find that 30% of
comets discovered since the year 1900 with $q < 1.3$ AU have $x < (10,000 \text{ AU})^{-1}$. We find that we cannot come close to reproducing this unless we assume some disruption. We applied a disruption law in which comets were removed after spending $\tau_{\text{disrupt}}$ years within 5 AU of the sun. For comets with $a_i = 50,000$ AU and $q < 5$ AU, the fraction seen with $x < 10^{-4}$ AU$^{-1}$ is $\{0.12 \pm 0.007, 0.04 \pm 0.004, 0.02 \pm 0.002\}$, for $\tau_{\text{disrupt}} = \{100, 1,000, 10,000 \text{ years}\}$. This confirms the well-known result that one cannot reproduce the Oort spike without some model for comet disruption or fading that removes comets after only a few appearances (e.g., Wiegert & Tremaine, 1999).

We also notice a broader “spike” of comets around the original energy for comets with $a_i = 5,000$ or 10,000 AU, and perihelion distances outside 30 AU. A typical kick in energy at 30 AU might be on the order of $\Delta x = 10^{-5}$ AU$^{-1}$ (Duncan et al., 1987), so comets with $a_i$ small enough to not be rapidly carried away by the tide will show a broad peak around their original energy at large perihelion distance.

### 4.6.6 Orbital Inclination

Figure 4.10 shows the orbital inclination distribution (in ecliptic coordinates) of the visible NICs. The longest period NICs are preferentially retrograde. In a random walk model, the density of comets varies inversely with the step size. The energy kick is 2-3 times larger for a prograde comet (Duncan et al., 1987), translating into an expected prograde fraction of 0.25 to 0.33.

We find that our error bars on the prograde fraction are large for the entire comet population, so for the following analysis we split the population into two groups depending on the semi-major axes of the comets. Once again we find that elements of comets with large values of $a_c$ have less statistical noise. We obtain prograde fractions among the visible comets of $0.27 \pm 0.03, 0.29 \pm 0.04, 0.29 \pm 0.03$, and $0.40 \pm 0.02$ for $a_i = 5,000, 10,000, 20,000$, and 50,000 AU respectively, for the comets with $a_c > 1,000$ AU. For comets with $a_c < 1,000$
Figure 4.10: Distribution of the cosine of the ecliptic inclination angle for different ranges of current semi-major axis (different color curves) and initial semi-major axis (different panels).
AU, we find prograde fractions of $0.42 \pm 0.15$, $0.26 \pm 0.13$, $0.32 \pm 0.09$, and $0.40 \pm 0.09$. These data are consistent with the random walk model.

While the simulation data agree with the random walk model, they contradict the observations. There is only a slight preference in the observational data for retrograde comets with high perihelion. 64 out of 110 comets (58%) with period greater than 200 years and perihelion greater than 5 AU in the database at http://ssd.jpl.nasa.gov/dat/ELEMENTS.COMET are retrograde.

4.6.7 Size Distribution

A bigger telescope enables us to see rare large comets because it can search more volume. It is impossible to say exactly what size distribution of comets to expect in the observed sample, however we can make an estimate based on extrapolation of the size distribution from Fernández & Sosa (2012). It is instructive to first consider the zero-planet model with a fixed power-law for the size distribution. In Appendix B, we calculate the number of NICs greater than a certain size $r$ expected to be seen in the zero-planet model. We find in Equation (4.32) that the number visible with size greater than $r$ is given by $N = 55.2/r^{1.50}$ assuming $10^{11}$ comets greater than 1 km at 10,000 AU, but the results depend sensitively on the assumed number density and semi-major axis. Additionally, due to the concentration at larger orbital radii, we actually expect to see objects much larger than one would infer from the zero-planet model used to derive Equation (4.32).

Figure 4.11 shows the expected number of observations of NICs larger than a given size in the distance range $5 \text{ AU} < R < 45 \text{ AU}$. This calculation assumes $10^{11}$ comets greater than 1 km at the given initial semi-major axis in the Oort cloud and slope of the size distribution given in Equation (4.16). Exclusion of comets with orbital radii less than 5 AU should have a negligible effect on the counts except at the smallest sizes, as the concentration factors are lower there, and it is a negligible fraction of the visible volume for the larger comets. Exclusion of comets with orbital radii greater than 45 AU has no effect for the size of comets
Figure 4.11: Number of predicted detections of bodies in the range $5 \text{ AU} < R < 45 \text{ AU}$ as a function of $r$ given the size distribution assumed in Equation (4.31). The blue and orange points are unmoved, and the green and red are shifted 3% left and right respectively.
that we have plotted, since none of them would be visible past 45 AU. We see that assuming
the size distribution from Equation (4.16) holds to these sizes, we observe on the order of
10 NICs greater than 80 km in size if the majority of the NICs are coming from $a_i = 5,000$
AU, and a few if the majority of the NICs are coming from $a_i \geq 10,000$ AU.

4.7 Conclusions

We simulated the evolution of NICs originating in the Oort cloud as they interact with the
giant planets. We used these simulations to create a catalog of simulated comet positions
and velocities. We observe different distributions of orbital elements including perihelion
distance, semi-major axis and inclination depending on the semi-major axes at which the
NICs originate. Observations by LSST will therefore let us determine the absolute numbers
of comets in the Oort cloud as a function of semi-major axis, and test Oort’s standard model
for the origin of comets. The distribution of NICs outside the orbits of Jupiter and Saturn
will provide direct evidence for the presence or absence of the hypothesized “inner Oort
cloud” corresponding to semi-major axes between 5,000 and 20,000 AU.

One surprising result is that we expect at least tens of percent of the comets observed
by LSST in the outer solar system to have been interacting with the giant planets for more
than 1 Gyr. This result makes the interpretation of the comet population detected by LSST
more difficult and interesting, since the population and spatial distribution of comets in the
Oort cloud almost certainly evolves on timescales of a few Gyr.

We will also get a better measurement of the Oort spike — the excess of comets in nearly
parabolic orbits — as we will be able to measure high-precision orbits in a regime where
comets are likely unaffected by non-gravitational forces. This will put constraints on models
of comet fading, as well as the original semi-major axes of comets.

We will furthermore be able to constrain the size distribution of Oort cloud comets out
conservatively to several tens of kilometers, and perhaps even to larger bodies, depending on
the size distribution and number density of comets. Our results are based on a relatively steep
slope for the size distribution of comets, \( dN \propto r^{-3.76} \) (Equation 4.16) and may substantially
underestimate the total number of comets that will be discovered at large distances by LSST.

We thank the referee, Ramon Brasser, for helpful and constructive comments and advice.

### 4.8 Appendix A: Maximum Torque on a Radial Orbit

We calculate the torque on a radial orbit from the Galactic tide. This calculation allows us
to exclude comets from the Oort cloud that will definitely not enter the buffer region during
the entry period \( \tau_{\text{entry}} \) as described in Section 4.3. The results are also used in Section 4.5.1.
While our method is approximate, we have checked that our prediction is conservative in
the sense that no comet can enter the buffer region which we predicted could not enter the
buffer region. For simplicity, we made the following approximations.

In calculating the torque, we assume that the orbit is completely radial, pointing in the
direction of apocenter. This is well-justified by the characteristic ratios of \( q/a_i \approx 60/20,000 \).
We ignore the components of the Galactic tidal field in the plane of the Galaxy, as they are
smaller than the out-of-plane \( (z) \) component by about a factor of 10.

Heisler & Tremaine (1986) give the tidal force (per unit mass) as

\[
F_z = \left[ -4\pi G \rho_0 + 2(B^2 - A^2) \right] z_{\text{gal}} \hat{z}_{\text{gal}},
\]

where \( \rho_0 = 0.09 \, \text{M}_\odot \, \text{pc}^{-3} \), and the Oort \( A \) and \( B \) constants are taken to be 14.6 km s\(^{-1}\) kpc\(^{-1}\) and \(-12.4 \, \text{km s}^{-1} \, \text{kpc}^{-1} \) respectively (Binney & Tremaine 2008). This leads to an
instantaneous torque given by

\[
\Gamma = \left[ -4\pi G \rho_0 + 2(B^2 - A^2) \right] \cdot \frac{R^2 \sin 2b}{2},
\]
where $b$ is the Galactic latitude of the comet (which is constant in the radial orbit approximation). The orbit averaged torque is calculated by noting that $\langle R^2 \rangle = 5a^2/2$.

4.9 Appendix B: Comets Visible Assuming a Uniform Distribution of Orbital Elements on an Energy Surface and no Planets

In this appendix, we calculate the number of comets observable at a snapshot in time as a function of heliocentric radius in the zero-planet model. We approximate comet orbits as parabolic in the observation region. If we have a population of comets with perihelion $q$, of which $s(R)$ pass perihelion per year that are large enough to be visible at $R$, then the density $N_{\text{vis}}(R)dR$ of comets visible in a radial interval at radius $R > q$, is

$$N_{\text{vis}}(R) = \frac{2s(R)}{v_R}, \quad (4.25)$$

where $v_R$ is the radial velocity (Equation (4.6)). Using Equations (4.16) and (4.18), we find that in the zero-planet model,

$$s(R) = -\frac{c_0}{\beta}e^{-\gamma\alpha/\beta} \int_{r_1}^{\infty} r^{\gamma/\beta - 1}dr = \frac{c_0}{\gamma}e^{-\gamma\alpha/\beta}r_1^{\gamma/\beta} R^{2\gamma/\beta} \quad (4.26)$$

where $r_1 = 0.0361 \cdot 10^{24.5-m_{\text{lim}}}/5$ is the size (in km) of a body that is marginally detectable at 1 AU in our model (see Equation (4.18)). Therefore, using Equations (4.6), (4.25) and (4.26) we can write the density of comets as an integral over $q$:

$$N_{\text{vis}}(R) = \frac{2c_0}{\gamma}e^{-\gamma\alpha/\beta}r_1^{\gamma/\beta} R^{2\gamma/\beta} \int_0^R \frac{\sqrt{R}dq}{2\pi \sqrt{2(1 - \frac{q}{R})}} =$$

$$\frac{\sqrt{2}c_0}{\pi\gamma}e^{-\gamma\alpha/\beta}r_1^{\gamma/\beta} R^{2\gamma/\beta+3/2} = 5.0 \cdot 10^{0.55(m_{\text{lim}}-24.5)} \left(\frac{R}{5}\right)^{-4.02}. \quad (4.27)$$
We also calculate how many comets we see as a function of perihelion given a magnitude cut at a given snapshot in time. We find that

\[ N_{\text{vis}}(q) = \frac{-c_0}{\beta} e^{\frac{-2\alpha}{\beta}} \int_{r_1 q^2}^{\infty} r^{\frac{3}{2}-1} \tau_{\text{vis}}(r, q) dr, \tag{4.28} \]

where \( \tau_{\text{vis}}(r, q) \) is the amount of time that a comet with radius \( r \) in kilometers and perihelion \( q \) in AU is visible during one perihelion passage. From Equation (4.6), we see that

\[ \tau_{\text{vis}}(r, q) = 2 \int_q^{R_{\text{max}}} dR(v_R)^{-1} = \frac{\sqrt{2}}{3\pi} (2q + R_{\text{max}}) \sqrt{R_{\text{max}} - q}, \tag{4.29} \]

where \( R_{\text{max}} = \sqrt{r/r_1} \) is the maximum heliocentric distance at which the comet is visible. Putting this together, we find that

\[ N_{\text{vis}}(q) = c_0 e^{-\gamma \alpha/\beta} \left[ \frac{r_1^3 \Gamma\left(\frac{3}{2} - \frac{2\gamma}{\beta}\right)}{\sqrt{2\pi} \gamma \Gamma(-1 - \frac{2\gamma}{\beta})} \right] q^{3/2+2\gamma/\beta} = 2.3 \cdot 10^{0.55(m_{\text{lim}} - 24.5)} \left( \frac{q}{5} \right)^{-4.02}. \tag{4.30} \]

Finally, we ask how many comets we expect to see above a certain size in the zero-planet model. In this calculation, we take the size distribution in Equation (4.16), but normalize the counts to those expected if there were to \( 10^{11} \) comets with semi-major axis \( a_i \) and radii greater than 1 km. A cloud of \( N \) comets with orbital elements uniformly distributed on a surface of constant energy in phase space will have \( 2/a_i^{5/2} \) comets passing perihelion per year per AU perihelion, for \( q \ll a_i \). Therefore, the correct normalization is

\[ dN_{\text{peri}}/dr = -\frac{2N \beta}{a_i^5 \gamma} r^{\frac{3}{2}-1} = 7.25 \cdot r^{-3.76} \frac{N}{10^{11}} \left( \frac{10^4 \text{ AU}}{a_i} \right)^{\frac{3}{2}}. \tag{4.31} \]
In this model, we can expect to see $N_L(r)$ comets larger than $r$ where $N_L$ is given by

$$N_L(r) = \int_{r'=r}^{\infty} \int_0^{R_{\text{max}}} \tau_{\text{vis}}(r', q) dq dr' =$$

$$\frac{16\sqrt{2}\beta^2 N r^{5/3 + \frac{\gamma}{\beta}}}{5\pi\gamma a_i^2 r_1^{3/2} (5\beta + 4\gamma)} = \frac{55.3}{r^{1.50}} \left( \frac{10^4 \text{ AU}}{a_i} \right)^{5/3} \cdot 10^{0.25(m_{\text{lim}}-24.5)}. \quad (4.32)$$

Setting this equal to 1, and solving for $r$, we find that we expect to see one comet larger than 14.4 km if $a_i = 10,000$ AU, but this estimate is clearly quite sensitive to the assumed number of comets and that value of $a_i$. 

193
Bibliography


Duncan, M., Quinn, T., & Tremaine, S. 1987, AJ, 94, 1330

Everhart, E. 1976, NASA Special Publication, 393, 445


—. 2014a, Icarus, 231, 110

194

Heisler, J., & Tremaine, S. 1986, Icar, 65, 13

Heisler, J., Tremaine, S., & Alcock, C. 1987, Icar, 70, 269


Kaib, N. A., & Quinn, T. 2009, Science, 325, 1234


Tsujii, T. 1992, Celestial Mechanics and Dynamical Astronomy, 54, 271


Wiegert, P., & Tremaine, S. 1999, Icar, 137, 84


Chapter 5

Producing Distant Planets by Mutual Scattering of Planetary Embryos

5.1 Abstract

It is likely that multiple bodies with masses between those of Mars and Earth (“planetary embryos”) formed in the outer planetesimal disk of the solar system. Some of these were scattered by the giant planets into orbits with semi-major axes of hundreds of AU. Mutual torques between the embryos may lift the perihelia of some of them beyond the orbit of Neptune, where they are no longer perturbed by the giant planets so their semi-major axes are frozen in place. We conduct N-body simulations of this process, and its effect on smaller planetesimals in the region of the giant planets and the Kuiper Belt. We find that (i) there is a significant probability that one, or possibly more, sub-Earth mass embryos is still present in the solar system; (ii) the orbit of the surviving embryo(s) typically has perihelion of 40–70 AU, semi-major axis less than 200 AU, and inclination less than 30°; (iii) it is likely that any surviving embryos could be detected by current or planned optical surveys or have a significant effect on solar-system ephemerides; (iv) whether or not an embryo has survived to the present day, their dynamical influence earlier in the history of the solar system can explain the properties of the detached disk (defined in this paper as containing objects with
perihelia > 38 AU and semi-major axes between 80 and 500 AU), and the low mass of, and high eccentricities and inclinations in, the Kuiper Belt.

5.2 Introduction

The standard model for the formation and evolution of comets (Oort, 1950) assumes that comets form in the protoplanetary disk, in the region of the giant planets, and are then excited onto highly eccentric planet-crossing orbits through planetary perturbations. The comets then undergo a random walk in energy space, taking one “step” at each perihelion passage when they interact with the giant planets. This process continues until the aphelion distance is a few times $10^4$ AU, at which point torques from passing stars and the tidal field of the Galaxy are sufficient to increase the perihelion beyond $\sim 38$ AU in less than the energy diffusion time. Once this occurs, the comet no longer interacts gravitationally with the planetary system and its semi-major axis is frozen in place, apart from a much slower random walk due to perturbations from passing stars. This model correctly predicts many properties of the Oort comet cloud (Wiegert & Tremaine, 1999; Fernández, 2005; Fouchard et al., 2013), but also incorrectly predicts that there should not be objects with perihelia substantially beyond the orbit of Neptune and aphelia less than a few thousand AU, since there is no dynamical pathway to this part of phase space (Duncan et al., 1987).

More specifically, the JPL small body database contains 29 objects on orbits with semi-major axes between 80 and 500 AU and perihelia well beyond the orbit of Neptune ($q > 38$ AU). Such bodies are often said to belong to the detached disk (Gladman et al., 2002). The best known member of this group is Sedna, a body with perihelion 76 AU, semi-major axis 480 AU, and a radius of about 1,000 km (Brown et al., 2004; Pál et al., 2012).

One model to explain the orbits of these bodies (Brasser et al., 2006, 2012) holds that the Sun was born in a cluster of stars with density in the range $10^4–10^5 M_\odot pc^{-3}$. Tides and close stellar encounters from this cluster exert torques on bodies with aphelia of several hundred

\footnote{https://ssd.jpl.nasa.gov/sbdb_query.cgi}
AU, raising their perihelia and thereby creating the detached disk. Several other models for producing the detached disk are reviewed by Morbidelli & Levison (2004).

Another idea, first posited in Gladman et al. (2002) but not explored in detail, is that multiple Mars-sized embryos formed interior to the orbit of Neptune, and, like the comets, gained semi-major axes of a few hundred AU via interactions with the giant planets. These embryos exerted torques on one another as well as on smaller bodies and many of them increased their perihelia as a result, decoupling them from the perturbations of the giant planets and forming a long-lived detached disk. In this paper, we investigate this scenario through direct N-body simulations and make predictions for the properties of the detached disk and the masses, numbers, and orbits of terrestrial-mass planets at semi-major axes of a few hundred AU.

5.3 Initial Conditions

We report on ten sets of N-body simulations containing the Sun and the four giant planets, as well as several embryos (0.05–2 $M_{\oplus}$) initialized on nearly circular and co-planar orbits within the orbits of the giant planets. The inner planets are ignored. The giant planets are initialized at their current semi-major axes, but with smaller inclinations and eccentricities (see below for details). The parameters of these simulations are summarized in the first five columns of Table 5.1. In each case, we represented the embryos by $N_{eb}$ massive extra bodies (MEBs) of mass $M_{eb}$ placed on orbits between Jupiter and Neptune. The semi-major axes of the MEBs satisfy the relation

$$a_{i+1} = a_i + \Delta a_i^{0.5}, \ i = 0, ... \quad (5.1)$$

where $a_i$ is the semi-major axis of the $i^{th}$ extra body. If $a_{i+1}$ lies within two Hill radii of any of the four giant planets, we assume such an orbit will not be stable and we do not put a planet there. We choose $a_0 = 5.91$ AU (corresponding to a position two Hill radii outside
Jupiter’s orbit) and choose the largest value of $\Delta$ such that the most distant body would still be within the orbit of Neptune (30 AU) given that there are $N_{\text{eb}}$ total bodies. Apart from the gaps at the giant planets, this spacing corresponds to a surface density proportional to $r^{-1.5}$ as in Hayashi’s (1981) minimum-mass solar nebula. We ran 30 simulations in each set, differing only in the random number seeds.

Each simulation also contains a set of 50 test particles on orbits with semi-major axes 5 to 50 AU, with radial distribution corresponding to a $r^{-1.5}$ surface density profile.

The initial inclinations and eccentricities of all bodies, including the giant planets, are drawn from Rayleigh distributions with $\sigma_i = 0.001$ radians ($= 0.057^\circ$) and $\sigma_e = 0.002$, having probability density

$$\rho(i) = \frac{i}{\sigma_i^2} \exp\left(-\frac{i^2}{2\sigma_i^2}\right); \quad \rho(e) = \frac{e}{\sigma_e^2} \exp\left(-\frac{e^2}{2\sigma_e^2}\right). \quad (5.2)$$

We integrated the orbits for 4.5 Gyr using the IAS15 package, developed by Rein & Liu (2012). Within this package, we used the IAS15 integrator (Rein & Spiegel, 2015), which is designed to handle close encounters and highly eccentric orbits.

We did not include the effects of Galactic tides, tides from the Sun’s birth cluster, or passing stars. We neglect Galactic tides and passing stars because we expect these to not have a significant effect on the few hundred AU scales of interest. For example, the torque from the Galactic tide would need $\sim 10^{12}$ years to impart enough angular momentum to convert a radial orbit to a circular one at 200 AU. We do not include cluster tides, as their magnitude and the time over which they act is uncertain, and in any case the goal of this paper is to focus on the effects of mutual interactions between the MEBs, rather than the effects of cluster tides.
Table 5.1

Table 5.1: Simulation Statistics

<table>
<thead>
<tr>
<th>Sim set</th>
<th>N_sims</th>
<th>N_{eb}</th>
<th>M_{eb} (M_{\odot})</th>
<th>M_{eff} (M_{\odot})</th>
<th>f_s</th>
<th>&lt;N_{remain}&gt;</th>
<th>&lt;f&gt;</th>
<th>DD with MEB</th>
<th>DD no MEB</th>
<th>Kuiper Belt</th>
<th>&lt;t_{run}&gt; (Gyr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>10</td>
<td>2.0</td>
<td>6.3</td>
<td>0.67</td>
<td>4.5%</td>
<td>0.5</td>
<td>0.6%</td>
<td>0.5%</td>
<td>4.5%</td>
<td>0.5\cdot10^{-5}</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>20</td>
<td>1.0</td>
<td>4.5</td>
<td>0.87</td>
<td>2.5%</td>
<td>0.5</td>
<td>0.5%</td>
<td>1.9%</td>
<td>4.7\cdot10^{-4}</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>40</td>
<td>0.5</td>
<td>3.2</td>
<td>0.97</td>
<td>3.2%</td>
<td>1.3</td>
<td>1.8%</td>
<td>0.9%</td>
<td>8.6\cdot10^{-4}</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>10</td>
<td>1.0</td>
<td>3.2</td>
<td>0.80</td>
<td>3.3%</td>
<td>0.33</td>
<td>1.3%</td>
<td>2.2%</td>
<td>2.2\cdot10^{-3}</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>20</td>
<td>0.5</td>
<td>2.2</td>
<td>1.0</td>
<td>3.5%</td>
<td>0.7</td>
<td>2.0%</td>
<td>0.8%</td>
<td>1.4\cdot10^{-3}</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>40</td>
<td>0.25</td>
<td>1.6</td>
<td>1.0</td>
<td>2.8%</td>
<td>1.0</td>
<td>1.0%</td>
<td>1.6%</td>
<td>1.2\cdot10^{-3}</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>20</td>
<td>20</td>
<td>0.25</td>
<td>1.1</td>
<td>1.0</td>
<td>1.0%</td>
<td>0.2</td>
<td>0.34</td>
<td>2.0%</td>
<td>5.7\cdot10^{-3}</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>40</td>
<td>0.1</td>
<td>0.6</td>
<td>1.0</td>
<td>0.8%</td>
<td>0.3</td>
<td>0.0%</td>
<td>0.2%</td>
<td>2.1\cdot10^{-2}</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>40</td>
<td>0.05</td>
<td>0.3</td>
<td>1.0</td>
<td>1.0%</td>
<td>0.4</td>
<td>0.0%</td>
<td>0.0%</td>
<td>5.7\cdot10^{-2}</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>40</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>35.5%</td>
<td>14.0</td>
<td>0.0%</td>
<td>N/A</td>
<td>2.1\cdot10^{-1}</td>
<td></td>
</tr>
</tbody>
</table>

Column 2: Number of simulations in this set.
Column 3: Number of massive extra bodies (MEBs).
Column 4: Mass of each MEB.
Column 5: \(M_{\text{eff}} \equiv M_{\text{eb}} \sqrt{N_{\text{eb}}}\) is proportional to the magnitude of the torques experienced by the MEBs if they are randomly distributed.
Column 6: Mean fraction of systems in which all four giant planets remained for 4.5 Gyr.
Column 7: Mean fraction of MEBs that remain in the simulations included in column 6.
Column 8: Mean number of MEBs that remain in the simulations included in column 6 (column 3 times column 7).
Column 9: Fraction of the systems with at least one remaining MEB, in which all MEBs would be expected to escape dynamical detection, and be fainter than a \(V\)-band magnitude of 19 (see Section 5.6 for details).
Column 10: Fraction of test particles that end up in the detached disk (\(q > 38\) AU and \(80\) AU < \(a < 500\) AU) for simulations with at least one remaining MEB.
Column 11: Fraction of test particles that end up in the detached disk (\(q > 38\) AU and \(80\) AU < \(a < 500\) AU) for simulations with no remaining MEBs.
Column 12: Fraction of test particles that end up in the Kuiper Belt (\(30\) AU < \(R < 50\) AU).
Column 13: Mean run time of the simulations.
We removed bodies from the simulation if they went outside a box with side length $5 \cdot 10^3$ AU centered on the Sun. Outside this box, our neglect of the Galactic tides is not accurate. If two particles collide, they are assumed to merge completely. In determining whether a collision occurred, we used the current radii of the four giant planets, and removed any particle that came within 0.1 AU of the Sun. Integrations were run for 4.5 Gyr, or until all of the MEBs and test particles, or one of the giant planets, were removed from the simulation. Currently a few of our simulations have not run for the full 4.5 Gyr. This is noted in Column 13 of Table 5.1. We expect that this only affects the results appreciably for simulation set 10, and this has been noted throughout the paper where we believe it to be relevant.

Simulation sets 1–3 and 4–6 have a total mass of $20 \, M_\oplus$ and $10 \, M_\oplus$ respectively in MEBs, with varying masses for the individual MEBs between 0.25 and $2 \, M_\oplus$. Simulation sets 7–9 explore smaller total masses of MEBs. Set 10 is a control set where the mass of the MEBs is set to zero. We also ran more simulations to explore parameter space; these informed our conclusions but will not be reported on explicitly.

When reporting orbital elements, we use Jacobi coordinates. Inclinations are reported relative to the fixed reference plane near which the bodies were initialized.

5.4 Simulation Results

Figure 5.1 shows the interdecile range ($10^{th}$ to $90^{th}$ percentile) of $\sqrt{e_N^2 + i_N^2}$ (where $e_N$ and $i_N$ are the eccentricity and inclination of Neptune after 4.5 Gyr) as a function of the effective mass of the MEB population, given by

$$M_{\text{eff}} = M_{eb} \sqrt{N_{eb}}.$$  \hspace{1cm} (5.3)

If the MEBs are randomly distributed azimuthally, then $M_{\text{eff}}$ is proportional to the net torque that the MEBs exert on each other and smaller planetesimals. The figure shows that
Figure 5.1: Excitation of Neptune’s eccentricity and inclination ($\sqrt{e_N^2 + i_N^2}$) after 4.5 Gyr vs. $M_{\text{eff}} = M_{\text{eb}}\sqrt{N_{\text{eb}}}$. We have plotted the interdecile range for all surviving simulations (simulations in which all 4 giant planets are still present after 4.5 Gyr) in each set. The horizontal line corresponds to the averaged (over secular oscillations) present-day value of $\sqrt{e_N^2 + i_N^2}$, where $i_N$ is measured relative to the invariable plane of the solar system. The two lines near $M_{\text{eff}} = 3.1M_\oplus$ have been horizontally offset slightly so as to both be visible. $M_{\text{eff}}$, which is proportional to the rms torque from the MEBs if they are randomly distributed, is a good predictor of the typical value of $\sqrt{e_N^2 + i_N^2}$. 

203
Figure 5.2: Location of massive extra bodies (MEBs) in the space of semi-major axis and perihelion after 4.5 Gyr. The color of the symbols corresponds to the effective mass of the population of MEBs. The black squares correspond to the case in which the MEBs have zero mass. The legend refers to the simulation numbers in column 1 of Table 5.1. We only show bodies in surviving systems (systems in which all four giant planets remain). We have labeled the lines corresponding to circular orbits, the boundary of the detached disk, and to orbits with aphelion equal to 140 AU (a rough bound to detectability, see Section 5.6). When $M_{eb} = 0$ (Simulation 10), there are no bodies with perihelion greater than 38 AU, as expected since the only torques come from the giant planets. While many bodies from simulation set 10 survive, with one exception, their orbits have semi-major axes less than 30 AU, and are not shown on this plot.
Figure 5.3: Location of MEBs in the space of semi-major axis and inclination after 4.5 Gyr. We have used the same color coding and point styles as in Figure 5.2. Inclinations remain moderate and do not vary greatly with semi-major axis or $M_{\text{eff}}$. 

205
Figure 5.4: Smoothed estimates of the density of the MEB population in semi-major axis, eccentricity, inclination and perihelion. The line color corresponds to $M_{\text{eff}}$, and the legend refers to the simulation numbers in column 1 of Table 5.1. The drop of all the curves $\rho(e)$, $\rho(i)$ to zero at $e = 0$ and $i = 0$ is due to our smoothing kernel and is expected for any smooth distribution in phase space (see text). The semi-major axis distributions peak between 50 and 150 AU. Higher values of $M_{\text{eff}}$ lead to higher mean values of the perihelion $q$. 

206
Figure 5.5: The top panel shows the fraction of remaining MEBs as a function of time. The bottom panel shows the mean fraction of MEBs in the detached disk (perihelion greater than 38 AU and aphelion in the range $80 \text{ AU} \leq a \leq 500 \text{ AU}$) as a function of time. The detached disk forms over a few times $10^8$ years, and decays slowly afterwards. Note that the precipitous decline of the black line around 2 Gyr is due to unfinished simulations.
$M_{\text{eff}}$ is a good predictor of the eccentricity and inclination excitation of Neptune. Similar results apply to Uranus.

All but the smallest values of $M_{\text{eff}}$ considered in our simulations excite the eccentricity and inclination of Neptune to values that exceed those in the real solar system. It seems we require $M_{\text{eff}} \lesssim M_{\oplus}$ in order for the eccentricity and inclination of Neptune to be roughly compatible with the current observations. The constraints from Uranus are much weaker due to its higher current eccentricity of 0.046. Only simulations 8, 9, and 10 have $M_{\text{eff}} \lesssim M_{\oplus}$. Subsequent damping of the eccentricities and inclinations (see Section 5.7) could relax this limit.

Column 6 of Table 5.1 shows the fraction of systems that survive in each simulation set. We define a simulation as “surviving” if all four giant planets remain on bound orbits after 4.5 Gyr. A significant fraction of simulations in the sets with the highest values of $M_{\text{eff}}$ did not survive. Since Figure 5.1 indicates that the degree of eccentricity and inclination excitation is roughly proportional to $M_{\text{eff}}$, we conclude that if $M_{\text{eff}} \gtrsim 5M_{\oplus}$ there is a significant chance that one of the giant planets will be ejected or collide with the Sun or another planet (unless the eccentricities and inclinations are subsequently damped, see Section 5.7). In the figures and discussion that follow, we only report results from surviving simulations.

Figure 5.2 is a scatter plot of perihelion vs. semi-major axis for MEBs remaining in the solar system after 4.5 Gyr. Points are color-coded according to the value of $M_{\text{eff}}$ for the corresponding simulation. The black squares correspond to our control simulation with massless extra bodies (simulation 10). We have labeled three special lines. The line at $q = a$ corresponds to circular orbits. The horizontal line labeled “boundary of detached disk” at $q = 38$ AU is roughly the distance outside which non-resonant perturbations from Neptune are unimportant over the age of the solar system [Gladman et al., 2002]. It is expected that after several Gyr, there will be few non-resonant bodies inside this line, as most would have been ejected by Neptune. Finally, we have drawn a line where the aphelion $Q$ is equal to 140
AU. This is a rough lower limit to the distance at which one of the MEBs could plausibly have escaped detection (see Section 5.6).

We see that most of the remaining MEBs are on orbits with perihelia less than 80 AU and semi-major axes less than a few hundred AU. In contrast, the test particles in the model of Brasser et al. (2012), where the torque was applied by tides and stellar flybys in the birth cluster, typically have semi-major axes greater than 1,000 AU.

Figure 5.3 is a scatter plot of orbital inclination against semi-major axis. There are no remaining MEBs with inclination greater than 50°. The median inclinations for our different simulation sets range from 4° to 15°, although much of this scatter is due to small number statistics. This is very different from the inclination distribution of detached objects that are formed in models involving cluster tides and flybys (Brasser et al., 2012)—even in the innermost part of their cloud, the median inclination is between 45° and 55°. Thus, the inclination distribution in the detached disk provides a clean test to distinguish our model from cluster tides.

Figure 5.4 shows the density of the MEBs in semi-major axis, eccentricity, inclination and perihelion. The curves for semi-major axis and perihelion were smoothed with a log-normal kernel, with a dispersion of 0.2 in \( \ln a \) or \( \ln q \). To determine the curves for eccentricity, we smoothed the eccentricity vector with a 2D Gaussian with \( \sigma_e = 0.1 \), and then integrated the resulting probability distribution over angle to recover the distribution of scalar eccentricity. This procedure ensures that \( \rho(e) \sim e \) for small values of \( e \)—consistent with a non-singular distribution in phase space around \( e = 0 \). We applied the same method to the inclination using \( \sigma_i = 5° \). Median eccentricities range from 0.28 to 0.40, although as can be seen from the figure, the distributions are quite broad. Values of eccentricity near unity are uncommon because particles are mostly limited to \( 0 < e < 1 - q_{\text{crit}}/a \), where \( q_{\text{crit}} \approx 38 \) AU is the perihelion below which a planet will be ejected by perturbations from Neptune. Systems with higher values of \( M_{\text{eff}} \) have higher mean values of the perihelion. This can be
understood qualitatively as being due to the larger typical torques experienced by particles in these simulations.

Figure 5.5 shows the evolution of the number of MEBs over time. The top panel shows the total fraction of remaining MEBs. Most of them are removed over $\sim 10^7$ year timescales. About a third of the MEBs survive for the duration of the simulation if their masses are negligible (black line, simulation set 10). The drop in the black line at $\sim 3$ Gyr is due to simulations that have not finished running. This is because the stability of test particles on near-circular orbits in the region of the giant planets (5-30 AU) depends strongly on the eccentricities of the planets. In our simulations the planets start on nearly circular orbits (Equation 2) but MEBs with non-zero masses rapidly excite the eccentricities, while those with zero masses do not. The bottom panel shows the fraction of MEBs on orbits with $q > 38$ AU. This number grows for a few times $10^8$ years, and slowly declines thereafter. As expected, there are no bodies with $q > 38$ AU if the MEBs have zero mass (i.e., there is no black line in the bottom panel of Figure 5.5). The approximate timescale for the creation of bodies with perihelion greater than 38 AU can be derived simply. Assuming a typical specific torque of $GM_{\text{eff}}/\langle R \rangle$, where $\langle R \rangle$ is the average orbital separation, then the timescale to produce orbits with perihelion greater than $q_{\text{crit}}$ can be approximated as

$$
\tau = \frac{\sqrt{2}GM_{\odot}q_{\text{crit}}}{GM_{\text{eff}}/\langle R \rangle} = 7 \cdot 10^7 \left( \frac{M_{\text{eff}}}{M_\oplus} \right)^{-1} \frac{\langle R \rangle}{150 \text{ AU}} \text{ years,} \tag{5.4}
$$

where we have used $q_{\text{crit}} = 38$ AU in the numerical estimate. Equation (5.4) agrees fairly well with the results shown in Figure 5.5. In particular, higher values of $M_{\text{eff}}$ lead to more rapid emplacement of bodies in high perihelion orbits. On the other hand, after a group of MEBs is placed on high perihelion orbits, higher values of $M_{\text{eff}}$ also lead to more rapid depletion over the remaining age of the solar system. Because the MEB population evolves on such long timescales, we expect the results of this model to not depend strongly on the details of the initial conditions and early evolution of the solar system.
Figure 5.6 shows a histogram of the number of remaining MEBs in the different runs of each simulation set. For a fixed number of MEBs, more remain after 4.5 Gyr in systems with intermediate-mass MEBs. There are few systems with more than one remaining MEB in which \( M_{eb} \) is greater than 0.5 \( M_{⊕} \), presumably because the mutual torques between MEBs of larger masses are sufficiently strong to lower the perihelia to the point that MEBs are ejected by one of the giant planets, until only one MEB is left.

To summarize, our simulations show that a population of MEBs with effective mass \( M_{eff} = \sqrt{N_{eb} M_{eb}} \) exceeding \( \sim 1 M_{⊕} \) will excite the eccentricities and inclinations of Uranus and Neptune to values larger than observed—thus either there has been subsequent damping or \( M_{eff} \lesssim M_{⊕} \). In the latter case, for \( M_{eb} \gtrsim 0.05 M_{⊕} \) we find a probability of up to \( \sim 40\% \) that one or occasionally more MEBs survive in bound orbits until the present time. Their orbits typically have perihelia of 40–70 AU, semi-major axes less than 200 AU, and inclinations \( \lesssim 30° \).

5.5 Fate of the Test Particle Population

As described in Section 5.3, in each set of simulations we also included 50 test particles that were originally in dynamically cold orbits (initial eccentricity and inclination less than 1%) between 5 and 50 AU, with surface density \( \propto r^{-1.5} \). These are included to monitor the effect of the MEBs on the Kuiper Belt, and to study the efficiency of injection of particles into the detached disk (defined in this paper as containing any body with perihelion greater than 38 AU and semi-major axis between 80 and 500 AU). Gladman & Chan (2006) showed that massive bodies exterior to Neptune lift test particles to high perihelion distances, thus providing a natural mechanism to create a disk of detached objects such as Sedna (Brown et al., 2004) or 2012VP\textsubscript{113} (Trujillo & Sheppard, 2014). We make predictions for the mass and orbital element distribution of a scattered disk produced in this way. Many of our results are similar to those of Gladman & Chan (2006).
Figure 5.6: Histograms of the number of MEBs remaining after 4.5 Gyr among the surviving simulations in each set. The mass and initial number of the extra bodies are labeled on each panel. Simulations with intermediate-mass MEBs (between 0.25 and 0.5 \( M_\oplus \)) tend to retain a higher number than simulations with either larger or smaller masses. It is rare for more than two MEBs to remain at 4.5 Gyr. We have currently not plotted the number of remaining MEBs for the control simulations (Simulation 10), as they have not run for the full 4.5 Gyr.
Figure 5.7 shows scatter plots of perihelion and inclination vs. semi-major axis for all test particles still present at the end of our simulations. We divide the results into top and bottom panels based on whether any of the MEBs remain after 4.5 Gyr. As expected, almost all of the test particles with perihelion less than 38 AU have been ejected. The orbits of the remaining test particles are similar to those of the remaining MEBs, and seem not to depend strongly on whether any of the MEBs remain, although many more test particles remain from simulations sets 1 and 2 (the sets with the highest values of $M_{\text{eff}}$) in the cases where all MEBs are ejected. The inclination distribution of the test particles is similar to that of the MEBs, but slightly broader. This inclination distribution is consistent with observations of the detached disk. Our results also imply that few bodies will be found in the detached disk with semi-major axes $\gtrsim 500$ AU.

Column 10 in Table 5.1 shows the fraction of test particles that end up in the detached disk for simulations in which at least one MEB remains after 4.5 Gyr. Column 11 shows this same fraction, but for simulations in which no MEBs remain. Many of these fractions are within a factor of a few of the fraction of test particles that are expected to be transferred to the present-day Oort cloud, estimated by Brasser et al. (2010) to be around 4.5%. Thus in our model the masses of the detached disk and the Oort cloud should be similar.

We now directly estimate the mass in the detached disk predicted by our model. Let us assume that the solids in the solar nebula were distributed with surface density equal to $30 \text{ g cm}^{-2} \cdot (r/\text{AU})^{-1.5}$ (Hayashi, 1981). This estimate assumes that all of the solids in the disk went into the planets. The test particles in our simulations are distributed with surface density distribution proportional to $r^{-1.5}$ from 5 to 50 AU. If we assume, for lack of a better estimate, that the same amount of solid mass was left over in planetesimals as was incorporated into the planets (implying that the solar nebula was actually twice as massive as Hayashi’s minimum estimate), then there were 70 Earth masses of planetesimals left over after planet formation in the region covered by our test particle population.
Assuming 2% of the test particles are transferred to the detached disk after 4.5 Gyr (see columns 10 and 11 of Table 5.1), then the detached disk should contain 1.4 Earth masses given the assumptions in the previous paragraph. For comparison, Brown et al. (2004) make a rough estimate that there are five Earth masses in bodies with Sedna-like orbits. This estimate was based on only one body (Sedna) and the assumption that the detached disk has the same mass distribution as the Kuiper Belt, so it is subject to much uncertainty.

Gladman et al. (2002) estimate that there are $10^6$ bodies greater than 100 km in the detached disk. This corresponds to a minimum mass of $0.175 M_\oplus$, assuming a density of 2 g cm$^{-3}$. For a plausible size distribution the actual mass could be larger by a factor of five or more.

Petit et al. (1999) suggest that large planetesimals scattered by the giant planets could be the reason that the Kuiper Belt is much less massive and more excited that it is thought to have been primordially. This suggestion is qualitatively consistent with our finding that the MEBs remove most of the test particles in our simulations. We now make a more quantitative comparison. Gladman et al. (2001) estimate a current mass of $0.1 M_\oplus$ for the Kuiper Belt. Their definition of the Kuiper Belt includes all bodies (except for Neptune) currently within 30-50 AU from the Sun. If their estimate is correct and we assume, as we did above, that the planetesimal population between 5 and 50 AU had a mass of $70 M_\oplus$, then about 0.14% of the original planetesimal population ended up in today’s Kuiper Belt.

To compare this result to our simulations, we considered two definitions of the Kuiper Belt. The first, strictly analogous to the Gladman paper, took any test particle with heliocentric distance between 30 and 50 AU to be in the Kuiper Belt. Thus, for every particle remaining after 4.5 Gyr, we calculated the fraction of its orbit that it spends between 30 and 50 AU, and weighted it accordingly. The fraction of test particles that end up in the Kuiper Belt is shown in Column 12 of Table 5.1. By this metric, it appears that simulations 2–7 do a reasonable job of reproducing the 0.14% target discussed in the previous paragraph. Simulations with too small a value of $M_{\text{eff}}$ leave behind a more massive Kuiper Belt than
observed, by more than a factor of 10 in some cases, so either some other process has eroded the mass, or our crude estimate of the initial mass is too high.

We also consider a more restricted definition of the Kuiper Belt that only includes particles with semi-major axes between 30 and 60 AU, and weights them by the fraction of the time they spend between 30 and 50 AU. This reduces contamination of the “Kuiper Belt” by objects in the scattered or detached disks. This definition causes modest reductions in many of the Kuiper belt populations, but does not change the qualitative picture. In this definition, fractions (0, $1.9 \cdot 10^{-4}, 2.2 \cdot 10^{-4}, 9.7 \cdot 10^{-4}, 1.0 \cdot 10^{-3}, 4.9 \cdot 10^{-4}, 5.2 \cdot 10^{-3}, 2.0 \cdot 10^{-2}, 5.6 \cdot 10^{-2}, 2.1 \cdot 10^{-1}$) survive (cf. column 12 of Table 5.1) for simulations 1–10. That said, in most cases we only have 1,500 test particles per simulation set, so a result below 0.14% could just be the result of small number statistics.

To summarize, our simulations show that MEBs can excite planetesimals into orbits with typical perihelia of 40–80 AU, semi-major axes as large as a few hundred AU, and inclinations 0–40°, consistent with the observed properties and (within large uncertainties) total mass of the observed detached disk. The MEBs may also be responsible for destroying most of the Kuiper Belt, which is much less massive than expected from extrapolating the solid content of the minimum solar nebula. These conclusions hold whether or not one or more MEBs is eventually found in the outer solar system, since the properties of the detached disk and Kuiper Belt are mostly independent of whether or not any MEBs survive.

5.6 Observational Constraints

MEBs can be detected dynamically or photometrically. We first consider dynamical detection. [Fienga et al. (2016)] find that a hypothetical planet of 10 Earth masses would be dynamically detectable, mostly from its perturbations to Saturn, if it were currently closer to the Sun than about 370 AU. This is a model-dependent result, as they were considering a planet on a particular orbit proposed by [Batygin & Brown (2016)]}, but this limit is probably
Figure 5.7: Orbital properties of the test particle population. The black squares represent known members of the detached disk ($q > 38$ AU, $80$ AU $< a < 500$ AU). Otherwise, the point properties are the same as for Figure 5.2 except that simulation 10 is not plotted. The top panels are for surviving runs in which at least one of the MEBs remains after 4.5 Gyr. The bottom panels are for surviving runs in which all of the MEBs were ejected. We chose not to plot observed objects with $a < 80$ AU as they are quite numerous and do not fall under our definition of detached disk objects. There is an observational selection bias that favors the discovery of objects with low semi-major axes and perihelia.
the best available for our purposes. To good approximation a planet at that distance only interacts with the known planets via a stationary tidal potential over the interval of modern observations. Thus we may assume that the detectability limit scales as $M/R^3$, where $M$ is the hypothetical planet’s mass and $R$ its current distance. Therefore an MEB should be dynamically detectable if its current heliocentric distance is less than about

$$R_{\text{crit}} = 170 \left( \frac{M}{M_{\oplus}} \right)^{1/3} \text{AU}. \quad (5.5)$$

We next consider photometric detection. The most extensive systematic survey is the Palomar Distant Solar System Survey (Schwamb et al., 2010), which covered 30% of the sky to a limiting $r$-band magnitude of 21.3. Using serendipitous discoveries from the Catalina Sky Survey and the Siding Spring Survey, Brown et al. (2015) estimate that there is a 30% chance that the solar system contains an undetected KBO brighter than a $V$-band magnitude of 19. In an abstract, Holman et al. (2016) report that Pan-STARRS has completed a search for slow-moving objects brighter than an $r$-band magnitude of 22.5 over the entire sky north of $-30^\circ$ in declination (75% of the sky). This survey area includes most, but not quite all of our bodies. Assuming a geometric albedo of 0.04 and a constant density of 2.0 g cm$^{-3}$ (appropriate for long-period comets; Lamy et al. 2004), an $r$-band limit of 19.0 (22.5) corresponds to a critical detection distance of $140 \,(313) (M/M_{\oplus})^{1/6}$ AU. This increases to $235 \,(526) (M/M_{\oplus})^{1/6}$ AU if we take the albedo to be 0.32, appropriate for Sedna (Pál et al., 2012). Thus, the Pan-STARRS survey has a high probability of detecting any remaining MEBs.

Column 9 in Table 5.1 shows the probability that none of the MEBs remaining in a given simulation after 4.5 Gyr are either dynamically detectable or brighter than magnitude 19 assuming an albedo of 0.04. The probability is only calculated for simulations in which at least one MEB remains after 4.5 Gyr. To compute this probability, we use the smaller of $170 \,(M/M_{\oplus})^{1/3}$ AU (Equation 5.5) and $140(M/M_{\oplus})^{1/6}$ AU to calculate $R_{\text{crit}}$ for each
remaining MEB. Then, given the orbital parameters of the MEB at the end of the simulation, we calculated the fraction of the orbital period that would be spent beyond $R_{\text{crit}}$. This is the probability that a given MEB would be undetected. Then, to find the probability that all the planets are undetectable, we multiply all the individual probabilities. This is not completely correct, even assuming the bodies’ phases to be uncorrelated, as their dynamical effect on the known planets could add constructively or destructively depending on their current locations. These probabilities are generally a few tens of percent, so if MEBs formed this way were still present in the solar system, it is likely but far from certain that they would have been detected.

Volk & Malhotra (2017) find that there is a statistically significant (at the 97% level) warp in the mean plane of the Kuiper Belt by comparing the orbits of non-resonant bodies with semi-major axes in the range from 42 to 48 AU to those with semi-major axes from 50 to 80 AU. They comment that this warp could be caused by an unseen Mars-mass body orbiting at 65–80 AU, consistent with the properties of the remaining MEBs in our simulations.

5.7 Damping

We found (Section 5.4) that the eccentricity and inclination of Neptune were excited above their observed values when $M_{\text{eff}} \gtrsim M_\oplus$. This constraint on $M_{\text{eff}}$ could be relaxed if the eccentricity and inclination were subsequently damped. Multiple groups (e.g., Kokubo & Ida 1995; Tsiganis et al. 2005) have found that the presence of many small bodies in a disk can damp the eccentricity of larger planetary-mass objects through dynamical friction. Directly modeling this process with an $N$-body planetesimal disk was impractical (even ignoring interactions between planetesimals), so we experimented with implementing a drag force

$$a_{r,i} = -\frac{v_{r,i}}{\tau_{\text{damp}}} - \frac{v_{z,i}}{\tau_{\text{damp}}}, \quad (5.6)$$
where $a_{r,i}$, $v_{r,i}$ and $a_{v,i}$, $v_{z,i}$ are the radial and vertical accelerations and velocities of the $i^{th}$ giant planet. We assumed the damping time $\tau_{\text{damp}}$ to be given by

$$\tau_{\text{damp}}(t) = \tau_1 \exp \left( t/\tau_2 \right).$$

(5.7)

We take $\tau_1 = 10^5$ years and $\tau_2 = 10^7$ years. $\tau_2$ is consistent with the time over which small bodies are removed from the system (see Figure 5.5). The damping force does not act upon the MEBs, as they are presumed to be insufficiently massive to be affected by dynamical friction with a planetesimal disk.

This formalism does produce some damping of the semi-major axes as well, but provided that the eccentricities and inclinations are much less than unity, one can show that under the influence of damping,

$$\frac{d \ln(a)}{dt} = \frac{d}{dt} \left( e^2 + i^2 \right).$$

(5.8)

Therefore, if the eccentricities and inclinations remain low, they are damped much faster than the semi-major axis.

We ran simulations analogous to our simulation sets 1–3 using the damping scheme discussed above and found little qualitative difference from the results presented in this paper. Even with $\tau_1 = 10^5$ years, large MEBs ($M_{eb} \geq 2M_\oplus$) could still eject one or more of the giant planets. As can be seen in the second panel of Figure 5.5, the MEB population still interacts with the giant planets after more than $10^8$ years, at which point one would expect most of the planetesimal disk to be gone, and the damping to therefore be negligible.

### 5.8 Conclusions

It is highly likely that multiple “planetary embryos” or massive extra bodies (MEBs) of up to a few Earth masses form among the giant planets. These are scattered outwards due to
gravitational perturbations from the giant planets. The MEBs exert torques on one another before they can be ejected from the solar system, thus increasing their perihelia beyond the gravitational reach of the giant planets. They can therefore remain in the solar system for 4.5 Gyr, outside the orbit of Neptune but far inside the Oort cloud. Additionally, these bodies exert torques on smaller planetesimals, thereby naturally creating a detached disk containing objects with orbits similar to Sedna and 2012VP_{113}.

The evolution of the population of MEBs is characterized by their effective mass $M_{\text{eff}} = M_{\text{eb}} \sqrt{N_{\text{eb}}}$. If $M_{\text{eff}}$ exceeds about an Earth mass, then the eccentricity and inclination of Neptune are usually excited above the observed values, which is inconsistent unless some process subsequently damps them. However, significant effects occur for values of the effective mass around this limit. For example, if 20 bodies of 0.25 Earth masses were present (Simulation 7; $M_{\text{eff}} = 1.1M_\oplus$), then in 20% of cases one MEB will remain in a moderate inclination orbit with semi-major axis between 50 and 150 AU, and on average 1.2% of the mass in the planetesimal belt between 5 and 50 AU is transferred to the detached disk. In many of our simulations with remaining extra bodies, at least one of them would be expected to have been detected either dynamically or photometrically, although this result depends strongly on studies done for other purposes or surveys that are still in progress. In a few tens of percent of simulations with remaining MEBs, the bodies were remote enough to escape dynamical detection and be fainter than 19th magnitude.

MEBs also transfer a few percent of the mass of the initial planetesimal belt into a detached disk composed of bodies on moderately inclined orbits with perihelia greater than 38 AU and semi-major axes between 80 and 500 AU. Like the observed sample of 29 objects in our definition of the detached disk, the population of test particles in our simulations are on moderately inclined orbits after 4.5 Gyr. The amount of material in the detached disk does not depend strongly on whether any of the MEBs remain in the system after 4.5 Gyr; thus the viability of this mechanism for producing the detached disk does not require the discovery of new terrestrial-mass planets in the outer solar system.

220
The original motivation for this model was to explain the properties of the detached disk, which is not predicted by the standard model of the formation and evolution of the Oort comet cloud. There are other tensions between observations and the standard model. These include a predicted mass for the Oort cloud that is too low given the likely mass in the planetesimal disk, a predicted mass for the scattered disk of comets that is too high, and a predicted inclination distribution for long-period comets with too many retrograde orbits (e.g., Wiegert & Tremaine (1999); Morbidelli (2008)). It is possible that the presence of massive extra bodies in the outer planetesimal disk will alleviate or resolve some of these tensions, but studying this possibility is beyond the ambition of the current paper.

We acknowledge helpful discussions with Matthew Holman, Brian Lacki, Matthew Payne, Roman Rafikov, and Yanqin Wu.
Bibliography


Brasser, R., Duncan, M. J., & Levison, H. F. 2006, Icarus, 184, 59


Duncan, M., Quinn, T., & Tremaine, S. 1987, AJ, 94, 1330


Hayashi, C. 1981, Progress of Theoretical Physics Supplement, 70, 35


Wiegert, P., & Tremaine, S. 1999, Icarus, 137, 84