Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **401,** 867–889 (2010) doi:10.1111/j.1365-2966.2009.15739.x

Long-lived planetesimal discs

Kevin Heng★ and Scott Tremaine★

Institute for Advanced Study, School of Natural Sciences, Einstein Drive, Princeton, NJ 08540, USA

Accepted 2009 September 14. Received 2009 September 7; in original form 2009 June 10

ABSTRACT

We investigate the survival of planetesimal discs over Gyr time-scales, using a unified approach that is applicable to all Keplerian discs of solid bodies – dust grains, asteroids, planets, etc. Planetesimal discs can be characterized locally by four parameters: surface density, semimajor axis, planetesimal size and planetesimal radial velocity dispersion. Any planetesimal disc must have survived all dynamical processes, including gravitational instability, dynamical chaos, gravitational scattering, physical collisions, and radiation forces, that would lead to significant evolution over its lifetime. These processes lead to a rich set of constraints that strongly restrict the possible properties of long-lived discs. Within this framework, we also discuss the detection of planetesimal discs using radial velocity measurements, transits, microlensing and the infrared emission from the planetesimals themselves or from dust generated by planetesimal collisions.

Key words: gravitational lensing – Kuiper Belt – minor planets, asteroids – planets and satellites: formation – Solar system: formation – stars: formation.

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

1 INTRODUCTION

Terrestrial planets and the cores of giant planets are generally believed to have formed hierarchically: small solid bodies ('planetesimals') condense from the gaseous circumstellar disc (Johansen et al. 2007), collide repeatedly, and accumulate into larger and larger assemblies (Safronov 1972; Goldreich, Lithwick & Sari 2004a; Reipurth, Jewitt & Keil 2007). Several aspects of this complex process remain obscure, in particular (but not limited to) the formation of planetesimals from dust grains (Blum & Wurm 2008), how to grow Uranus and Neptune in the short time available before the gaseous disc is dissipated (Goldreich et al. 2004a; Goldreich, Lithwick & Sari 2004b), the origins of the large eccentricities of the extrasolar planets (Tremaine & Zakamska 2004), the role of planetary migration (Goldreich & Tremaine 1980; Papaloizou & Terquem 2006), how the residual planetesimals were cleaned out of the Solar system (Goldreich et al. 2004b), and why the Solar system is so different from known extrasolar planetary systems (Beer et al. 2004).

Given these large gaps in our understanding, it is worthwhile to investigate not just the difficult question of how planets form but also the simpler question of whether they can survive once formed. Planetary systems in the Galaxy have presumably been formed at a more-or-less constant rate, so it is reasonable to assume that most of today's planetary systems are at least several Gyr old. They must therefore have survived all dynamical processes – gravitational instabilities, collisions, viscous stirring or two-body relaxation, etc. – that would lead to a substantial change in their properties on time-scales less than about 3 Gyr. Understanding what long-lived planetary systems are possible should help us to understand to what extent the properties of actual planets are shaped by the formation process as opposed to evolution ('nature versus nurture') and may guide observers in searching for novel types of planetary systems.

Orbiting solid bodies are often called 'planetesimals', 'planetary embryos' or 'planets' depending on their mass, but for simplicity we will use the term 'planetesimal' to describe any body, whether solid like a terrestrial planet or gas-dominated like a giant planet, that is large enough so that gas drag and radiation effects (Poynting–Robertson drag, radiation pressure, Yarkovsky effect, etc.) are negligible.

We will find it useful to classify discs as 'hot' or 'cold' depending on whether or not the planetesimal orbits cross. Within the Solar system, the planets form a cold system (except for Pluto) while the asteroid and Kuiper belts are hot. Among hot planetesimal discs, an important special case is the 'collision-limited' disc, in which the collision time between planetesimals is equal to the age of the disc. Collision-limited discs are likely to arise from discs in which there is a distribution of planetesimal sizes: smaller planetesimals have shorter collision times and therefore are destroyed first, so the dominant planetesimal population (by mass) always has a collision time that is roughly equal to the disc age. We will also use the term 'warm' to describe discs in which the planetesimal orbits cross, but the impact velocities are so low that the cumulative effect of collisions does not substantially damage the planetesimals (Section 3.2.2).

★E-mail: heng@ias.edu (KH); tremaine@ias.edu (ST)

© 2009 The Authors. Journal compilation © 2009 RAS

Table 1. Sample planetesimal discs.

	A	B	C	D	E	F
a	1 au	1 au	10 au	10 au	100 au	100 au
$\mu \equiv \Sigma a^2/\mathrm{M}_\odot$	10^{-4}	10^{-6}	10^{-4}	10^{-6}	10^{-4}	10^{-6}
$M_{\rm disc} = \pi f_m \Sigma a^2$	$\simeq 100 M_\oplus$	$\simeq 1 M_\oplus$	$\simeq 100 M_\oplus$	$\simeq 1 M_\oplus$	$\simeq 100 M_\oplus$	$\simeq 1 M_\oplus$
$m_{\rm min}^{\rm cold}$	$60 M_\oplus$	$0.1 M_\oplus$	$60 M_\oplus$	$0.1 M_\oplus$	$60 M_\oplus$	$0.1 M_\oplus$
$N_{\rm max}^{\rm cold}$	1–2	12	1–2	12	1–2	12
$m_{\rm min}^{\rm hot}$	–	–	–	10^{24} g	4×10^{19} g	4×10^{13} g
$N_{\rm max}^{\rm hot}$	–	–	–	5×10^3	2×10^{10}	2×10^{14}
$m_{\rm max}^{\rm warm}$	–	–	–	3×10^{25} g	10^{21} g	10^{15} g
$N_{\rm min}^{\rm warm}$	–	–	–	200	5×10^8	5×10^{12}

Note: – $m_{\rm min}^{\rm cold}$ (equation 65), $m_{\rm min}^{\rm hot}$ (equation 74) are the minimum planetesimal mass for cold and hot discs, respectively; $m_{\rm max}^{\rm warm}$ is the maximum mass for warm discs (equation 75). $N_{\rm max}^{\rm cold}$ (equation 67) and $N_{\rm max}^{\rm hot}$ (equation 74) are the maximum number of planetesimals per octave for cold and hot discs, respectively; $N_{\rm min}^{\rm warm}$ (equation 75) is the minimum number for warm discs.

The most important observational signature of many planetesimal discs arises from dust formed in recent planetesimal collisions. These 'debris discs' were first detected from the thermal emission of the dust, which creates an infrared (IR) excess in the spectral energy distribution of the otherwise normal stars that they surround (Aumann et al. 1984). Spatially resolved debris discs are sometimes also visible from their scattered light. The IR excess (bolometric) luminosity relative to the luminosity from the parent star depends on the distance and spectral type of the host star as well as the disc radius, but is typically $\gtrsim 10^{-5}$ (see Zuckerman 2001 and Wyatt 2008 for reviews). The asteroid and Kuiper belts in our own Solar system can be thought of as debris discs, although they would not be detectable around other stars with current technology since the bolometric IR excess is $\sim 10^{-7}$ for both belts.

Debris discs have been detected around stars with a wide range of spectral types (A to M) and ages ($\sim 10^7$– 10^{10} yr).[1] The dust masses inferred from these observations are $10^{-3} \lesssim M_{\rm dust}/M_\oplus \lesssim 1$ (fig. 3 of Wyatt 2008), although this result depends on the assumed size distribution of the dust. The term 'debris' emphasizes that the lifetime of the dust grains from grain–grain collisions, Poynting–Robertson drag or radiation pressure is robustly and considerably less than the stellar age, so the dust cannot be primordial and must be continuously regenerated, presumably by ongoing planetesimal collisions.

We stress that the discs we consider in this paper are far more general than debris discs: they include hot discs in which the rate of dust generation may be undetectably small, cold discs in which there are no collisions, planetary systems, asteroid belts, planetary rings, etc.

We begin by constructing a simple model for a planetesimal disc in Section 2. In Section 3, we describe the dynamical processes that act on planetesimal discs. Collision-limited discs, which are a special subset of hot discs, are described in Section 4. Non-gravitational forces on dust are briefly reviewed in Section 5. The properties of long-lived discs are discussed in Section 6, where we also study six sample discs (Table 1). We discuss possible techniques for detecting and studying planetesimal discs in Section 7, and we summarize and discuss our results in Section 8.

[1] IR excesses characteristic of debris discs have also been detected around white dwarfs, and they appear to be strongly correlated with metal contamination in the white-dwarf photosphere, presumably arising from accreted planetesimals (Farihi, Jura & Zuckerman 2009).

2 A SIMPLE MODEL FOR A PLANETESIMAL DISC

We consider a system of planetesimals orbiting a solar-type star of mass $\mathrm{M}_\odot$. The extension to other types of stars is straightforward, but at this stage adds excessive complication. We will focus on discs with an age $t_0 = 3$ Gyr since these are likely to be more common than younger discs. We assume that the discs are gas-free, which is consistent with observations for most discs older than a few Myr (see fig. 2 of Wyatt 2008).

The surface density of the disc at semimajor axis a is written $\Sigma(a)$; more precisely, the mass with semimajor axes in the range $[a, a + \mathrm{d}a]$ is $\mathrm{d}M_{\rm disc} = 2\pi\Sigma a\,\mathrm{d}a$. The orbital period is $2\pi/\Omega$, where

$$\Omega = \sqrt{\frac{G\mathrm{M}_\odot}{a^3}}. \tag{1}$$

In most of the cases, we will assume that this material is collected in identical spherical bodies of density $\rho_{\rm p}$, radius r and mass $m = 4\pi\rho_{\rm p} r^3/3$ – a monodisperse planetesimal system (see the end of Section 4.3 for further discussion of this approximation). The densities of the planets in the Solar system range from 5.5 g cm^{-3} (Earth) to 0.7 g cm^{-3} (Saturn); in general gas-giant planets have smaller densities than terrestrial planets, but we will sacrifice accuracy for simplicity and assume that all planetesimals have $\rho_{\rm p} = 3$ g cm^{-3}. The surface number density of planetesimals is

$$\mathcal{N} = \frac{\Sigma}{m}. \tag{2}$$

We assume that the eccentricities e and inclinations i of the planetesimals follow a Rayleigh distribution,

$$\mathrm{d}^2 n \propto ei \exp\left[-(e/e_0)^2 - (i/i_0)^2\right]\,\mathrm{d}i\,\mathrm{d}e, \tag{3}$$

where e_0 and i_0 are the root mean square (rms) eccentricity and inclination. The density of the disc in the direction normal to its symmetry plane (which we call the z-direction) is given by

$$\mathrm{d}\mathcal{N} = n(z)\,\mathrm{d}z = n_0 \exp\left(-\frac{1}{2}z^2/h^2\right)\,\mathrm{d}z, \tag{4}$$

where $h \equiv ai_0/\sqrt{2}$ is the rms z-coordinate. The mid-plane number density n_0 is related to the surface number density by

$$n_0 = \frac{\mathcal{N}}{\sqrt{2\pi}h} = \frac{\Sigma}{\sqrt{\pi}mai_0}. \tag{5}$$

The radial velocity dispersion is

$$\sigma_r = \frac{\Omega a e_0}{\sqrt{2}}. \tag{6}$$

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

The ratio i_0/e_0 can in principle have a wide range of values depending on the dynamical history of the disc, but in a variety of theoretical models and observed astrophysical discs $i_0/e_0 \simeq 0.5$ (Dones & Tremaine 1993; Stewart & Ida 2000) so we will adopt this value throughout the paper (see also discussion following equation 29).

With these assumptions, the local properties of the planetesimal disc are specified by four parameters: the semimajor axis a, the surface density Σ, the planetesimal radius r or mass m and the rms eccentricity e_0. (We take the planetesimal density ρ_p, the disc age t_0 and the stellar mass $\mathrm{M}_\odot$ to be fixed.) Collision-limited discs are specified by three parameters, since the requirement that the collision time equals the age provides one constraint on the four parameters.

We will find it useful to express many of our results in terms of the following three dimensionless parameters:

$$\begin{aligned}
\mu &\equiv \frac{\Sigma a^2}{\mathrm{M}_\odot}, \\
\nu &\equiv \frac{\rho_\mathrm{p} a^3}{\mathrm{M}_\odot} = 5.0 \times 10^9 \left(\frac{\rho_\mathrm{p}}{3\,\mathrm{g\,cm^{-3}}}\right)\left(\frac{a}{10\,\mathrm{au}}\right)^3, \\
\tau &\equiv \frac{\Sigma a^2}{\mathrm{M}_\odot}\Omega t_0 = 6.0 \times 10^4 \left(\frac{\mu}{10^{-4}}\right)\left(\frac{a}{10\,\mathrm{au}}\right)^{-3/2}\left(\frac{t_0}{3\,\mathrm{Gyr}}\right). \qquad (7)
\end{aligned}$$

The parameter μ is a dimensionless mass, of order the ratio of the disc mass to the stellar mass. The parameter ν is a dimensionless density, which is almost always quite large for planetesimal discs, so we may assume $\nu \gg 1$ when necessary. The parameter τ is a dimensionless age, scaled by Ω and by $\Sigma a^2/\mathrm{M}_\odot$ since most evolutionary processes run slower if the orbital time is longer or the disc contains less mass.

We parametrize discs in terms of surface density Σ and semimajor axis a since these are likely to vary much less than planetesimal mass m or rms eccentricity e_0 during the planet formation process. As points of reference, we will refer to a set of six discs (Table 1) with semimajor axes $a = 1$, 10 and 100 au (a typical radius for spatially resolved debris discs), and with dimensionless mass $\mu = 10^{-4}$ and 10^{-6}. The former mass corresponds to $M_\mathrm{disc} \sim \pi\Sigma a^2 = \pi\mu\mathrm{M}_\odot \simeq 0.3 M_\mathrm{Jupiter} \simeq 100 M_\oplus$, about the solid mass needed to form the giant planets and comets in the Solar system (Goldreich et al. 2004b), while the latter mass corresponds to $M_\mathrm{disc} \simeq M_\oplus$.

3 DYNAMICAL PROCESSES IN THE PLANETESIMAL DISC

There are a few general criteria that planetesimal discs must satisfy. Firstly, we require that the planetesimal eccentricities and inclinations are not too large, which can be interpreted as the condition that most of the planetesimals are bound ($e < 1$), or that the disc is thin ($h < a$), or that the radial velocity dispersion is less than the circular speed ($\sigma_r < \Omega a$). We will write this criterion as

$$e_0 \lesssim f_e \quad \text{or} \quad \frac{h}{a} \lesssim 0.35 f_e \quad \text{or} \quad \frac{\sigma_r}{\Omega a} \lesssim \frac{f_e}{\sqrt{2}}. \qquad (8)$$

In this paper, we adopt $f_e = 0.5$. (See Table 2 for a summary of the dimensionless numbers used in our study.)

Naturally, non-trivial planetesimal discs must contain more than one body. Assuming that the surface density does not vary strongly with radius and the disc is not too extended, the disc mass may be written

$$M_\mathrm{disc} = \pi f_m \Sigma a^2 \qquad (9)$$

Table 2. Summary of dimensionless numbers.

Symbol	Assumed value	Defining equation and/or section
i_0/e_0	0.5	(3)
f_e	0.5	(8)
f_m	1	(9)
f_d	0.3	(11)
f_c	1	(14)
f_Q	1	(21), (23)
f_1	0.690	(27), Appendix A
f_2	1.521	(27), Appendix A
f_3	0.28	(37)
f_4	22.67	(30), Appendix A
f_5	12.94	(30), Appendix A
f_6	1	(42)
f_7	0.46	(56)
μ	various	(7)
ν	various	(7)
τ	various	(7)
a	−0.13	(44)
b	0.44	(44)

with f_m of order unity, so the criterion that the number of planetesimals $N = M_\mathrm{disc}/m > 1$ becomes

$$m < \pi f_m \Sigma a^2 \quad \text{or} \quad \pi f_m \mathcal{N} a^2 > 1. \qquad (10)$$

In this paper, we adopt $f_m = 1$, which corresponds to a disc with a radial width of about $0.5a$.

We will also require that the total disc mass is small compared to the stellar mass,

$$\frac{M_\mathrm{disc}}{\mathrm{M}_\odot} < f_d, \qquad (11)$$

where we adopt $f_d = 0.3$.

The total cross-sectional area of the disc is $\pi r^2 N$. Thus, in the absence of mutual shadowing the geometrical optical depth of the disc as seen from the host star is

$$\tau_\mathrm{p} = \frac{N r^2}{4a^2} = \frac{f_m}{4}\mathcal{N}\pi r^2. \qquad (12)$$

Note that this differs by a factor of order unity from the normal geometrical optical depth of the ring, $\mathcal{N}\pi r^2$.

3.1 Cold discs

A 'cold' disc is one in which planetesimal orbits do not cross. A planetesimal in an orbit with eccentricity e has a total radial excursion of $2ae$. The typical radial separation between planetesimals is

$$\Delta a = \frac{m}{2\pi\Sigma a}. \qquad (13)$$

Thus, most orbits do not cross if the rms eccentricity is

$$e_0 < e_\mathrm{hot} \equiv f_\mathrm{c}\frac{m}{4\pi\Sigma a^2}, \qquad (14)$$

and we will choose $f_\mathrm{c} = 1$, at which point just over half of the particles cross if their semimajor axes have a Poisson distribution. (A refinement of the preceding condition is to include the radius of the planetesimal in the crossing condition, that is, $2(ae + r) < \Delta a$, but in the cases of interest to us this correction is unimportant.) Equivalently, one can describe cold discs as those for which $\sigma_r < \sigma_\mathrm{hot}$, where

$$\sigma_\mathrm{hot} \equiv \frac{f_\mathrm{c}}{2^{5/2}\pi}\frac{m\Omega}{a\Sigma} = \frac{f_\mathrm{c}}{3\sqrt{2}}\frac{\sqrt{G\mathrm{M}_\odot}\rho_\mathrm{p} r^3}{a^{5/2}\Sigma}. \qquad (15)$$

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

3.1.1 Gravitational stability

Gravitational instability in cold discs is associated with dynamical chaos, which leads to growth in the eccentricities and inclinations of the planetesimals. The growth rate can be extremely slow – for example, tens of Gyr for Mercury in the current Solar system configuration (Laskar 2008). No rigorous analytical formulae for the rate of chaotic evolution in cold discs are available. However, N-body experiments suggest that cold discs can survive for millions of orbits if the separation (13) is typically a few times larger than the Hill radius

$$r_{\rm H} \equiv a \left(\frac{m}{3\,{\rm M}_\odot} \right)^{1/3}. \tag{16}$$

Thus, stability requires

$$\frac{\Delta a}{a} > k \left(\frac{m}{3\,{\rm M}_\odot} \right)^{1/3} \quad \text{or} \quad \frac{m}{{\rm M}_\odot} > 9.1\,(k\mu)^{3/2}. \tag{17}$$

Chambers, Wetherill & Boss (1996) find $k = 7$–11 from integrations of planetesimals with masses between 10^{-5} and $10^{-7}\,{\rm M}_\odot$ lasting $\sim 10^7$ yr.[2] They also find that the factor k varies slowly with planetesimal mass, so a stability criterion that fits their results more accurately is

$$\frac{\Delta a}{a} > 4 \left(\frac{m}{{\rm M}_\odot} \right)^{0.3} \quad \text{or} \quad \frac{m}{{\rm M}_\odot} > 100\mu^{1.43}. \tag{18}$$

This result depends on the duration of the integrations, but only weakly: increasing the duration by a factor of 10 typically increases the minimum stable separation by about one Hill radius.

The criterion (18) was derived from simulations with planetesimals of mass 10^{-5}–$10^{-7}\,{\rm M}_\odot$; the extrapolation to larger planets is somewhat uncertain but probably not a major source of error. For Jupiter-mass planets, the formula predicts $\Delta a/a > 0.5$, corresponding to $k \simeq 7$ in the notation of equation (17). For comparison, Jurić & Tremaine (2008) estimate $k \simeq 12$–14 from orbit integrations of planets with masses between 0.1 and 10 Jupiter masses; their larger value of k probably arises because they used a range of masses rather than a single common mass for the planets. In any event, equation (18) should be correct to within a factor of 2 or so.

When the planetesimals in a cold disc have non-zero eccentricities, larger separations are required for stability (e.g. Yoshinaga, Kokubo & Makino 1999); however, we will not include this refinement since the detectability of cold discs does not depend on the rms eccentricity (equation 14).

3.2 Hot discs

A hot disc is one in which planetesimal orbits cross. In hot discs, the rms eccentricity or radial velocity dispersion exceeds $e_{\rm hot}$ or $\sigma_{\rm hot}$, respectively (equations 14 and 15). As we discuss below (equation 40), 'warm' discs may also be defined, in which the orbits cross but the collisions do not substantially damage the planetesimals over the lifetime of the disc.

3.2.1 Gravitational stability

The disc must be gravitationally stable to axisymmetric perturbations (Toomre 1964). The Toomre stability criterion is derived from the Wentzel–Kramers–Brillouin (WKB) dispersion relation for axisymmetric density waves. For a fluid disc in a Keplerian potential, this may be written as (Binney & Tremaine 2008)

$$\frac{\omega^2}{\Omega^2} = 1 - \frac{\lambda_{\rm crit}}{\lambda} + \frac{Q^2}{4}\frac{\lambda_{\rm crit}^2}{\lambda^2}, \tag{19}$$

where ω is the frequency, λ is the radial wavelength, and

$$Q \equiv \frac{\sigma_r \Omega}{\pi G \Sigma} = \frac{e_0}{\sqrt{2}\pi\mu},$$
$$\lambda_{\rm crit} = \frac{4\pi^2 G\Sigma}{\Omega^2} = 4\pi^2 a\mu. \tag{20}$$

This dispersion relation was derived for a barotropic fluid whereas the planetesimal disc more closely approximates a collisionless fluid, but for our purposes the results should be accurate enough.

The disc is stable to perturbations at a given wavelength λ if $\omega^2 > 0$, which requires

$$Q^2 > 4\frac{\lambda}{\lambda_{\rm crit}}\left(1 - \frac{\lambda}{\lambda_{\rm crit}}\right). \tag{21}$$

The maximum of the right-hand side occurs at $\lambda = \lambda_{\rm crit}/2$ and equals unity, so the disc is stable to perturbations of all wavelengths if $Q > 1$. However, this result needs to be modified if the number of planetesimals is so small that the disc cannot be approximated as a continuous fluid. The typical separation between planetesimals is given by equation (13), so the number of planetesimals in one wavelength λ is $N_\lambda = \lambda/\Delta a = 2\pi a\Sigma\,\lambda/m$. The continuum approximation should be valid if $N_\lambda \gg 1$ or $\lambda > \lambda_{\rm c} \equiv f_Q m/(2\pi a\Sigma)$ with f_Q of order unity. We adopt $f_Q = 1$ when a choice is necessary. Then, a necessary requirement for stability is that equation (21) is satisfied for all $\lambda > \lambda_{\rm c}$ or

$$Q^2 > \begin{cases} 1 & \text{if } \lambda_{\rm c} < \lambda_{\rm crit}/2, \\ 4(\lambda_{\rm c}/\lambda_{\rm crit}) - 4(\lambda_{\rm c}/\lambda_{\rm crit})^2 & \text{if } \lambda_{\rm c} > \lambda_{\rm crit}/2. \end{cases} \tag{22}$$

This may be rewritten as

$$e_0 > \sqrt{2}\pi\mu, \qquad \text{if} \quad \frac{m}{{\rm M}_\odot} < \frac{4\pi^3\mu^2}{f_Q},$$
$$e_0 > \sqrt{\frac{f_Q m}{\pi {\rm M}_\odot}\left[1 - \frac{f_Q m}{(2\pi)^3\,{\rm M}_\odot\mu^2}\right]}, \quad \text{otherwise;} \tag{23}$$

there is no constraint if the square root in the second equation is negative.

The relation between the stability criteria for hot and cold discs can be clarified using Fig. 1, which plots allowed regions as a function of the dimensionless parameters $m/{\rm M}_\odot$ (ratio of planetesimal mass to stellar mass; horizontal axis) and e_0 (rms eccentricity, vertical axis). The axes are logarithmic. We aim for clarity at the sake of accuracy by neglecting all factors of order unity for the rest of this subsection. The diagram shows the following constraints.

(i) There must be at least one planetesimal in the disc (equation 10), so $m/\Sigma a^2 \lesssim 1$ or $m/{\rm M}_\odot \lesssim \mu$ where μ is defined in equation (7). The boundary $m/{\rm M}_\odot = \mu$ is represented by a green, vertical solid line; the excluded region to the right-hand side of this line is labelled '$N < 1$'.

(ii) The division between hot and cold discs (14) may be written $e_0 \lesssim m/\mu{\rm M}_\odot$, which is marked by a slanted solid line.

(iii) The condition (18) for the gravitational stability of cold discs is $m/{\rm M}_\odot \gtrsim \mu^{1.43}$, which appears in the figure as a vertical dotted line; the unstable region to the left-hand side of this line is labelled 'unstable' (in blue).

[2] Note that Chambers et al. (1996) define the Hill radius as $(2m/3\,{\rm M}_\odot)^{1/3}$, a factor of $2^{1/3}$ larger than our definition.

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

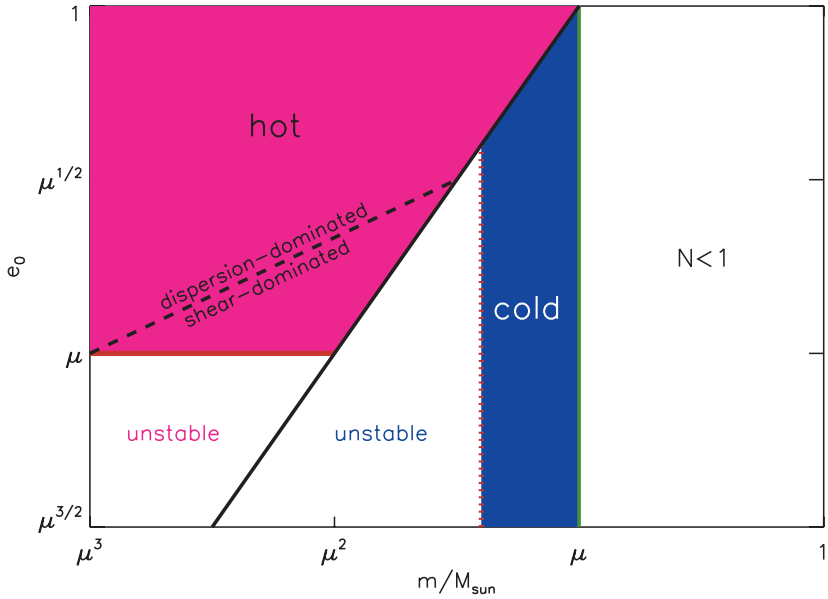

Figure 1. Stability and other properties of planetesimal discs, as a function of planetesimal mass m and rms eccentricity e_0. The parameter $\mu \equiv \Sigma a^2/\mathrm{M}_\odot \sim M_{\rm disc}/\mathrm{M}_\odot$ is approximately the disc mass relative to the stellar mass. The region to the right-hand side of the green, vertical solid line, labelled '$N < 1$', is not allowed because the planetesimal mass exceeds the assumed disc mass. The slanted solid line divides hot discs, in which planetesimal orbits cross, from cold discs (equation 14). The regions labelled 'unstable' represent hot discs that are unstable (in pink) and cold discs that are unstable (in blue) according to the approximate equations derived at the end of Section 3.2.1. The slanted dashed line separates hot discs in which the encounters are shear-dominated from those in which encounters are dispersion-dominated. The pink and blue shaded regions represent the allowed parameter values for hot and cold discs, respectively.

(iv) The condition (23) for gravitational stability of hot discs becomes $e_0 \gtrsim \mu$ if $m/\mathrm{M}_\odot \lesssim \mu^2$. The second of equations (23) is neglected because it applies only over a range of a factor of 2 in planetesimal mass. The unstable region is bounded by a solid horizontal line and labelled 'unstable' in pink. Although this derivation was carried out for hot discs, it should apply to cold discs as well so long as $m/\mathrm{M}_\odot \lesssim \mu^2$; however, it adds no new restrictions on cold discs since these are already unstable by condition (18).

The pink and blue shaded regions represent the allowed parameter values for hot and cold discs, respectively.

3.2.2 *Collisions*

The collision time in a system with isotropic velocity dispersion σ and number density n is given by (Binney & Tremaine 2008, equation 7.195)

$$t_c^{-1} = 16\sqrt{\pi} n \sigma r^2 (1 + \Theta), \tag{24}$$

where

$$\Theta \equiv \frac{Gm}{2\sigma^2 r} \tag{25}$$

is the Safronov number. The factor $(1 + \Theta)$ reflects the enhancement in the collision rate due to gravitational focusing. If the velocity-dispersion tensor is anisotropic, we define Θ by replacing σ by the radial dispersion σ_r, and then with equation (6) we have

$$\Theta = \frac{m}{\mathrm{M}_\odot}\frac{a}{r}\frac{1}{e_0^2} = 1.61\left(\frac{m}{\mathrm{M}_\odot}\right)^{2/3}\frac{\nu^{1/3}}{e_0^2}, \tag{26}$$

where the dimensionless density ν is defined in equation (7).

Equation (24) requires several corrections for use in discs. First, the velocity-dispersion tensor in a Keplerian disc is not isotropic; for the Rayleigh distribution (3) and our choice $i_0/e_0 = 0.5$, this requires replacing σ with the radial dispersion σ_r and the factor $(1 + \Theta)$ with $(f_1 + f_2\Theta)$, where $f_1 = 0.690$ and $f_2 = 1.521$. Secondly, we must account for the falloff in density above the disc mid-plane, which we do by replacing the number density n in equation (24) with the number-weighted average $\int n^2(z)\mathrm{d}z / \int n(z)\mathrm{d}z = n_0/\sqrt{2}$, where $n(z)$ is given by equation (4). Thus, we replace equation (24) by

$$\begin{aligned} t_c^{-1} &= 2^{7/2}\pi^{1/2} n_0 \sigma_r r^2 (f_1 + f_2\Theta) \\ &= 16\mathcal{N}\Omega r^2 (f_1 + f_2\Theta) \end{aligned} \tag{27}$$

for $i_0/e_0 = 0.5$, as we assume throughout. See Appendix A for details on how to modify the more general formulae of Dones & Tremaine (1993) to arrive at the preceding result.

Equation (27) neglects the Keplerian shear in the disc and thus is only valid for dispersion-dominated encounters, for which

$$\sigma_r \gtrsim \Omega \max\{r, r_{\rm H}\}, \tag{28}$$

where the Hill radius $r_{\rm H}$ is defined in equation (16). Note that $r_{\rm H} > r$ if the dimensionless parameter ν (equation 7) exceeds 0.7, which is almost always true, so in practice condition (28) reduces to

$$\sigma_r \gtrsim \Omega r_{\rm H} \quad \text{or} \quad e_0 \gtrsim (m/\mathrm{M}_\odot)^{1/3}. \tag{29}$$

This boundary is marked as a slanted dashed line in Fig. 1. The allowed region in parameter space below this line, in which encounters are shear-dominated, is relatively small but still requires consideration. Note that condition (29) can be rewritten with the help of (26) as $\Theta \lesssim \nu^{1/3}$ so shear-dominated encounters occur only if the rms eccentricity is so small that the Safronov number is greater than unity by the large factor $\nu^{1/3}$. When the condition in equation (29) is violated, the disc is so flat that encounters may excite out-of-plane motions less efficiently than in-plane motions, so that i_0/e_0 may be less than our assumed value of 0.5. We do not, however, attempt to model variations in i_0/e_0 in this paper.

Formulae for the shear-dominated collision rate are given by Greenzweig & Lissauer (1992) and Dones & Tremaine (1993).

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

Adapting these formulae to the present model and notation (see Appendix A), we have

$$t_c^{-1} = \begin{cases} f_4 \mathcal{N} \Omega^2 a^2 r \sigma_r^{-1} \left(m/\mathrm{M}_\odot\right)^{2/3}, & \Omega r_\mathrm{H}/\nu^{1/6} \lesssim \sigma_r \lesssim \Omega r_\mathrm{H}, \\ f_5 \mathcal{N} \Omega r^{1/2} a^{3/2} \left(m/\mathrm{M}_\odot\right)^{1/2}, & \sigma_r \lesssim \Omega r_\mathrm{H}/\nu^{1/6}, \end{cases} \tag{30}$$

where $f_4 = 22.67$, $f_5 = 12.94$ and ν is defined in equation (7).

Collisions between equal-mass planetesimals may have various outcomes. If the gravitational binding energy of the planetesimals is negligible compared to their relative kinetic energy (Safronov number $\Theta \ll 1$) then (i) high-speed collisions will shatter the planetesimals and disperse the fragments, while (ii) low-speed collisions will leave the planetesimals unaffected or produce small craters. If the gravitational binding energy is much larger than the kinetic energy then (iii) high-speed collisions will still shatter and disperse the planetesimals, while (iv) the outcome of low-speed collisions will be a single gravitationally bound object containing most of the mass of the two planetesimals (see Section 4.1 for further detail). For our purposes, any of the outcomes (i), (iii) or (iv) leads to a substantial change in the mass distribution of planetesimals if the collision time is less than the age of the disc, $t_c \lesssim t_0$ – and therefore is inconsistent with our requirement that the disc is long-lived in its present state.

However, many collisions of type (ii) could occur without substantially altering the planetesimal mass distribution. We therefore differentiate between dynamically hot discs, in which the collision time is longer than the age, $t_\mathrm{c} \gtrsim t_0$, and 'warm' discs, in which the collision time is shorter than the age but the velocity dispersion σ_r is small enough that the collisions do not substantially damage the planetesimals over the lifetime of the disc. The survival criteria for warm discs are discussed in Section 3.3.

3.2.3 Gravitational scattering

The time-scale for a substantial change in the rms eccentricity and inclination due to gravitational scattering of planetesimals must also be longer than the disc age. We write this as $t_g > t_0$, where

$$\begin{aligned} t_g^{-1} &= \frac{\mathrm{d} \log e_0^2}{\mathrm{d}t} \\ &= \frac{\Omega}{e_0^4} \frac{\mu m}{\mathrm{M}_\odot} \left[S_1(i_0/e_0)\mathcal{C} + S_2(i_0, e_0, m/\mathrm{M}_\odot) \right]. \end{aligned} \tag{31}$$

Here, the term $S_1(i_0/e_0)\mathcal{C}$ incorporates the effects of close or dispersion-dominated encounters, in which the relative velocity is dominated by the velocity dispersion σ_r, while $S_2(i_0, e_0, m/\mathrm{M}_\odot)$ represents the effects of distant or shear-dominated encounters. The factor $\mathcal{C}$ is usually called the Coulomb logarithm. We evaluate these terms by specializing the formulae of Stewart & Ida (2000) to the case where the colliding planetesimals have the same mass and the same eccentricity and inclination distributions. From equations (3.29), (3.31) and (6.6)–(6.8) of that paper, we find $S_1(0.5) = 4.50$ and

$$\begin{aligned} 2\mathcal{C} &= \log(\Lambda^2 + 1) - \log\left(\Lambda_\mathrm{c}^2 + 1\right) + \frac{1}{\Lambda^2 + 1} - \frac{1}{\Lambda_\mathrm{c}^2 + 1}, \\ \Lambda &= \frac{\mathrm{M}_\odot}{m} \left(e_0^2 + i_0^2\right) \left[\sqrt{2} i_0 + \left(\frac{2}{3} m/\mathrm{M}_\odot \right)^{1/3} \right], \\ \Lambda_\mathrm{c} &= \frac{\mathrm{M}_\odot}{m} \frac{2r}{a} \left(e_0^2 + i_0^2\right) \\ &\quad \times \left\{ 1 + \frac{ma}{\mathrm{M}_\odot r \left[e_0^2 + i_0^2 + (1/2)\left(\frac{2}{3} m/\mathrm{M}_\odot\right)^{2/3} \right]} \right\}^{1/2}. \end{aligned} \tag{32}$$

From equations (4.8) and (6.5) of Stewart & Ida (2000),

$$S_2(i_0, e_0, m/\mathrm{M}_\odot) = \frac{3.8\xi}{1 - i_0^2/e_0^2} \left[W\left(\epsilon/e_0^2\right) - W\left(\epsilon/i_0^2\right) \right], \tag{33}$$

where $\epsilon \equiv (1/2)\, \xi [(2/3) m/\mathrm{M}_\odot]^{2/3}$,

$$W(Y) \equiv \int_Y^\infty \exp(Y - y) \frac{\mathrm{d}y}{y}, \tag{34}$$

and $\xi \simeq 2$ is determined by an empirical fit to N-body simulations. Using equation (26), equation (31) may be rewritten as

$$t_g^{-1} = \mathcal{N} \Omega r^2 \Theta^2 S_1 \mathcal{C}', \tag{35}$$

where

$$\mathcal{C}' \equiv \mathcal{C} + S_2/S_1 \tag{36}$$

is the correction factor arising from the Coulomb logarithm and shear-dominated encounters.

The characteristic time-scales (27) and (35) for physical collisions and gravitational scattering can be combined into a single relaxation time-scale,

$$t_\mathrm{relax}^{-1} = t_\mathrm{c}^{-1} + t_g^{-1} = 16 \mathcal{N} \Omega r^2 \left(f_1 + f_2 \Theta + f_3 \mathcal{C}' \Theta^2 \right), \tag{37}$$

where $f_3 = S_1/16 = 0.28$. A long-lived disc must have $t_\mathrm{relax} \gtrsim t_0$.

The effects of collisions and gravitational scattering can be clarified using Fig. 2, which plots allowed regions as a function of $m/\mathrm{M}_\odot$ and e_0, in logarithmic coordinates. We neglect all factors of order unity for the rest of this subsection. In this approximation, the requirement for survival of hot discs simplifies to

$$(\Omega t_0)^{-1} \gtrsim \frac{\Sigma}{m^{1/3} \rho_\mathrm{p}^{2/3}} \max \left\{ 1, \nu^{2/3} \left(\frac{m}{\mathrm{M}_\odot} \right)^{4/3} e_0^{-4} \right\}. \tag{38}$$

This result holds only if the collisions are dispersion-dominated, but if they are shear-dominated the relaxation time becomes shorter than this formula would predict so the criterion (38) remains necessary (but not sufficient) for survival of the disc.

The results can be rewritten in terms of the dimensionless time τ defined in equation (7). Discs with Safronov number $\Theta > 1$ have $m/\mathrm{M}_\odot > e_0^3/\sqrt{\nu}$ and the survival criterion (38) is $m/\mathrm{M}_\odot \gtrsim \tau^3/\nu^2$ ($\Theta \lesssim 1$) or $m/\mathrm{M}_\odot \lesssim e_0^4/\tau$ ($\Theta \gtrsim 1$). These constraints are shown in Fig. 2. The excluded regions are bounded by the cyan, vertical solid ($m/\mathrm{M}_\odot = \tau^3/\nu^2$) and green, slanted dot–dash ($e_0^4 = m\tau/\mathrm{M}_\odot$) lines, and are labelled '$t_\mathrm{relax} < t_0$'. Note that for a given value of the rms eccentricity e_0 or velocity dispersion σ_r, there is only a finite range of planetesimal masses or radii in which the relaxation time exceeds the age: the disc cannot survive if the planetesimal mass is either too small or too large. This behaviour arises because for small masses m or radii r ($\Theta \ll 1$), we have $t_\mathrm{relax} \propto r \propto m^{1/3}$ at a given surface density and eccentricity, while for large masses ($\Theta \gg 1$), we have $t_\mathrm{relax} \propto m^{-1} \propto r^{-3}$.

3.3 Warm discs

As described in Section 3.2.2, planetesimals in warm discs may suffer collisions but the velocity dispersion is low enough that these collisions do not destroy the planetesimals. Thus, a necessary condition for the survival of warm discs is that the squared velocity dispersion σ_r^2 must be less than Q_D^*, the energy per unit mass required to disperse the planetesimal into fragments, which we obtain from equation (44) below.

This criterion is not sufficient, for two reasons. Firstly, if the gravitational binding energy of the planetesimals is much larger than the kinetic energy, an inelastic collision is likely to leave the colliding

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

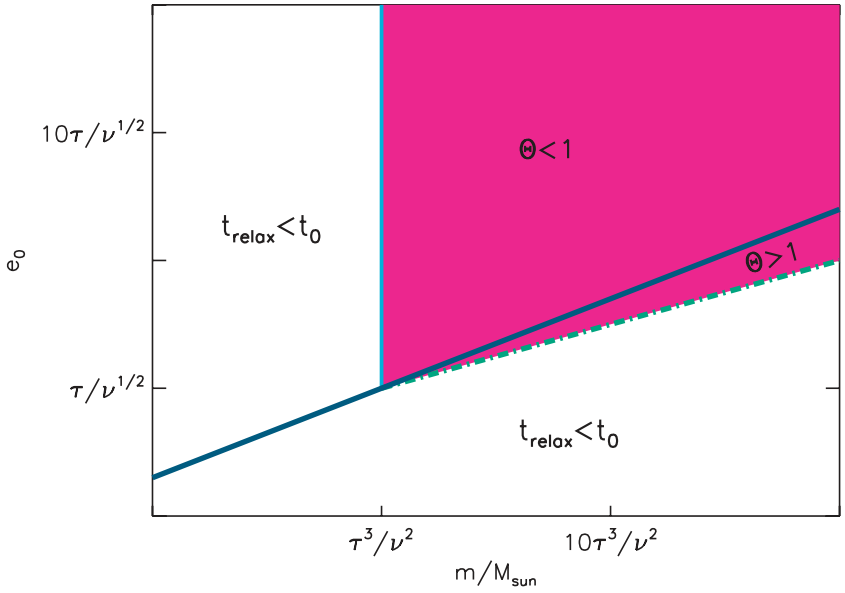

Figure 2. Survival of hot discs as a function of planetesimal mass m and rms eccentricity e_0. The parameters τ and ν are defined in equation (7). The slanted solid line divides discs according to their Safronov number Θ. The regions marked by '$t_{\rm relax} < t_0$' are not allowed because the collision time (when $\Theta < 1$) or the gravitational scattering time (when $\Theta > 1$) is shorter than the disc age. The pink shaded region represents the allowed parameter values for hot discs.

planetesimals as a gravitationally bound pair, which alters the mass distribution and therefore is inconsistent with our assumption that the disc has not evolved. Therefore, we require that the Safronov number $\Theta \lesssim 1$ for warm discs.

Secondly, low-velocity collisions can chip or crater the planetesimals even if they are not disrupted. Thus, if the collision time $t_{\rm c}$ is much less than the age t_0 the planetesimals may be gradually eroded away even if they are not dispersed in a single collision. A simple parametrization of the erosive process is to assume that the mass lost in a typical collision of two objects of mass m at relative velocity v is (e.g. Thébault & Augereau 2007)

$$\Delta m = 0.5m \left(\frac{v^2}{Q_D^*}\right)^{\gamma}, \quad v^2 \lesssim Q_D^* \tag{39}$$

(the factor 0.5 arises because the usual definition of a dispersive impact is one in which the mass of the largest fragment is less than half of the target mass; see Section 4.1).

Erosion through impact involves a number of complicated processes such as crack propagation in brittle materials and plastic flow in ductile materials, and melting or sublimation at high impact velocities. Our present state of knowledge is derived from a variety of approximate physical models, experiments and numerical simulations, most intended for situations far removed from planetesimal collisions. These typically yield values of γ between 1 and 1.5. For example (i) impacts of small pellets into targets composed of ice-silicate mixtures at speeds up to 12 km s^{-1} yield $\gamma \simeq 1.2$ (Koschny & Grün 2001), (ii) numerical simulations of collisions between rocky and icy bodies yield $\gamma \simeq 1$ (Benz & Asphaug 1999; Stewart & Leinhardt 2009). We will adopt $\gamma = 1$ recognizing that this is (a) oversimplified; (b) conservative (in that the actual erosion rate is expected to be smaller for low-velocity collisions if γ is larger).

The criterion for survival is that the cumulative mass loss $\Delta m(t_0/t_{\rm c}) \lesssim m$. Thus warm discs must satisfy

$$t_{\rm c} \lesssim t_0, \quad \Theta \lesssim 1, \quad \text{and} \quad \sigma_r^2 \lesssim (t_{\rm c}/t_0)\, Q_D^*. \tag{40}$$

where Q_D^* is given by equation (44).[3]

An additional constraint for warm discs is that collisions do not result in excessive 'viscous' spreading of the disc. A disc initially localized at semimajor axis a that loses a small amount of energy δE must spread by $\delta a \ll a$ (Brahic 1977), where

$$\delta E = -\frac{G{\rm M}_\odot M_{\rm disc}}{32a^3}\delta a^2. \tag{41}$$

If the typical energy lost per unit mass in a collision is $f_6\sigma_r^2$, where $f_6 \sim 1$, then the rate of energy loss is $f_6\sigma_r^2\, M_{\rm disc}/t_{\rm c}$. Therefore, in a time t_0 the disc spreads to

$$\frac{\delta a}{a} = \sqrt{\frac{32 f_6 a \sigma_r^2}{G{\rm M}_\odot}\frac{t_0}{t_{\rm c}}}. \tag{42}$$

Requiring $\delta a/a \lesssim 1$ and assuming $\Theta \ll 1$ yields

$$\sigma_r \lesssim 8\times10^2\,{\rm cm\,s^{-1}}\left(\frac{f_m}{f_6}\right)^{1/2}\left(\frac{f_1}{0.69}\right)^{-1/2}\left(\frac{\rho_{\rm p}}{3\,{\rm g\,cm^{-3}}}\frac{r}{10^5\,{\rm cm}}\right)^{1/2} \times \left(\frac{\mu}{10^{-4}}\frac{t_0}{3\,{\rm Gyr}}\right)^{-1/2}\left(\frac{a}{10\,{\rm au}}\right)^{5/4}. \tag{43}$$

This constraint does not restrict the allowed range for warm discs among the six sample discs considered in this paper, i.e. it is satisfied so long as the warm discs satisfy all of the other constraints we have already discussed.

4 COLLISION-LIMITED DISCS

4.1 The effects of collisions

Collisions or impacts may crater, shatter or disperse the target, as discussed in Sections 3.2.2 and 3.3. The distinction between shattering and dispersive impacts arises because the pieces of the

[3] The distinction between 'hot' and 'warm' discs based on whether or not the collisions are destructive is moot when the collision time is longer than the age. Our (arbitrary) convention is that such discs are 'hot', not 'warm'.

© 2009 The Authors. Journal compilation © 2009 RAS, MNRAS **401,** 867–889

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

target may remain gravitationally bound even after the target is shattered (Benz & Asphaug 1999).

The outcome of an impact depends mainly on the relative speed, masses and composition of the impacting bodies. Consider an impact between two planetesimals of masses m_1 (the target) and $m_2 \leq m_1$ (the projectile), at relative velocity Δv. The kinetic energy of relative motion is $E_k = (1/2)\tilde{m}\Delta v^2$, where $\tilde{m} \equiv m_1 m_2/(m_1 + m_2)$ is the reduced mass. The impact is shattering if the kinetic energy per unit target mass $E_k/m_1 > Q_S^*$, and dispersive if $E_k/m_1 > Q_D^*$, where Q^* is a function of the mass and composition of the target. Roughly speaking, $Q^*(m)$ is the binding energy per unit mass of the target – the energy needed to rupture its internal chemical bonds for Q_S^* and this energy plus its gravitational potential energy for Q_D^*. In the strength-dominated regime, when the target is small enough that its self-gravity is negligible, we expect $Q_S^* \sim Q_D^*$, and both should be independent of m in the idealized case in which the strength of the target is independent of its size. In practice, both simulations and experiments find that Q^* declines slowly with m in the strength-dominated regime. In the gravity-dominated regime, $Q_S^* \ll Q_D^*$ (an impact that shatters a target may not impart escape speed to the fragments, so they re-accumulate as a gravitationally bound rubble pile) and we might expect that $Q_D^* \propto m^{2/3}$ since the gravitational binding energy per unit mass of a homogeneous body scales as $m^{2/3}$.

The dependence of Q_D^* on target mass is typically parametrized as

$$Q_D^*(m) = Q_0 \left(x_0^a + x_0^b\right), \tag{44}$$

where $x_0 \equiv m/m_0$; Q_0, m_0, a and b are parameters to be fitted to experiments or simulations. We will adopt $m_0 = 10^{14}$ g, $Q_0 = 6 \times 10^5$ erg g^{-1}, $a = -0.13$ and $b = 0.44$, these values being compromises between the results for ice and basalt given by Benz & Asphaug (1999). This simple form fails for very small mass, since it predicts $Q_D^*(m) \to \infty$ as $m \to 0$; experiments with high-velocity impacts of small bodies suggest $Q_D^* \approx 10^7$ erg g^{-1} independent of mass (Flynn & Durda 2004), so for $x_0 \ll 1$ we use the smaller of this value and the prediction of equation (44) – the transition occurs at $m \approx 10^4$ g.

4.2 The collisional cascade

When the collision time is much less than the age of the disc, the mass distribution of the smaller planetesimals is likely to be established by a 'collisional cascade', in which large bodies are dispersed by collisions into smaller bodies, these in turn being dispersed into smaller ones, until bodies of size $\lesssim 1$ μm ('dust grains') are removed by Poynting–Robertson drag and radiation pressure (see Section 5). We now derive an approximate form for the steady-state mass distribution in a collisional cascade (Dohnanyi 1969; O'Brien & Greenberg 2003; Pan & Sari 2005), using the following assumptions.

(i) The number density of planetesimals per unit mass is a power-law function of mass, at least over a limited range,

$$\frac{\mathrm{d}n}{\mathrm{d}m} \propto m^{-p}. \tag{45}$$

(ii) The rms eccentricity of the planetesimals is independent of mass. (This assumption is made for simplicity, but appears to hold approximately in the asteroid and Kuiper belts.)

(iii) The velocity dispersion (6) is large enough that a typical impact between two bodies of equal mass has much more than enough energy to disperse the two bodies, i.e. $\sigma_r^2 \gg Q_D^*$, which requires $\sigma_r \gg 3 \times 10^3$ cm s^{-1} $(Q_D^*/10^7$ erg g$^{-1})^{1/2}$ or $e_0 \gg 0.005(a/10\,\mathrm{au})^{1/2} \times (Q_D^*/10^7$ erg g$^{-1})^{1/2}$.

(iv) The specific impact energy required for a dispersive impact is a power-law function of mass, $Q_D^* \propto m^j$. According to equation (44), we expect $j = a$ for $m \ll m_0$ (strength-dominated) and $j = b$ for $m \gg m_0$ (gravity-dominated).

(v) The cross-section for a collision between two bodies of masses m_1 and $m_2 \ll m_1$ is $\sigma_{\mathrm{coll}}(m_1, m_2) \propto m_1^l$. From equation (24), we expect $l = 2/3$ for $\Theta \ll 1$ ($\sigma_{\mathrm{coll}} \propto r^2 \propto m^{2/3}$) and $l = 4/3$ for $\Theta \gg 1$ ($\sigma_{\mathrm{coll}} \propto r^2\Theta \propto rm \propto m^{4/3}$).

According to assumption (iii), the smallest projectile mass that can disperse a target of mass m in an impact is $m_{\min}(m) \approx m Q_D^*/\sigma_r^2 \ll m$. The rate of collisions in which planetesimals of mass $>m$ are dispersed is approximately

$$\Phi_n(m) \propto \int_m^\infty \mathrm{d}m_1 \int_{m_{\min}(m_1)}^{m_1} \mathrm{d}m_2 \frac{\mathrm{d}n\,(m_1)}{\mathrm{d}m_1} \frac{\mathrm{d}n(m_2)}{\mathrm{d}m_2} \sigma_{\mathrm{coll}}\,(m_1, m_2)\,. \tag{46}$$

This formula is accurate to a factor of order unity only, since it neglects the fact that one object of mass $>m$ is dispersed when $m_1 > m$ and $m_2 < m$, while two are dispersed when m_1, $m_2 > m$; it also neglects the possibility that one or more of the collision fragments is more massive than m. These inaccuracies do not affect our final result.

Since most of the fragments in the collisions that dominate this rate will have masses $< m$, the mass per unit volume in planetesimals larger than m decreases at a rate given approximately by

$$\Phi_m(m) \approx m\Phi_n(m) \propto m \int_m^\infty \mathrm{d}m_1 \times \int_{m_{\min}(m_1)}^{m_1} \mathrm{d}m_2 \frac{\mathrm{d}n\,(m_1)}{\mathrm{d}m_1} \frac{\mathrm{d}n\,(m_2)}{\mathrm{d}m_2} \sigma_{\mathrm{coll}}\,(m_1, m_2)\,. \tag{47}$$

With assumptions (i), (iv) and (v), we find

$$\Phi_m(m) \propto m^{3+j+l-p(2+j)}. \tag{48}$$

In a steady state, the mass flux $\Phi_m(m)$ must be independent of mass, so

$$p = \frac{3 + l + j}{2 + j}. \tag{49}$$

With the parameters used in this paper, almost all dispersive collisions have $\Theta \ll 1$ so we can set $l = 2/3$ and obtain (O'Brien & Greenberg 2003)

$$p = \frac{11 + 3j}{6 + 3j}. \tag{50}$$

Matching the power-law behaviour for $m \ll m_0$ and $m \gg m_0$, we have (Löhne, Krivov & Rodmann 2008)

$$\frac{\mathrm{d}n(m)}{\mathrm{d}m} = \begin{cases} (m/m_0)^{-(11+3b)/(6+3b)}, & m > m_0, \\ (m/m_0)^{-(11+3a)/(6+3a)}, & m < m_0. \end{cases} \tag{51}$$

These expressions are only valid if (i) the collision time at mass m is short compared to the age of the disc, (ii) the typical impact velocity is sufficient to disperse a body of mass m, $\sigma_r^2 \gg Q_D^*(m)$, (iii) the velocity dispersion is independent of mass, (iv) the mass m is sufficiently large that the lifetime to radiation pressure and Poynting-Robertson drag is much larger than the collisional lifetime.

For the exponents $a = -0.13$, $b = 0.44$ given after equation (44), we have $p = 1.89$ for $m \ll m_0$ and $p = 1.68$ for $m \gg m_0$. With these exponents the total mass in the collisional cascade ($\propto \int m\,\mathrm{d}n \propto m^{2-p}$) is dominated by the largest bodies, while

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

the total cross-section ($\propto \int m^{2/3}\,\mathrm{d}n \propto m^{5/3-p}$) is dominated by the smallest bodies.

4.3 Properties of collision-limited discs

Collision-limited discs can arise if we assume that there is an initial distribution of planetesimal masses in which the total disc mass is dominated by small bodies, say $\mathrm{d}n \propto m^{-P}\mathrm{d}m$ with $P > 2$.[4] In such discs, the collision time is shorter for planetesimals of smaller mass; thus a collisional cascade is established for all masses below some maximum $m_{\max}$. We have seen that the mass in the collisional cascade is dominated by the largest bodies in the cascade, while if $P > 2$ the mass in the 'primordial' regime $m > m_{\max}$ is dominated by the smallest bodies. Thus, the overall mass in the disc is dominated by bodies with mass $\sim m_{\max}$, and for many purposes we may treat the disc as a monodisperse system composed of bodies with a single mass $m_{\max}$. The difference from our earlier discussions is that now the planetesimal mass is not a free parameter; rather, it is determined by the condition that the lifetime of a planetesimal of mass $m_{\max}$ subject to dispersive impacts is equal to the disc age. We now derive this condition.

Following the discussion in the preceding section, we will assume that dispersive impacts have Safronov number $\Theta \ll 1$ and are dominated by impactors of mass much less than the target mass. Then if we equate the collision time from equation (27) to the disc age t_0, we have

$$t_0^{-1} = 8\sqrt{2\pi} f_1 \sigma_r \left(\frac{3m_{\max}}{4\pi\rho_{\mathrm{p}}}\right)^{2/3} \int_{m_{\min}(m_{\max})}^{m_{\max}} \frac{\mathrm{d}n_0\,(m')}{\mathrm{d}m'}\,\mathrm{d}m', \tag{52}$$

where $m_{\min} = m_{\max} Q_D^*(m_{\max})/\sigma_r^2$ is the smallest impactor that will disperse a target of mass $m_{\max}$. Assuming that the mass distribution in the mid-plane of the disc is

$$\frac{\mathrm{d}n_0\,(m)}{\mathrm{d}m} = \frac{n_{\max}}{m_{\max}} \begin{cases} (m/m_{\max})^{-p} & m \le m_{\max}, \\ (m/m_{\max})^{-P} & m \ge m_{\max}, \end{cases} \tag{53}$$

we have

$$t_0^{-1} \simeq \frac{5.32}{p-1}\frac{f_1}{0.69}\frac{n_{\max} m_{\max}^{2/3}\sigma_r}{\rho_{\mathrm{p}}^{2/3}}\left[\frac{Q_D^*(m_{\max})}{\sigma_r^2}\right]^{1-p}. \tag{54}$$

The total mid-plane mass density ρ_0 is related to the surface density by $\Sigma = \sqrt{\pi}\rho_0 a i_0$ (equation 5) and ρ_0 is given by

$$\begin{aligned}\rho_0 &= \frac{n_{\max}}{m_{\max}}\left(m_{\max}^p \int_0^{m_{\max}} m^{1-p}\mathrm{d}m + m_{\max}^P \int_{m_{\max}}^{\infty} m^{1-P}\mathrm{d}m\right) \\ &= n_{\max} m_{\max}\left(\frac{1}{2-p} + \frac{1}{P-2}\right),\end{aligned} \tag{55}$$

assuming $p < 2 < P$. If we divide the second equation by the first and use equations (5) and (6) we obtain

$$\Omega t_0 \simeq 1.25 f_7 \left(\frac{f_1}{0.69}\right)^{-1} \frac{m_{\max}^{1/3}\rho_{\mathrm{p}}^{2/3}}{\Sigma}\left[\frac{Q_D^*\,(m_{\max})}{\sigma_r^2}\right]^{p-1}, \tag{56}$$

which is an implicit equation for the characteristic planetesimal mass $m_{\max}$. Here, $f_7 \equiv (p-1)[1/(2-p) + 1/(P-2)]/5.32$. We choose $p = 1.68$ and $P = 4$ so $f_7 = 0.46$; a steeper high-mass slope $P = 6$ would change this only to $f_7 = 0.43$. Alternatively, we may write

$$\sigma_r \simeq \left[1.25 f_7 \left(\frac{f_1}{0.69}\right)^{-1} \frac{m_{\max}^{1/3}\rho_{\mathrm{p}}^{2/3}}{\Sigma\Omega t_0}\right]^{1/[2(p-1)]} \sqrt{Q_D^*\,(m_{\max})}. \tag{57}$$

By replacing the mass m in our discussion of monodisperse discs with $m_{\max}$, most of the results of Sections 2 and 3 can be applied to collision-limited discs without major errors. For example (i) the gravitational stability of hot discs depends on the surface density $\Sigma \propto \int m\,\mathrm{d}n$, which is dominated by masses near $m_{\max}$ when $p < 2 < P$. (ii) For most purposes the appropriate replacement for the collision rate $t_{\mathrm{c}}^{-1} \propto nr^2(f_1 + f_2\Theta)$ (equation 27) in a disc with a distribution of masses is the mass-weighted collision rate $\propto \int mr^2(f_1 + f_2\Theta)\,\mathrm{d}n$, which in turn is proportional to $\int m^{5/3}\,\mathrm{d}n$ for $\Theta \ll 1$ and $\int m^{7/3}\,\mathrm{d}n$ for $\Theta \gg 1$. These integrals are dominated by masses near $m_{\max}$ when $p < 8/3 < P$ and $p < 10/3 < P$, respectively. (iii) The gravitational scattering rate is $t_g^{-1} \propto \int r^2\Theta^2\,\mathrm{d}n \propto \int m^2\,\mathrm{d}n$ so this is dominated by masses near $m_{\max}$ when $p < 3 < P$. All of these inequalities are satisfied for our nominal values $p = 1.68$ and $P = 4$.

Finally, we note an interesting inconsistency in the results we have derived so far. Neglecting factors of order unity, the collision time $t_{\mathrm{c,m}}$ for a monodisperse disc with $\Theta \ll 1$ is given by equation (27) as $t_{\mathrm{c,m}}^{-1} \sim n_0\sigma_r m^{2/3}/\rho_{\mathrm{p}}^{2/3}$. The analogous collision time $t_{\mathrm{c,cl}}$ for a collision-limited disc is given by equation (54) as $t_{\mathrm{c,cl}}^{-1} \sim n_{\max}\sigma_r m_{\max}^{2/3}/\rho_{\mathrm{p}}^{2/3}\,(Q_D^*/\sigma_r^2)^{1-p} \sim t_{\mathrm{c,m}}^{-1}\,(Q_D^*/\sigma_r^2)^{1-p} \gtrsim t_{\mathrm{c,m}}^{-1}$. The difference arises because the collision-limited disc has a large population of bodies with $m_{\min} \lesssim m \lesssim m_{\max}$ that can collide with and disperse the large planetesimals, whereas in the monodisperse disc these are destroyed only by collisions among themselves. What then is the state of a disc in which $t_{\mathrm{c,cl}} \lesssim t_0 \lesssim t_{\mathrm{c,m}}$? Should it be regarded as monodisperse or collision-limited? We assume here that such discs are monodisperse but this assumption may be oversimplified.

5 NON-GRAVITATIONAL FORCES ON DUST

We also describe the most important non-gravitational forces on dust grains (Burns, Lamy & Soter 1979), since the distribution of dust grains determines the IR flux from debris discs; these forces can also be relevant for the planetesimals in warm discs. Gas drag on the dust is unimportant since we are focusing on discs older than a few Myr, at which point the gas in the protoplanetary disc has disappeared. The ratio of repulsive forces from the stellar wind and radiation pressure to the attractive gravitational force is

$$\beta = \frac{3}{16\pi}\frac{\mathrm{L}_\odot \mathcal{P}_r}{G\mathrm{M}_\odot c\rho_{\mathrm{p}} r} = 0.19\,\mathcal{P}_r\left(\frac{\rho_{\mathrm{p}}}{3\,\mathrm{g\,cm^{-3}}}\right)^{-1}\left(\frac{r}{1\,\mu\mathrm{m}}\right)^{-1}, \tag{58}$$

in which we have assumed that the host star has the solar mass $\mathrm{M}_\odot$ and luminosity $\mathrm{L}_\odot$, ρ_{p} is the dust grain density, r is the grain radius, and

$$\mathcal{P}_r = Q_{\mathrm{pr}} + \frac{\dot{M} v_{\mathrm{w}} c}{\mathrm{L}_\odot}, \tag{59}$$

where $\dot{M}$ is the rate of mass loss in the stellar wind, v_{w} is the wind speed, and Q_{pr} is the radiation pressure efficiency factor (averaged over the stellar spectrum) as defined by Burns et al. (1979). For stars with the Sun's luminosity and age, the contribution of the stellar wind to $\mathcal{P}_r$ is negligible ($\lesssim 10^{-3}$). Grains created by collisions on circular orbits with the Keplerian speed are unbound if $\beta \ge 1/2$ – because their eccentricity is $\beta/(1-\beta)$ (Burns et al.

[4] The condition $P \gtrsim 2$ appears to hold for most planetesimal populations in the Solar system. For the classical and excited Kuiper belts, $P \simeq 3.3$ and 2.1, respectively, and for trans-Neptunian objects $P \simeq 2.5$ (Bernstein et al. 2004). Numerical models of the formation of the Kuiper belt give $P = 2.3$ (Kenyon & Bromley 2004). Asteroid observations yield P between 1.9 and 2.3 (Parker et al. 2008) and Jupiter-family comets have $P = 1.9$ though with large uncertainties (Fernández 2005).

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

1979) – and thus $\beta = 1/2$ defines the 'blow-out radius',

$$r_b = 0.38\,\mu\mathrm{m}\,\mathcal{P}_r \left(\frac{\rho_\mathrm{p}}{3\,\mathrm{g\,cm^{-3}}}\right)^{-1}. \quad (60)$$

The rate of orbital decay from Poynting–Robertson and stellar-wind drag is (Burns et al. 1979)

$$\frac{1}{t_\mathrm{PR}} \equiv -\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t} = \frac{3}{8\pi}\frac{\mathrm{L}_\odot \mathcal{P}_\phi}{c^2 a^2 \rho_\mathrm{p} r} = \frac{2\mathcal{P}_\phi}{\mathcal{P}_r}\frac{G\mathrm{M}_\odot}{ca^2}\beta, \quad (61)$$

where

$$\mathcal{P}_\phi = Q_\mathrm{pr} + \frac{\dot{M}c^2}{\mathrm{L}_\odot}. \quad (62)$$

Numerically, we have

$$t_\mathrm{PR} = 8.0\times10^4\,\mathrm{yr}\,\beta^{-1}\frac{\mathcal{P}_r}{\mathcal{P}_\phi}\left(\frac{a}{10\,\mathrm{au}}\right)^2. \quad (63)$$

The contribution of the stellar wind to $\mathcal{P}_\phi$ is negligible for particles with $r \gtrsim 1\,\mu$m but grows as r^{-1} for $r \lesssim 0.3\,\mu$m, and equals the contribution due to Poynting-Robertson drag at $r \approx 0.1\,\mu$m (Burns et al. 1979).

The thermal emission from the dust is determined by its absorption efficiency factor Q_a, which is similar in magnitude to Q_pr and equal to it if scattering is neglected. In the geometric optics limit, Q_a is independent of wavelength and close to unity for typical dust grains; in the limit of long wavelength, where $X \equiv 2\pi r/\lambda \ll 1$, $Q_\mathrm{a} \sim X$ (equation 92). The contribution of a grain to the thermal emissivity at wavelength λ is proportional to $\pi r^2 Q_\mathrm{a}$. Thus for a power-law mass distribution with exponent $-p$ (equation 45) in the range $5/3 < p < 2$ (as is the case for a collisional cascade; see Section 4.2)[5] the total thermal emissivity is dominated by grains with $X = 2\pi r/\lambda \sim 1$ or $r \simeq r_\mathrm{IR}(\lambda) \equiv \lambda/(2\pi)$. More precisely, a fraction f of the emission comes from particles with radii that exceed r_IR/κ, where $f = 1 - (3p-5)\kappa^{3(p-2)}$. For $p = 11/6$, the expected value for a collisional cascade (we use $p = 11/6 = 1.83$ rather than $p = 1.89$ as at the end of Section 4.2 for reasons given after equation 94), 75 per cent of the emission comes from particles with radii that exceed $0.25r_\mathrm{IR}$. For observations at $\lambda = 20\,\mu$m, $r_\mathrm{IR} = 3\,\mu$m and 75 per cent of the emission comes from particles with radii that exceed $0.8\,\mu$m, where $\beta = 0.24$ (assuming $\rho_\mathrm{p} = 3\,\mathrm{g\,cm^{-3}}$ and $Q_\mathrm{a} = 1$, since most of the stellar emission is at shorter wavelengths). At longer observational wavelengths β is even smaller for the particles dominating the emission. Thus radiation pressure is negligible, except perhaps for accurate modelling at the shortest observational wavelengths.

Poynting–Robertson drag is also negligible, at least for detectable debris discs, as shown by the following argument. Assume for simplicity that the dust particles have a single size. Using equations (12) and (27) the collision time t_c for dust is related to the geometrical optical depth as seen from the host star τ_p by

$$\Omega t_\mathrm{c} = \frac{\pi f_m}{64 f_1 \tau_\mathrm{p}} \quad \text{or} \quad t_\mathrm{c} = 3.6\times10^4\,\mathrm{yr}\,f_m\left(\frac{f_1}{0.69}\right)^{-1} \times \left(\frac{\tau_\mathrm{p}}{10^{-5}}\right)^{-1}\left(\frac{a}{10\,\mathrm{au}}\right)^{3/2}. \quad (64)$$

[5] The assumption of a power-law mass distribution neglects the oscillations that appear in the mass distribution of a collisional cascade at radii that are not too far from the blow-out radius (Krivov, Löhne & Sremčević 2006); these oscillations can change the number density at a given radius or mass by a factor of ∼3.

For particles with absorption efficiency $Q_\mathrm{a} \simeq 1$, the optical depth τ_p is equal to the bolometric luminosity of the disc relative to the star, which exceeds 10^{-5} in almost all observed debris discs (Wyatt 2008). Therefore, the collision time $t_\mathrm{c} \lesssim 4\times10^4\,\mathrm{yr}\,(a/10\,\mathrm{au})^{3/2}$ in observed discs, shorter than the Poynting–Robertson drag time (63), so the grains are destroyed by collisions before they experience significant orbital decay.

Based on these arguments, we neglect non-gravitational forces on the dust distribution when calculating its emission properties.

6 THE PROPERTIES OF LONG-LIVED DISCS

The primary goal of this paper is to explore the properties of planetesimal discs that can survive for most of the age of the Galaxy. Even in our highly simplified model, the local properties of discs are specified by four parameters: semimajor axis a, surface density Σ, planetesimal radius r or mass m, and rms eccentricity e_0, while collision-limited discs are specified by three parameters. It is challenging to visualize the properties of a four-dimensional parameter space. As a first step, we outline some general conclusions that arise from the discussion of the previous section. We later examine six sample discs in Figs 3–5.

6.1 Cold discs

6.1.1 The minimum planetesimal mass in cold discs

Eliminating the separation Δa between equations (13) and (18) yields the minimum mass in cold discs,

$$m > m_\mathrm{min}^\mathrm{cold} \equiv 100\,\mathrm{M}_\odot\mu^{1.43} \simeq 60 M_\oplus\left(\frac{\mu}{10^{-4}}\right)^{1.43}. \quad (65)$$

Thus, the minimum planetesimal mass in a cold disc with $\mu = 10^{-4}$ is 0.2 Jupiter masses, while if $\mu = 10^{-6}$ the minimum mass is 0.1 Earth masses.

6.1.2 The maximum number of planetesimals in cold discs

If we write the disc mass as $M_\mathrm{disc} = \pi f_m \Sigma a^2$ (with $f_m = 1$ this is roughly the mass per octave in semimajor axis; a disc extending over multiple octaves in semimajor axis could have $f_m \gg 1$), then with equation (65) we have

$$N = \frac{M_\mathrm{disc}}{m} < N_\mathrm{max}^\mathrm{cold} \equiv \frac{f_m}{8}\left(\frac{\mathrm{M}_\odot}{m}\right)^{0.3} \quad (66)$$

or $N \sim 1$ for Jupiter-mass planets, 6 for Earth-mass planets and 20 for lunar-mass planets.

This result can be re-cast in terms of the surface density,

$$N < N_\mathrm{max}^\mathrm{cold} \equiv 1.6 f_m\left(\frac{\mu}{10^{-4}}\right)^{-0.43}. \quad (67)$$

Thus cold discs with $\mu = 10^{-4}$ can host no more than one or two equal-mass planetesimals per octave in radius, while cold discs with $\mu = 10^{-6}$ can host up to 12.

6.2 Hot discs

6.2.1 The maximum surface density for hot discs

Comparing Figs 1 and 2 shows that a necessary condition for the survival of a hot disc is that the allowed regions in the two figures overlap, and that this requires (i) $\tau^3/\nu^2 \lesssim \mu$ (the minimum planetesimal mass for which the collision time is longer than the age must

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

be smaller than the disc mass) and (ii) $\tau/\nu^{1/2} \lesssim 1$ (the minimum eccentricity for which the collision and gravitational scattering times are less than the age must be less than unity). The first of these can be written more accurately using equations (10) and (27) as

$$\mu < \mu_{\rm max,c}, \tag{68}$$

where

$$\mu_{\rm max,c} = \frac{\pi^{3/2} f_m^{1/2}}{48 f_1^{3/2}} \frac{\nu}{(\Omega t_0)^{3/2}}, \\ = 7.0 \times 10^{-5} f_m^{1/2} \left(\frac{\rho_{\rm p}}{3\,{\rm g\,cm^{-3}}}\right) \left(\frac{t_0}{3\,{\rm Gyr}}\right)^{-3/2} \left(\frac{a}{10\,{\rm au}}\right)^{21/4}. \tag{69}$$

To describe condition (ii) more accurately, we re-write the relaxation time (37) using equation (26):

$$t_{\rm relax}^{-1} = \frac{16\Sigma\Omega}{e_0} \left(\frac{3a}{4\pi\rho_{\rm p} {\rm M}_\odot}\right)^{1/2} (f_1 \Theta^{-1/2} + f_2 \Theta^{1/2} + f_3 \Theta^{3/2} \log\Lambda), \tag{70}$$

in which we have assumed $\Lambda \gg 1$, as is usually the case, and neglected the contribution of shear-dominated encounters. The minimum of the expression in brackets occurs when $\Theta = f_1^{1/2}/(3 f_3 \log\Lambda)^{1/2}$ for $\log\Lambda \gg 1$, and equals $0.97(\log\Lambda)^{1/4}$. Since $e_0 < f_e$ (equation 8), the relaxation time cannot be greater than the age t_0 for any rms eccentricity unless

$$\mu < \mu_{\rm max,relax} \tag{71}$$

where

$$\mu_{\rm max,relax} = 4.4 \times 10^{-6} \left(\frac{f_e}{0.5}\right) \left(\frac{f_1}{0.69}\right)^{-3/4} \left(\frac{f_3}{0.28}\right)^{-1/4} \\ \times \left(\frac{\log\Lambda}{10}\right)^{-1/4} \left(\frac{\rho_{\rm p}}{3\,{\rm g\,cm^{-3}}}\right)^{1/2} \left(\frac{t_0}{3\,{\rm Gyr}}\right)^{-1} \left(\frac{a}{10\,{\rm au}}\right)^{3}. \tag{72}$$

The existence of a maximum surface density implies a maximum value for the IR excess emission due to dust (see Section 8.2).

6.2.2 *The maximum number of planetesimals in hot discs*

When a hot disc satisfies the constraints (68) and (71), for a given surface density and semimajor axis there is a minimum planetesimal mass and maximum number of planetesimals, given approximately by (cf. Fig. 2)

$$\frac{m_{\rm min}^{\rm hot}}{{\rm M}_\odot} \sim \frac{\tau^3}{\nu^2} = \frac{(\Sigma\Omega t_0)^3}{{\rm M}_\odot \rho_{\rm p}^2}, \quad N_{\rm max}^{\rm hot} \sim \left(\frac{\rho_{\rm p} a}{\Sigma}\right)^2 \frac{1}{(\Omega t_0)^3}. \tag{73}$$

More precisely, by evaluating $t_{\rm c} = t_0$ ($\Theta \ll 1$) we get:

$$m_{\rm min}^{\rm hot} = \frac{2304 f_1^3}{\pi^2} \frac{(\Sigma\Omega t_0)^3}{\rho_{\rm p}^2} \\ = 210 M_\oplus \left(\frac{\mu}{10^{-4}}\right)^3 \left(\frac{f_1}{0.69}\right)^3 \left(\frac{\rho_{\rm p}}{3\,{\rm g\,cm^{-3}}}\right)^{-2} \\ \times \left(\frac{t_0}{3\,{\rm Gyr}}\right)^3 \left(\frac{a}{10\,{\rm au}}\right)^{-21/2}, \\ N_{\rm max}^{\rm hot} = \frac{\pi^3 f_m}{2304 f_1^3} \left(\frac{\rho_{\rm p} a}{\Sigma}\right)^2 \frac{1}{(\Omega t_0)^3} \\ = 0.49 f_m \left(\frac{f_1}{0.69}\right)^{-3} \left(\frac{\mu}{10^{-4}}\right)^{-2} \left(\frac{\rho_{\rm p}}{3\,{\rm g\,cm^{-3}}}\right)^2 \\ \times \left(\frac{t_0}{3\,{\rm Gyr}}\right)^{-3} \left(\frac{a}{10\,{\rm au}}\right)^{21/2}. \tag{74}$$

6.3 Warm discs

Similarly, for given values of the surface density and semimajor axes there is a maximum mass ($m_{\rm max}^{\rm warm}$) – and thereby a minimum number ($N_{\rm min}^{\rm warm}$) – of planetesimals in warm discs. Generally, the maximum mass and minimum number of planetesimals need to be evaluated numerically, but for the cases considered below (Section 6.4), these extremes are attained when the line $t_{\rm c} = t_0$ intersects the line $\Theta = 1$ (cf. Figs 4 and 5). At this point,

$$m_{\rm max}^{\rm warm} = 0.007 M_\oplus \left(\frac{f_1 + f_2}{2.211}\right)^3 \left(\frac{\mu}{10^{-6}}\right)^3 \left(\frac{\rho_{\rm p}}{3\,{\rm g\,cm^{-3}}}\right)^{-2} \\ \times \left(\frac{t_0}{3\,{\rm Gyr}}\right)^3 \left(\frac{a}{10\,{\rm au}}\right)^{-21/2}, \\ N_{\rm min}^{\rm warm} = 150 f_m \left(\frac{f_1 + f_2}{2.211}\right)^{-3} \left(\frac{\mu}{10^{-6}}\right)^{-2} \left(\frac{\rho_{\rm p}}{3\,{\rm g\,cm^{-3}}}\right)^2 \\ \times \left(\frac{t_0}{3\,{\rm Gyr}}\right)^{-3} \left(\frac{a}{10\,{\rm au}}\right)^{21/2}. \tag{75}$$

6.4 Sample discs

As described at the end of Section 2, we examine six possible planetesimal discs (Table 1), with semimajor axes $a = 1$, 10 and 100 au, and dimensionless masses $\mu = 10^{-4}$ and 10^{-6}. The allowed values of velocity dispersion σ_r and planetesimal radius r are shown in Figs 3–5. In all cases, we assume that the disc age is $t_0 = 3$ Gyr and the planetesimal density is $\rho_{\rm p} = 3\,{\rm g\,cm^{-3}}$.

(A) $a = 1$ au, $\mu = 10^{-4}$: hot and warm discs cannot survive (see equations 40, 68, 71 and 75). Cold discs *can* survive, but only for a narrow range of planetesimal masses (the thin blue trapezoid in the upper panel of Fig. 3): there can be at most one or two planetesimals per octave of semimajor axis, of mass $m_{\rm min}^{\rm cold} \simeq 60 M_\oplus \simeq 0.2 M_{\rm Jupiter}$ (equation 65). Such discs are rather similar to some of the many extrasolar planetary systems already detected by radial-velocity variations in the host star.

(B) $a = 1$ au, $\mu = 10^{-6}$: hot and warm discs cannot survive. Cold discs can have planetesimal masses in the range 0.1–1$M_\oplus$ (equation 65); the lower limit corresponds to about a dozen planetesimals per octave of semimajor axis (equation 67). These discs may be detectable with space-based transit surveys and are reminiscent of the terrestrial planets in our own Solar system.

(C) $a = 10$ au, $\mu = 10^{-4}$: hot and warm discs cannot survive. As in the case of disc A, only one or two planetesimals (or planets) per octave of mass $m_{\rm min}^{\rm cold} \simeq 60 M_\oplus$ can survive in a cold disc. We suggest in Section 7.5.3 that some discs of this type may be detectable by gravitational microlensing.

(D) $a = 10$ au, $\mu = 10^{-6}$: hot, warm, and cold discs can all survive. The hot discs may contain up to about 5000 planetesimals per octave with minimum masses of $m_{\rm min}^{\rm hot} \approx 10^{24}$ g, about the mass of Ceres. The dynamical constraints on cold discs are the same as for disc B; such discs are not detectable with current or planned transit surveys because the probability of transits is too small and the orbital period is too large but could be detected by targeted searches for gravitational microlensing. A wide range of warm discs is possible, with at least 200 planetesimals per octave and masses at most 0.4 times that of the Moon. A more typical warm disc might have 5×10^5 planetesimals per octave, of radius 100 km, with velocity dispersion $\sigma_r \simeq 0.1\,{\rm km\,s^{-1}}$ and a collision time of 0.6 Gyr.

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

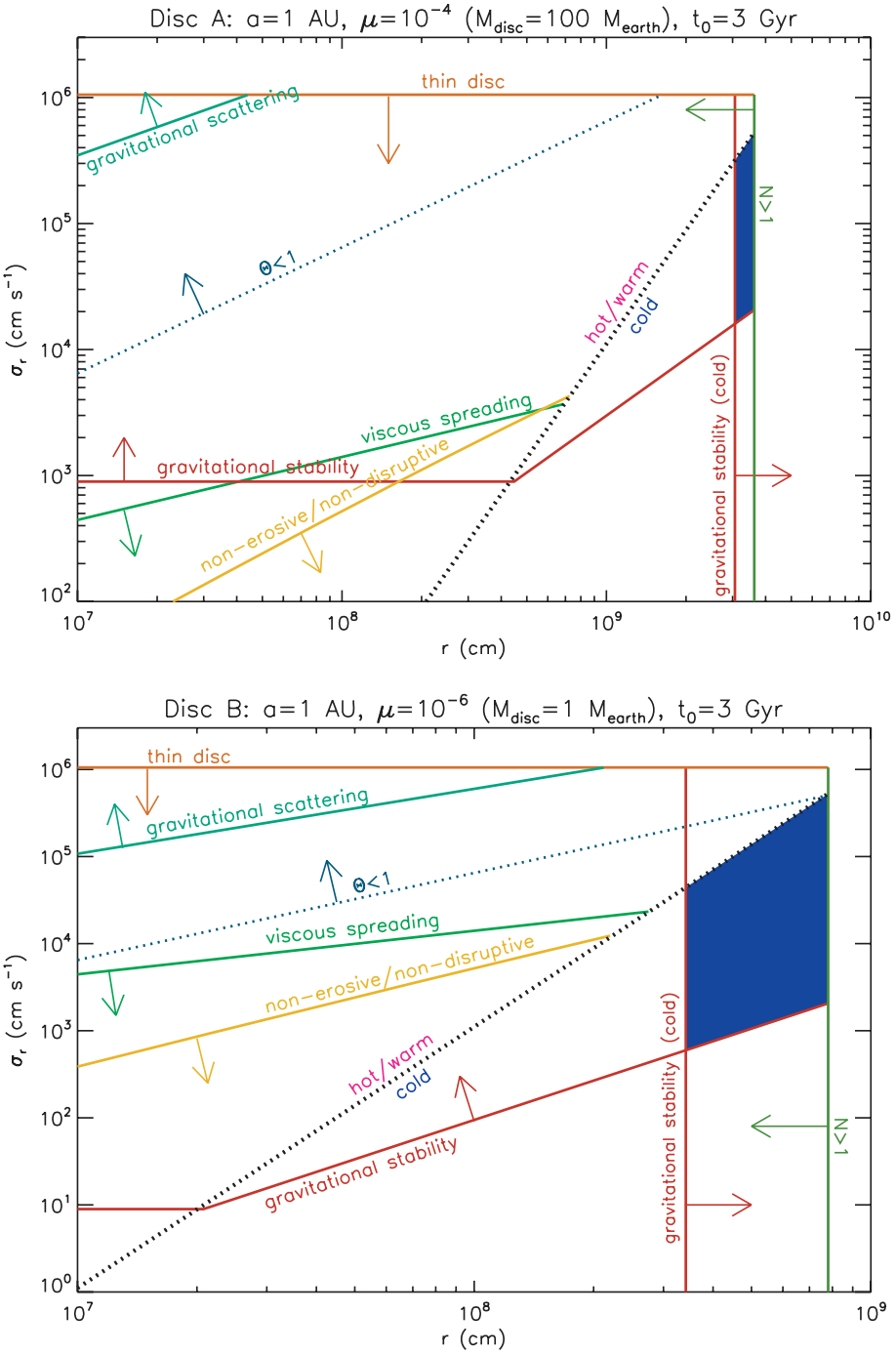

Figure 3. Allowed values of planetesimal velocity dispersion and radius for discs with age $t_0 = 3\,\mathrm{Gyr}$, semimajor axis $a = 1\,\mathrm{au}$, and dimensionless mass $\mu = 10^{-4}$ ($M_{\rm disc} \simeq 100\,M_\oplus$) and $\mu = 10^{-6}$ ($M_{\rm disc} \simeq M_\oplus$), i.e. discs A and B of Table 1. The blue-shaded region denotes allowed cold discs. There are no allowed hot or warm discs. The various lines represent conditions for: gravitational stability (equations 18 and 23), gravitational scattering time exceeds disc age (equation 31), thin disc (equation 8), $N > 1$ (equation 10), collisions are not erosive/disruptive (equations 40 and 44), and viscous spreading time exceeds disc age (equation 43). The arrows attached to each line indicate the region in which long-lived discs could exist. The dividing line between hot/warm and cold discs is given by equation (14) and the dividing line between collision speeds greater than or less than the escape speed from the planetesimal surface ($\Theta < 1$ or $\Theta > 1$ respectively) is given by equation (26).

(E) $a = 100\,\mathrm{au}$, $\mu = 10^{-4}$: hot, warm, and cold discs can all survive. As in the case of discs A and C, the cold discs can have only one or two planetesimals per octave. Hot discs can have up to 2×10^{10} objects per octave with masses of at least 4×10^{19} g. Warm discs contain at least 5×10^{8} objects per octave with maximum masses of 10^{21} g ($r \simeq 40$ km; approaching the largest sizes of comet nuclei).

(F) $a = 100\,\mathrm{au}$, $\mu = 10^{-6}$: hot discs have at most 2×10^{14} objects per octave with $m_{\rm min}^{\rm hot} \simeq 4 \times 10^{13}$ g. Warm discs contain objects with $m \lesssim 10^{15}$ g ($r \simeq 400$ m), of which there are at least

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

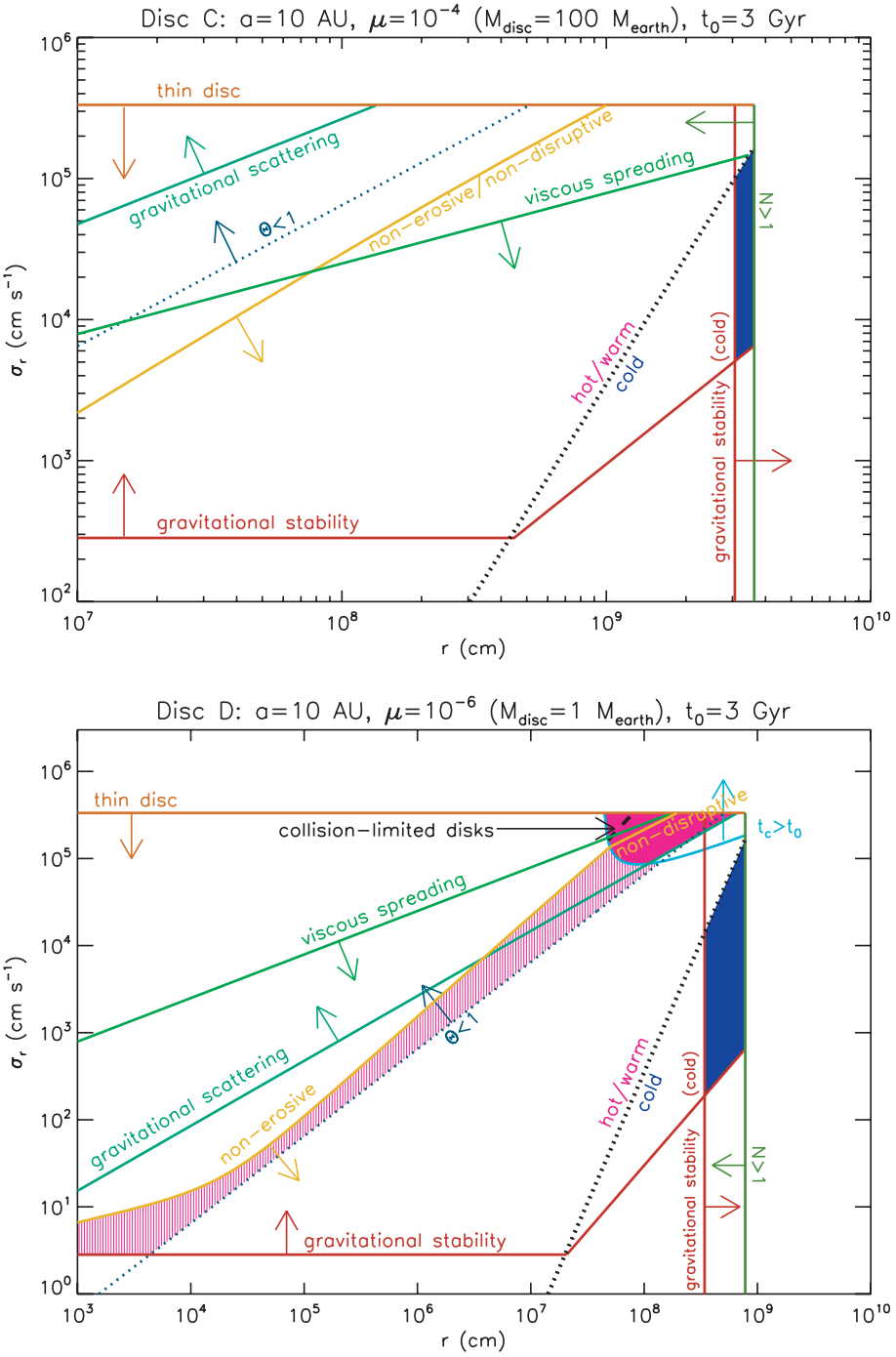

Figure 4. Same as Fig. 3, but for $a = 10$ au, i.e. discs C and D. For disc D, the pink shaded regions denote allowed hot discs (solid colour) and warm discs (vertical hatching). An additional constraint for disc D is $t_c \gtrsim t_0$ (equations 27 and 30). The dashed curve represents collision-limited discs (equation 57).

5×10^{12} per octave. The constraints on cold discs are the same as for discs B and D.

It is worthwhile to emphasize that Figs 3–5 do not show how a planetesimal disc of a given initial mass will evolve, but rather whether a disc with a current mass $M_{\rm disc}$ can survive in approximately its current state for 3 Gyr; the 'allowed' regions in these figures can be interpreted as the allowed regions for discs with an age $t_0 = 3$ Gyr on the assumption that it is unlikely to find objects in states that evolve on a time-scale much less than their age.

6.5 Collision-limited discs

Collision-limited discs are shown by dashed curves in Figs 3 and 4 for discs D, E and F (equation 57). The allowed ranges of masses are 10^{24} g $\lesssim m_{\rm max} \lesssim 8 \times 10^{24}$ g (disc D), 4×10^{19} g $\lesssim m_{\rm max} \lesssim 3 \times 10^{21}$ g (disc E) and 4×10^{13} g $\lesssim m_{\rm max} \lesssim 4 \times 10^{18}$ g (disc F).

© 2009 The Authors. Journal compilation © 2009 RAS, MNRAS **401,** 867–889

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

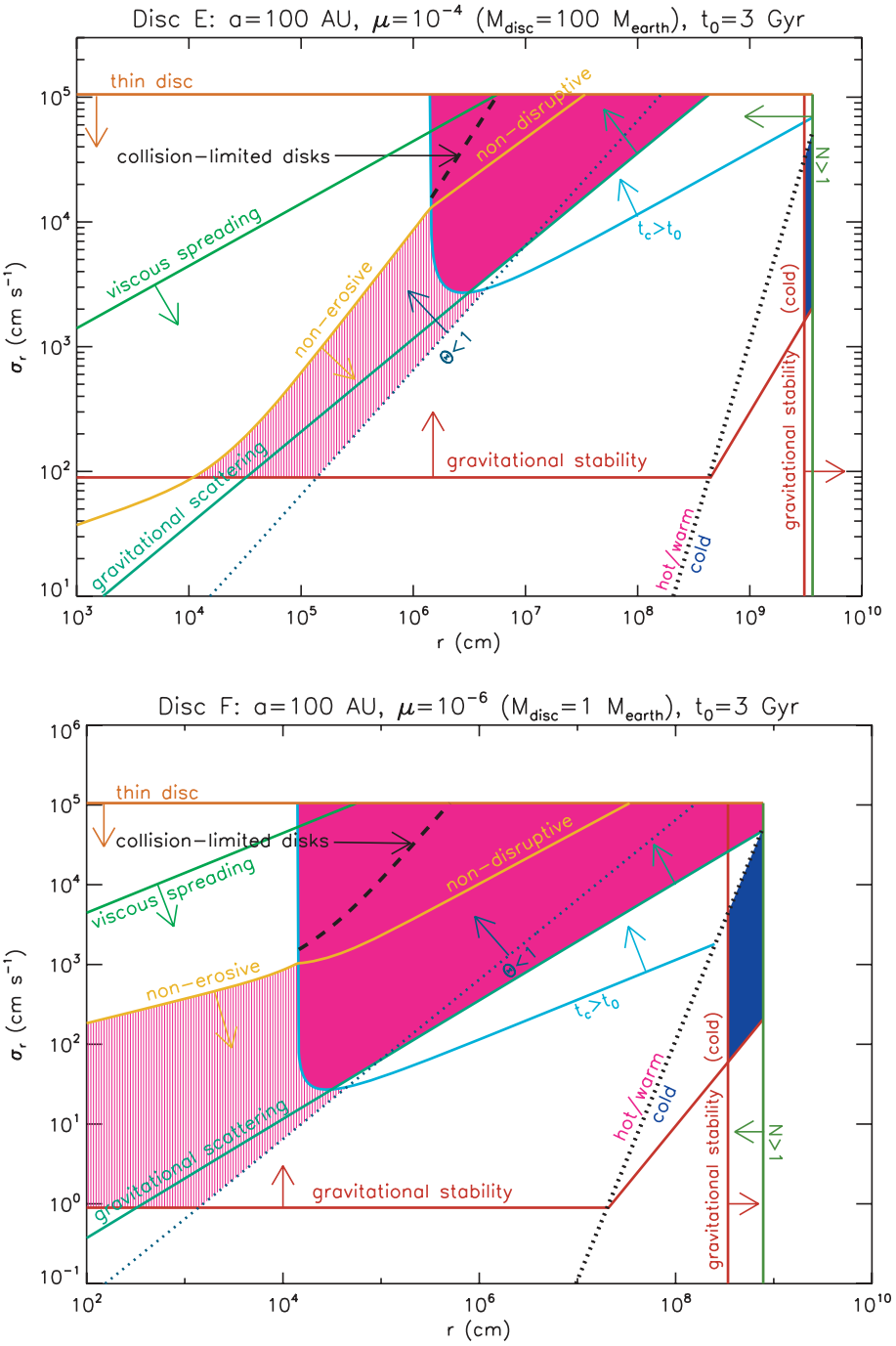

Figure 5. Same as Figs 3 and 4, but for $a = 100$ au, i.e. discs E and F.

7 DETECTION TECHNIQUES

7.1 Radial velocity measurements

The stellar wobble or reflex radial velocity induced by an edge-on disc containing N planetesimals of mass m is approximately

$$\begin{aligned} v_{\rm wobble} &\simeq \sqrt{\frac{GM_\odot N}{a}}\frac{m}{M_\odot}, \\ &= \pi f_m \mu \sqrt{\frac{GM_\odot}{aN}}, \\ &= 3.0\,{\rm m\,s^{-1}}\, f_m N^{-1/2}\left(\frac{a}{10\,{\rm au}}\right)^{-1/2}\left(\frac{\mu}{10^{-4}}\right). \end{aligned} \tag{76}$$

With current technology we can detect reflex velocities as small as $v_{\rm wobble} \sim 1\,{\rm m\,s^{-1}}$ with orbital periods as long as $\sim$10 yr, corresponding to $a \simeq 4.6$ au. For hot discs with ages of several Gyr, equations (71) and (76) imply $v_{\rm wobble} \lesssim 0.1\,{\rm m\,s^{-1}}\,(a/10\,{\rm au})^{5/2}$, too small to be detectable.

Cold discs are detectable if the semimajor axis is small: combining equation (67) with the second of the equations above, we have

$$v_{\rm wobble} > 2.3\,{\rm m\,s^{-1}}\, f_m^{1/2}\left(\frac{a}{10\,{\rm au}}\right)^{-1/2}\left(\frac{\mu}{10^{-4}}\right)^{1.22}. \tag{77}$$

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

Thus, *all* long-lived cold discs with $\mu \gtrsim 10^{-4}$ and $a \lesssim 10\,\mathrm{au}$ can be detected by current radial-velocity surveys; this of course is because gravitational stability requires that they have only a few large planets.

7.2 Transits

Space-based transit surveys such as NASA's *Kepler* mission are capable of detecting photometric variations as small as $\sim 10^{-5}$, corresponding to the transit of a planetesimal of radius $2000\,\mathrm{km} = 0.35 R_\oplus$. Thus, edge-on discs containing lunar-mass planetesimals may be detectable by transit surveys. Reliable detection of transits of such small objects requires that the stellar variability is negligible; this is likely to be true for at least some stars as the solar variability on the hourly time-scales relevant to transit detection is only a few times 10^{-5} (Batalha et al. 2002).

Other criteria for detectability of edge-on planetesimal discs by transits include the following. (i) The orbital period must be less than a year or so, so that several transits of a given object can be detected in a mission of reasonable duration. (ii) There must not be too few transits, that is, at least one planetesimal in the disc must transit the star. If the disc is nearly edge-on and the characteristic thickness h (equation 4) is small compared to the stellar radius $R_\star$ then most planetesimals transit the stellar disc in the course of an orbit, while if $h \gg R_\star$ the fraction of transiting planetesimals is $(2/\pi)^{1/2}\, R_\star/h$, so the expected number of transiting planetesimals is roughly

$$N_t = N \min\left[1, \left(\frac{2}{\pi}\right)^{1/2} \frac{R_\star}{h}\right]. \tag{78}$$

(iii) There must not be *too many* transits: if multiple planetesimals are transiting the disc at any one time, the fluctuations in stellar flux will be difficult to distinguish from normal stellar variability. The average number of planetesimals in transit at a given time is $N R_\star/(\pi a)$ if $h \ll R_\star$ and $N R_\star^2/(\sqrt{8\pi} h a)$ if $h \gg R_\star$ so the average number of planetesimals in transit at any instant is

$$n_t = N \min\left(\frac{R_\star}{\pi a}, \frac{R_\star^2}{\sqrt{8\pi} h a}\right). \tag{79}$$

7.3 Microlensing

7.3.1 Microlensing by individual planetesimals

The classical lensing equation is (e.g. Schneider, Ehlers & Falco 1992)

$$\alpha d_\star = \frac{d_\star L}{d} - \frac{2 R_\mathrm{s} d_{\mathrm{l}\star}}{L}, \tag{80}$$

where $R_\mathrm{s} = 2Gm/c^2$ is the Schwarzschild radius of the lens and L is the projected separation between the light ray and the lens in the lens plane. The distance to the source and lens and the separation between them are denoted by $d_\star$, d and $d_{\mathrm{l}\star} = d_\star - d$, respectively. The quantity α is the subtended angle between the lines of sight to the lens and the source. The Einstein radius is defined by the value of L when $\alpha = 0$ (i.e. lens and source are aligned),

$$r_\mathrm{E} = \frac{2}{c}\sqrt{G m d_\star \zeta\,(1-\zeta)} \tag{81}$$

where $\zeta \equiv d/d_\star$.

A planetesimal can be far from or near to its parent star, where 'far' and 'near' are defined with respect to the stellar Einstein radius,

$$R_\mathrm{E} = 4.0\,\mathrm{au}\left[\frac{M_\star}{\mathrm{M}_\odot}\frac{d_\star}{8\,\mathrm{kpc}}\frac{\zeta(1-\zeta)}{0.25}\right]^{1/2}, \tag{82}$$

with $M_\star$ being the stellar mass.

The magnification of the total flux from the source is

$$\mathcal{A} = \frac{u^2+2}{u\sqrt{u^2+4}}, \tag{83}$$

where $u \equiv \alpha d/r_\mathrm{E}$. In our simple treatment, we assume that a planetesimal can produce a detectable lensing event of non-negligible magnification when the distance αd between the lines of sight to the planetesimal and the source star, measured in the lens plane, is less than the Einstein radius; this corresponds to $u = 1$ or amplification $\mathcal{A} = 3/\sqrt{5} \simeq 1.34$.

We must check that the planetesimal radius is small compared to the Einstein radius to ensure that the magnified light curve is not blocked (Agol 2002). We have

$$\begin{aligned}\frac{r}{r_\mathrm{E}} &= \left[\frac{3c^2}{16\pi G \rho_\mathrm{p} r d_\star \zeta(1-\zeta)}\right]^{1/2}, \\ &= 0.66\left[\frac{r}{1\,\mathrm{km}}\frac{\rho_\mathrm{p}}{3\,\mathrm{g\,cm^{-3}}}\frac{d_\star}{8\,\mathrm{kpc}}\frac{\zeta\,(1-\zeta)}{0.25}\right]^{-1/2},\end{aligned} \tag{84}$$

so obscuration by the planetesimals is unimportant if they are much larger than 1 km.

In addition, we require that the stellar (source) radius as projected on the lens plane – equal to the stellar radius $R_\star$ multiplied by ζ – cannot be much larger than the Einstein radius, to ensure that the magnified light curve is not smeared out. We have

$$\begin{aligned}\frac{R_\star \zeta}{r_\mathrm{E}} &= \left(\frac{3c^2 R_\star^2}{16\pi G \rho_\mathrm{p} r^3 d_\star}\frac{\zeta}{1-\zeta}\right)^{1/2}, \\ &= 7.3\left(\frac{R_\star}{R_\odot}\right)\left(\frac{r}{1000\,\mathrm{km}}\right)^{-3/2}\left(\frac{\rho_\mathrm{p}}{3\,\mathrm{g\,cm^{-3}}}\frac{d_\star}{8\,\mathrm{kpc}}\frac{1-\zeta}{\zeta}\right)^{-1/2}.\end{aligned} \tag{85}$$

We have parametrized $R_\star$ in terms of the solar radius since this is the typical size of the source star in existing planetary microlensing events (at 8 kpc the corresponding angular size is $0.6\,\mu$as). With the nominal parameters and $\zeta = 0.5$, $R_\star\,\zeta/r_\mathrm{E} < 1$ only for planetesimal radii $r > 3750\,\mathrm{km}$, corresponding to mass $m > 0.11 M_\oplus$ (Paczyński 1996 gives a similar estimate, $0.07 M_\oplus$). This limit is conservative because the magnification of extended sources remains substantial when the projected source radius is as large as several times the Einstein radius – for example, a uniform source whose centre is separated from the lens by one Einstein radius is magnified *more* than a point source so long as $R_\star\,\zeta/r_\mathrm{E} < 2.17$ (Gould 1994; Witt & Mao 1994). Thus, microlensing searches are likely to be sensitive to planetesimals as small as $\sim 10^{-1.5} M_\oplus$ or a few times the mass of the Moon (but see Heng & Keeton 2009).[6]

[6] Events associated with lower amplifications have larger microlensing cross-sections, that is $\pi(\phi r_\mathrm{E})^2$ where $\phi > 1$. Planetesimals with Einstein radii smaller than the projected size of the source star may contribute appreciably to the expected number of events per planetesimal disc crossing because the range of masses involved in microlensing now extends down to much lower values.

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

If the transverse velocity of the lens relative to the source is $v_\perp$, the characteristic duration of the event is

$$t_{\rm E,d} \sim \frac{2r_{\rm E}}{v_\perp} = 16\,{\rm min}\left(\frac{v_\perp}{100\,{\rm km\,s^{-1}}}\right)^{-1}\left(\frac{r}{1000\,{\rm km}}\right)^{3/2} \times\left[\frac{\rho_{\rm p}}{3\,{\rm g\,cm^{-3}}}\frac{d_\star}{8\,{\rm kpc}}\frac{\zeta(1-\zeta)}{0.25}\right]^{1/2}. \quad (86)$$

In the most cases, the transverse velocity is dominated by the apparent angular speed of the source star relative to the host star of the planetesimal, rather than the motion of the planetesimal around its host star.

If we assume that the surface density of the disc is uniform over a circle of radius a, the probability of lensing at any given moment for a star whose image lies within the disc (i.e. the optical depth) is

$$\tau_{\rm lens} \simeq N\left(\frac{r_{\rm E}}{a}\right)^2, = 5\times10^{-5} f_m\left(\frac{\mu}{10^{-4}}\frac{d_\star}{8\,{\rm kpc}}\right)\left(\frac{a}{10\,{\rm au}}\right)^{-2}\frac{\zeta(1-\zeta)}{0.25}. \quad (87)$$

The optical depth of a planetesimal disc of a given size a and mass $M_{\rm disc} = \pi f_m \Sigma a^2$ is independent of the mass of the individual planetesimals. Thus (for example) the optical depth for a disc composed of 100 Earth-mass planets ($M_{\rm disc}/{\rm M}_\odot = 3\times10^{-4}$, $\mu = 10^{-4}$) is the same as the optical depth of a single 0.3 Jupiter-mass planet at the same radius.

A related quantity is the probability that at least one lensing event by a planetesimal will be seen at some time during the passage of the source star near the host star of the planetesimal disc. If the impact parameter associated with this passage is small compared to the size of the planetesimal disc, this probability is $1-\exp(-\tilde{N}_{\rm lens})$, where

$$\tilde{N}_{\rm lens} = \frac{4Nr_{\rm E}}{\pi a}, = \frac{8}{c}\left[\frac{f_m}{\pi}G\Sigma d_\star N\zeta(1-\zeta)\right]^{1/2} = 9.1\times10^{-3}\left(\frac{10\,{\rm au}}{a}\right)\left[f_m N\frac{d_\star}{8\,{\rm kpc}}\frac{\mu}{10^{-4}}\frac{\zeta(1-\zeta)}{0.25}\right]^{1/2}. \quad (88)$$

Note that the probability of observing an event goes up as the planetesimal mass goes down, since $N\sim\Sigma a^2/m$ (although the duration of the event is shorter). In this respect, a disc consisting of many small planetesimals may actually be *easier* to detect than a single large planet.

The short duration (86) of the events is one of the principal challenges in reliably observing microlensing by planetesimals of an Earth mass or less. To avoid being swamped by noise, it is useful to focus on source stars that are experiencing – or have recently experienced – microlensing by an intervening star. Strong amplification by the host star of a planetesimal disc requires that the impact parameter is less than the Einstein radius of the host star, given by equation (82). Since many planetesimal discs may be substantially larger than $R_{\rm E}$ the source star should be monitored for short-duration events for some time after the amplification by the host star has returned to unity.

7.3.2 Other microlensing effects

Planetesimal discs can produce other signals in microlensing surveys. Zheng & Ménard (2005) point out that if the source star for a microlensing event hosts a debris disc, the mid-/far-IR light curve will contain a component determined by the surface-brightness profile of the thermal emission from the debris disc. Similarly, the optical and near-IR light curve will contain a component from the scattered light from the disc. In both cases we may expect that the light curve is no longer wavelength independent. These effects are challenging to detect since (i) many debris discs are much larger than the stellar Einstein radius $R_{\rm E}\simeq4\,{\rm au}$ (equation 82), so the maximum magnification is only $\sim(R_{\rm E}/a)^2$, (ii) accurate mid-/far-IR photometry is exceedingly difficult, except from space and (iii) the fractional flux of scattered light is small, typically 10^{-3}–10^{-5} in observed debris discs (e.g. Wyatt 2008).

Other signals may arise if the lens star hosts a planetesimal disc. The overall mass distribution in the disc will contribute to the magnification and thereby distort the microlensing light curve, but this distortion will be difficult to detect because the magnification due to the disc will only be of order $\mu R_{\rm E}/a$ where μ is defined in equation (7). We have assumed that the disc is not far from face-on and that its semimajor axis a is larger than the Einstein radius of the host star (see Hundertmark, Hessman & Dreizler 2009 for a discussion of lensing by edge-on discs). A potentially more sensitive probe is high-magnification events. The gravitational field from distant stellar companions or other external mass distributions can produce a characteristic double-peak structure near the point of maximum magnification; for example, Kim et al. (2008) estimate that typical high-magnification events can detect stellar companions with mass m and separation Δd such that $m/{\rm M}_\odot \gtrsim (\Delta d/100\,{\rm au})^2$. Unfortunately, these events are insensitive to distant discs in most cases, for the following reason: the deflection angle from a surface mass density distribution $\Sigma_{\rm proj}(\mathbf{x})$ on the sky plane is (Schneider et al. 1992)

$$\alpha(\mathbf{x}) = \int\frac{4G\Sigma_{\rm proj}(\mathbf{x}')}{c^2}\frac{\mathbf{x}-\mathbf{x}'}{|\mathbf{x}-\mathbf{x}'|^2}{\rm d}^2\mathbf{x}'. \quad (89)$$

which is proportional to the gravitational field from a cylindrical mass distribution with density $\rho(x_1, x_2, x_3)\propto\Sigma_{\rm proj}(x_1, x_2)$. The deflection angle for light rays passing inside an inclined ring of material is therefore proportional to the gravitational field inside an elliptical, cylindrical shell, which is zero from Newton's theorem. Only discs in which $\Sigma_{\rm proj}$ is non-zero near the host star (e.g. nearly edge-on discs with a significant thickness) will affect the light curve near the peak magnification.

7.4 Infrared emission due to dust generated from collisions

Most extrasolar detections of planetesimal discs are based on measurements of IR excesses, i.e. the presence of a debris disc (Wyatt 2008) in which a steady supply of dust is generated by planetesimal collisions; the dust is heated by the host star and the heated dust generates IR emission. At a given wavelength, the IR excess, $f_{\rm IR}$, is the dust luminosity measured relative to the stellar luminosity.[7] In this section we estimate the IR excess due to planetesimal collisions in discs. The discs considered here differ from the collision-limited discs of Section 4 in that the collision time exceeds the age of the discs, $t_{\rm c}\gtrsim t_0$.

[7] Confusingly, f is also used by many authors to denote the bolometric luminosity of the disc relative to the star. For particles with absorption efficiency $Q_{\rm a}=1$ (equation 92), this ratio is equal to the geometrical optical depth $\tau_{\rm p}$ defined in equation (12).

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

If the dust grains are treated as grey bodies, their equilibrium temperature is

$$T_{\rm dust} = \left(\frac{\mathcal{L}_\star}{16\pi a^2 \sigma_{\rm SB}}\right)^{1/4} = 279\,{\rm K}\left(\frac{a}{\rm au}\right)^{-1/2}\left(\frac{\mathcal{L}_\star}{\mathcal{L}_\odot}\right)^{1/4}, \quad (90)$$

where $\sigma_{\rm SB}$ is the Stefan–Boltzmann constant. The peak wavelength of the blackbody spectrum $\lambda B_\lambda(\lambda, T)$ at this temperature is $\lambda_{\rm max} = 13\,\mu{\rm m}\,(a/\,{\rm au})^{1/2}(\mathcal{L}/\mathcal{L}_\odot)^{-1/4}$. The IR excess is then

$$f_{\rm IR}(\lambda) = \frac{B_\lambda(\lambda, T_{\rm dust})}{B_\lambda(\lambda, T_\star)} \int Q_{\rm a}\left(\frac{r}{R_\star}\right)^2 {\rm d}N_{\rm dust}, \quad (91)$$

where $T_\star$ and $R_\star$ are the stellar temperature and radius, $Q_{\rm a}$ is the absorption efficiency, and ${\rm d}N_{\rm dust}(r)$ is the number of dust particles as a function of their radius r. As discussed at the end of Section 5, the absorption efficiency can be approximated as

$$Q_{\rm a} = \min\{1, X\} \quad \text{where} \quad X \equiv \frac{2\pi r}{\lambda}, \quad (92)$$

so if the number of dust particles is a power law in radius,

$${\rm d}N_{\rm dust} = K r^{-q}\,{\rm d}r, \quad (93)$$

we have

$$f_{\rm IR}(\lambda) = \frac{B_\lambda(\lambda, T_{\rm dust})}{B_\lambda(\lambda, T_\star)} \frac{K}{(4-q)(q-3)R_\star^2}\left(\frac{\lambda}{2\pi}\right)^{3-q}. \quad (94)$$

The radius exponent q is related to the mass exponent p defined in equation (45) by $q = 3p - 2$. For small particles such as dust, we expect that the specific kinetic energy required for disruption, Q_D^*, is independent of mass (cf. Section 4.1). In this case, $p = 11/6$ (equation 50) so $q = 7/2$, and we will use this value in evaluating equation (94) numerically. Note that equation (94) is only valid if $3 \le q \le 4$ and if the minimum grain size in the distribution is much smaller than $\lambda/2\pi$.

The physical processes governing the effects of collisions are outlined in Section 4.1. In a steady state, the rate of dust mass production in a monodisperse planetesimal disc of the kind we are considering is

$$\Phi_m \approx \frac{mN}{t_{\rm c}}, \quad (95)$$

where as usual m and N are the mass and total number of planetesimals and $t_{\rm c}$ is the collision time, given by equation (27). To estimate the corresponding dust mass, we use the conservation of mass flux.

First, we generalize the collision time (27) to the case where particles of radius r_1 are colliding with particles of radius r_2. We have

$$t_{\rm c}^{-1}(r_1) = 2^{3/2}\pi^{1/2} f_1 \sigma_r \int_{r_{\rm min}}^{r_{\rm max}} (r_1 + r_2)^2 \frac{{\rm d}n_0(r_2)}{{\rm d}r_2}\,{\rm d}r_2, \quad (96)$$

here we have assumed that self-gravity is negligible ($\Theta \ll 1$). If the number density is a power law in radius, ${\rm d}n_0(r)/{\rm d}r \propto r^{-q}$, and the integral is dominated by projectiles with radii r_2 much less than the target radius r_1, we have

$$t_{\rm c}^{-1}(r_1) = \frac{2^{3/2}\pi^{1/2} f_1}{q-1}\frac{{\rm d}n_0(r_1)}{{\rm d}r_1}\sigma_r r_1^3 \theta^{q-1}, \quad (97)$$

where $\theta \equiv r_1/r_{\rm min} \gg 1$, that is $r_1 \gg r_{\rm min}$. If we define $r_{\rm min}$ to be the minimum projectile radius that will disrupt a grain of radius r_1, then the mass flux in the disc is roughly

$$\Phi_m(r_1) \approx \frac{m(r_1)}{t_{\rm c}(r_1)}\left(\frac{{\rm d}N_{\rm dust}}{{\rm d}\log r}\right)_{r_1}, \quad (98)$$

where $m(r_1) = 4\pi\rho_{\rm p} r_1^3/3$ is the mass of a grain of radius r_1. The number density and the total number of grains are related by equations (5) and (6),

$$\left(\frac{{\rm d}N_{\rm dust}}{{\rm d}\log r}\right)_{r_1} = K r_1^{1-q} = 2^{1/2}\pi^{3/2} f_m \frac{i_0}{e_0}\frac{\sigma_r a^2}{\Omega} r_1 \frac{{\rm d}n_0(r_1)}{{\rm d}r_1}. \quad (99)$$

Equating equations (95) and (98), using equation (99) to eliminate ${\rm d}n_0(r_1)/{\rm d}r_1$ in favour of K, and assuming that the velocity dispersion, internal density and radial distribution of the dust and planetesimals are equal, we have

$$\theta^{5/2} K^2 \approx 10 N^2 r^5 (1 + f_2\Theta/f_1), \quad (100)$$

which relates the properties of the dust distribution on the left-hand side of the equation to those of the parent planetesimals on the right-hand side. Although we have assumed $q = 7/2$ for the dust particles (i.e. the particles that dominate the IR emission), which is equivalent to the assumption that the specific dispersion energy Q_D^* is independent of mass for these particles, this derivation does not require any assumption about the dependence of Q_D^* on mass for larger bodies – the argument relates the mass flux from the destruction of equal-mass planetesimals (equation 95) to the mass flux in dust (equation 98) and since mass flux is conserved the properties of intermediate-mass bodies are irrelevant. The value of θ is estimated from the specific dispersion energy Q_D^*; following the discussion after equation (44) a typical value is $Q_D^* = 10^7$ erg g^{-1} and we equate $m(r_1)Q_D^*$ to $m(r_{\rm min})\sigma_r^2/2$ to obtain

$$\theta = 28\, e_0^{2/3}\left(\frac{a}{10\,{\rm au}}\frac{Q_D^*}{10^7\,{\rm erg\,g^{-1}}}\right)^{-1/3}. \quad (101)$$

The approximations that we have made to derive equation (100) are invalid unless $\theta \gg 1$.

Equation (100) determines the normalization K of the dust distribution, which is substituted into equation (94) to determine the IR excess. This derivation is for hot discs, and implicitly assumes a monodisperse disc in which collisions slowly feed a population of smaller debris. These assumptions are only valid for discs in which the collision time exceeds the age, $t_{\rm c} \gtrsim t_0$. Once $t_{\rm c} \sim t_0$ the appropriate model is a collision-limited disc (Section 4.3), and the factor $t_{\rm c}$ in equation (95) should be replaced by t_0. The analogue to equation (100) is then

$$\theta^{5/2} K^2 \approx \frac{15}{32\Omega t_0}\frac{f_m}{f_1}\frac{M_{\rm disc} a^2}{\rho_{\rm p}}. \quad (102)$$

As discussed in Section 3.2.2, collisions also occur in warm discs, but in this case the collisions do not disrupt the planetesimals and a collisional cascade is not established. We evaluate the IR emission properties of warm discs in Section 7.5 by assuming that most of the emission comes from the planetesimals themselves.

For calibration-limited detections – in which the limiting factor is the accuracy of the extrapolation of the photospheric flux to long wavelengths[8] – the minimum detectable flux produced by the dust, normalized by the stellar flux, is $f_{\rm det}$. For illustration, we set $f_{\rm det} = 0.1$ ($\lambda = 24\,\mu$m) and 0.55 ($\lambda = 70\,\mu$m), similar to the limits in Su et al. (2006). We show examples of hot planetesimal discs with detectable IR excesses in Fig. 6. It is apparent that the IR excess is a poor diagnostic for the disc mass $M_{\rm disc}$: the IR flux from discs of a given mass and semimajor axis in Fig. 6 can vary by more than an order of magnitude.

[8] See section 2.4 of Wyatt (2008) for a discussion of calibration- versus sensitivity-limited surveys.

© 2009 The Authors. Journal compilation © 2009 RAS, MNRAS **401,** 867–889

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

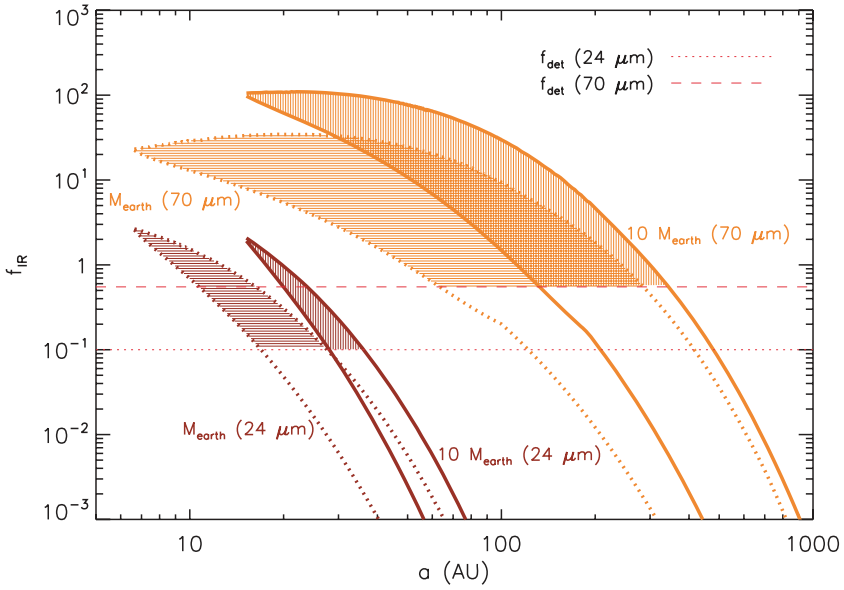

Figure 6. Allowed values of the infrared excess, $f_{\rm IR}(\lambda)$, for hot discs around solar-type stars with masses $M_{\rm disc} = M_\oplus$ and 10 $M_\oplus$. The shaded regions denote hot discs that are detectable via their IR excesses; the assumed detection thresholds $f_{\rm det}$ at $\lambda = 24$ and 70 μm are shown as horizontal lines.

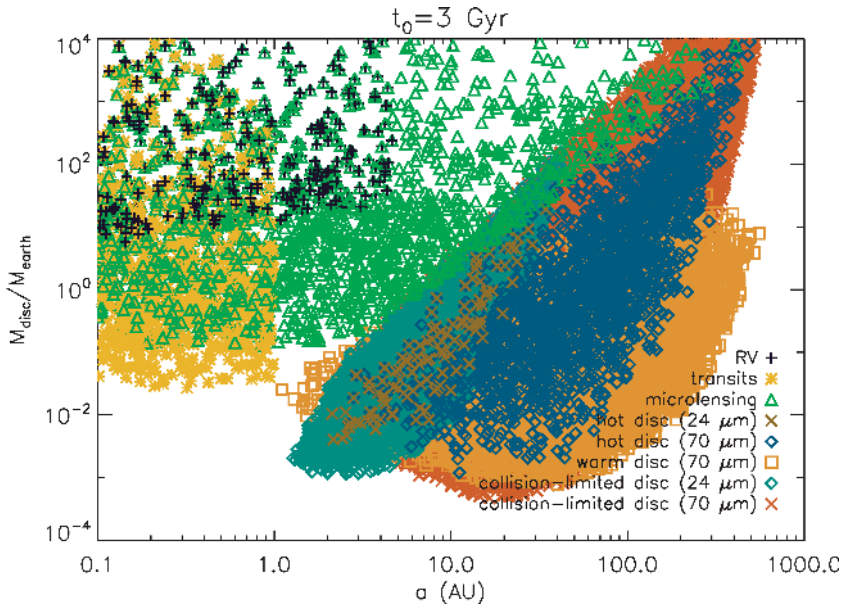

Figure 7. Different detection techniques probe different ranges of planetesimal disc mass and semimajor axis. Each of the four disc parameters (mass, semimajor axis, planetesimal radius, and velocity dispersion) is randomly generated and the various detection criteria are checked (see text). Only discs that survive for 3 Gyr are shown. The detectability criteria are summarized in Sections 7.5.1–7.5.5.

7.5 Probing disc mass and size

We now ask what long-lived planetesimal discs are detectable by the methods we have discussed in Sections 7.1–7.4. To efficiently explore the four-parameter space of disc mass and semimajor axis, planetesimal radius, and velocity dispersion ($M_{\rm disc}$, a, r, σ_r), we randomly generate 3×10^6 discs, uniformly sampled on logarithmic scales: $10^{-4} \leq M_{\rm disc}/M_\oplus \leq 10^4$, $0.1 \leq a/\,{\rm au} \leq 1000$, $10^{-6} \leq r/{\rm cm} \leq 10^{12}$ and $10^{-5} \leq \sigma_r/{\rm cm\ s^{-1}} \leq 10^7$. We then ask whether each disc can survive for $t_0 = 3\,{\rm Gyr}$ and is detectable by one or more methods using the detection thresholds described below. Fig. 7 shows the detectable planetesimal discs as projected onto the $M_{\rm disc}$–a plane.[9]

[9] Note that the density of generated points in Fig. 7 is generally lower at higher disc masses ($M_{\rm disc} \gtrsim 100\ M_\oplus$), which is a surprising result since massive discs should be easier to detect. The low density comes about because the range of allowed planetesimal radii for cold discs becomes narrower for higher disc masses (see Section 6.4 and Figs 3–5), and we are sampling $\log_{10} r$ uniformly.

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

7.5.1 Radial velocities

We consider a planetesimal disc to be detectable by this method if the orbital period $2\pi(a^3/G\,\mathrm{M}_\odot)^{1/2}$ is less than 10 yr and $v_{\rm wobble} \geq v_{\rm det}$, where $v_{\rm wobble}$ is given by equation (76) and the detection threshold $v_{\rm det} = 1\ \mathrm{m\ s^{-1}}$. Black crosses in Fig. 7 denote discs that survive for 3 Gyr and are detectable by this method. The minimum detectable mass scales $\propto a^{1/2}$, as expected. No warm or hot discs of age 3 Gyr were detectable by this method. Not surprisingly, planetesimal discs that are detectable by radial velocity variations in the host star tend to be massive and contain a small number of large bodies, i.e. planets; they resemble Disc A of Section 6.4.

7.5.2 Transits

As described in Section 7.2, planetesimals in nearly edge-on discs can be detected transiting their parent star if $r \gtrsim 0.35 R_\oplus$, and $a \lesssim 1$ au. We also require that $N_t > 1$ (equation 78) and $n_t < 1$ (equation 79). It could be argued that the condition $N_t > 1$ is unnecessarily stringent, since even if $N_t \ll 1$ a fraction of discs with these properties could be detected in a large transit survey.

No warm or hot discs of age 3 Gyr were detectable via transits. Discs A and B of Section 6.4 are detectable via transits.

7.5.3 Microlensing

We consider a planetesimal disc to be detectable via microlensing if the microlensing optical depth $\tau_{\rm lens}$ (equation 87) exceeds 10^{-6} (for comparison, the measured microlensing optical depth towards the Galactic bulge is 2–3 $\times\ 10^{-6}$). Since $\tau_{\rm lens}$ is maximized when the planetesimal is halfway to the source, we adopt $\zeta = d/d_\star = 0.5$ for illustration. We also assume a solar-type source star at a distance $d_\star = 8$ kpc. With these parameters, no warm or hot discs are detectable by microlensing. In Fig. 7, the cut-off for $M_{\rm disc} \lesssim 0.1 M_\oplus$ arises because the projected source size becomes larger than the Einstein radius, while for $a \gtrsim 1$ au we have the detectable disc mass scaling $\propto a^2$ (equation 87). Discs A, B, and C of Section 6.4 are detectable by microlensing.

7.5.4 Debris discs

Debris discs are dynamically hot discs that produce a collisional cascade of dust whose associated IR excess exceeds the detection threshold, $f_{\rm det}$. We consider both hot and collision-limited discs. As discussed in Section 7.4, we take $f_{\rm det} = 0.1$ (24 μm) and 0.55 (70 μm). In Fig. 7, there is a minimum and maximum detectable disc mass at a given semimajor axis. This feature appears to arise because the relaxation time $t_{\rm relax}$ (equation 37) at fixed disc mass and velocity dispersion has a minimum near $\Theta = 1$. There is also a cut-off at small semimajor axes, which arises because the maximum allowable surface density for hot discs (equations 68 and 71) is a strongly increasing function of semimajor axis. Discs E and F of Section 6.4 are detectable via their IR excesses at 70 μm.

It is remarkable that the discs detectable by radial-velocity/transit surveys or microlensing do not overlap with those detectable from their IR excesses. This result is consistent with the observational findings of Beichman et al. (2005), Bryden et al. (2006), Greaves, Fischer & Wyatt (2006), Moro-Martín et al. (2007) and Kóspál et al. (2009) that there is little or no correlation between the occurrence of planets and debris discs.

This lack of overlap does not preclude the possibility that a single host star may have planetesimals that are detectable by both methods, so long as the planetesimal disc extends over several octaves in semimajor axis. Beichman et al. (2005) and Kóspál et al. (2009) list six and 10 planet-bearing stars with debris discs, and planets have been imaged in the debris-disc systems HR 8799 (Marois et al. 2008) and Fomalhaut (Kalas et al. 2008).

7.5.5 Warm discs

In warm discs, collisions may cause cratering of the planetesimals but do not shatter them. Although cratering collisions produce significant amounts of dust, they do not establish a collisional cascade of the kind described in Section 4.2, so it is likely that IR emission is dominated by the planetesimals themselves. Based on this assumption, we show the detectability of warm discs from their IR emission in Fig. 7. We again adopt $f_{\rm det} = 0.55$ at 70 μm (we choose not to show detectable warm discs at 24 μm so as not to over-crowd Fig. 7).

In the figure, we see that there is again a minimum and maximum detectable disc mass at a given semimajor axis. This feature is now associated with the constraints set by non-gravitational forces ($t_{\rm PR} \gtrsim t_0$ and $\beta < 0.5$; see Section 5) and condition (40) for warm discs, respectively.

In principle, cold discs may also be detectable through the IR emission from the planetesimals, but we found no such discs given our assumed detection limits at 24 and 70 μm.

7.5.6 Undetectable discs

Many planetesimal discs that survive for 3 Gyr are not detectable using any of the methods described in this section.

8 DISCUSSION AND SUMMARY

8.1 Can warm discs mimic debris discs?

It is generally believed that the IR excesses around main-sequence stars older than a few Myr are due to dust that is produced in collisions between large solid bodies orbiting the star (hence the term 'debris disc'). Direct evidence that the emitting material is dust comes from several sources: (i) submillimetre observations of a handful of debris discs show that the absorption efficiency $Q_{\rm a}$ (equation 91) declines roughly as λ^{-1} for wavelengths $\gtrsim 100$ μm (Dent et al. 2000; Williams & Andrews 2006; Backman et al. 2009), suggesting grain sizes of a few tens of μm (equation 92). (ii) Chen et al. (2006) obtained *Spitzer Space Telescope* IR spectra of 59 stars with IR excesses and found five with 10–20 μm features that imply the presence of micron-sized silicate grains. (iii) The polarization of scattered light from the debris discs around β Pictoris and au Microscopii is consistent with simple models of scattering by dust (Gledhill, Scarrott & Wolstencroft 1991; Graham, Kalas & Matthews 2007).

Despite this evidence, it is instructive to consider the possibility that in some stars the IR excess arises not from dust produced by a collisional cascade but rather from a population of planetesimals with much larger radii. The most likely candidates are warm planetesimal discs, in which the collision time is less than the disc age but the collision velocities are too small to destroy the planetesimals over the lifetime of the disc. To simplify the calculations, we consider the lowest possible radial velocity dispersion for warm discs,

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

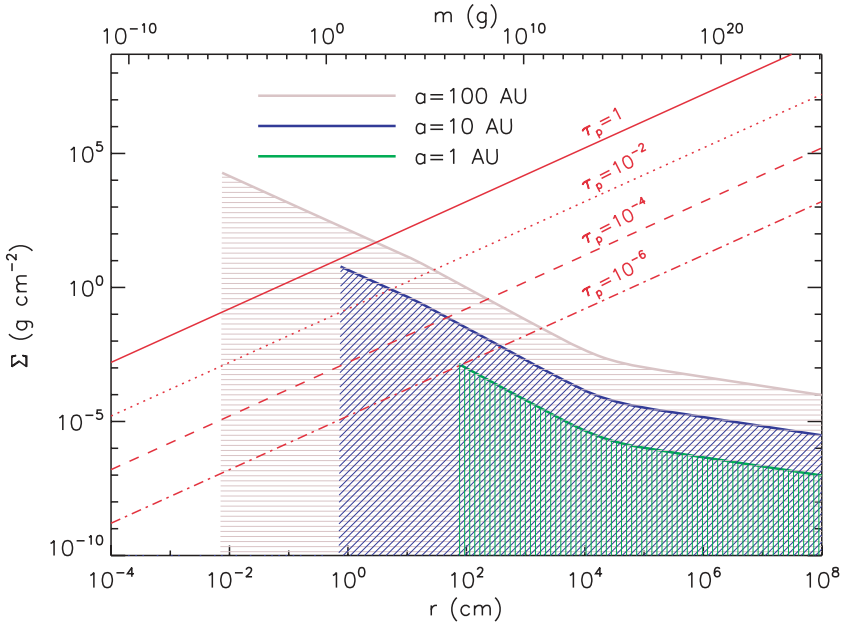

Figure 8. Allowed surface density of warm discs as a function of the planetesimal radius according to equations (103) and (104). The allowed region is shaded in different colours for different values of the semimajor axis $a = 1$, 10 and 100 au. The disc age is assumed to be 3 Gyr. Also shown are lines of constant optical depth $\tau_{\rm p}$ (equation 12).

which occurs when the Safronov number $\Theta = 1$ (cf. discs D, E, and F in Figs 4 and 5). Using equations (25), (27) and the third condition in equation (40), we obtain

$$r \leq \frac{1}{8\,(f_1 + f_2)} \left(\frac{Q_D^*}{G \Sigma \Omega t_0} \right). \tag{103}$$

If $\Theta < 1$, then the preceding constraint becomes stronger, that is the numerical coefficient in equation (103) becomes larger.

The minimum planetesimal size in warm discs is typically set by Poynting–Robertson drag. Using equations (58) and (63), the condition $t_{\rm PR} \gtrsim t_0$ yields

$$r \gtrsim 0.7\,{\rm cm}\,\mathcal{P}_\phi \left(\frac{t_0}{3\,{\rm Gyr}} \right) \left(\frac{\rho_{\rm p}}{3\,{\rm g\,cm^{-3}}} \right)^{-1} \left(\frac{a}{10\,{\rm au}} \right)^{-2}. \tag{104}$$

In Fig. 8, we show the constraints (103) and (104) for $a = 1$, 10 and 100 au; also shown are lines of constant optical depth $\tau_{\rm p}$ (equation 12). This optical depth is equal to the ratio of the bolometric disc luminosity to the bolometric stellar luminosity and hence provides a convenient measure of the detectability of the disc. Known discs typically have $\tau_{\rm p} \gtrsim 10^{-5}$ (Wyatt 2008). We conclude from Fig. 8 that the IR emission from so-called 'debris discs' at $a = 100$ au could in some cases be coming from planetesimals as large as $r \approx 10$ m ($m \approx 10^{10}$ g). A strong test of this possibility is that the emission spectrum from such a disc should resemble a blackbody spectrum, even at submm wavelengths (or a superposition of blackbody spectra if the emission originates from a range of disc semimajor axes).

8.2 The maximum optical depth of a debris disc

Wyatt et al. (2007) argue that a simple model for the collisional evolution of planetesimal discs implies that the maximum optical depth or fractional bolometric disc luminosity is (their equation 21)

$$\tau_{\rm p}^{({\rm max})} = 1.6 \times 10^{-4} \left(\frac{a}{\rm au} \right)^{7/3} \left(\frac{t_0}{\rm Myr} \right)^{-1}. \tag{105}$$

This result is based on several plausible but arbitrary assumptions (planetesimal radius $r = 2000$ km; rms eccentricity $e_0 = 0.05$, strength $Q_D^* = 2 \times 10^6$ erg g^{-1}, etc). To examine the applicability of this result, we employ the same Monte Carlo procedure used to produce Fig. 7 to randomly generate hot, collision-limited, and warm planetesimal discs and calculate the corresponding values of $\tau_{\rm p}$. For warm discs, the optical depth is given by equation (12) with N and r equal to the number and radius of the planetesimals. For hot and collision-limited discs, the optical depth is given by

$$\tau_{\rm p} = \frac{1}{4a^2} \int Q_{\rm a} r^2 \, {\rm d}N_{\rm dust}; \tag{106}$$

taking the absorption efficiency $Q_{\rm a}$ from equation (92) and the number of particles ${\rm d}N_{\rm dust}$ from equation (93) with $q = 7/2$, we find

$$\tau_{\rm p} = \frac{K}{a^2} \left(\frac{\lambda}{2\pi} \right)^{-1/2}, \tag{107}$$

where K is taken from equations (100) or (102) for hot and collision-limited discs, respectively. In the discussion below we assume $\lambda = 70\,\mu$m.

The results are shown in Fig. 9 for disc ages $t_0 = 30$ Myr, 300 Myr and 3 Gyr. We also show Wyatt et al.'s estimate (105) for $t_0 = 30$ Myr as a dashed white line; this is easily scaled to other ages since $\tau_{\rm p}^{({\rm max})} \propto 1/t_0$. Wyatt et al. estimate the uncertainties involved to span ~2 orders of magnitude and this is reflected in the light blue band shown in Fig. 9.

We are able to generate hot and collision-limited discs with optical depths substantially larger than the estimate of equation (105); however, these still lie within the estimated range of uncertainty given by Wyatt et al. (2007). A major source of uncertainty is in the planetesimal strength Q_D^*. In the calculations shown in Fig. 9, we used a mass-dependent Q_D^* as defined in equation (44). As a check, we carried out additional calculations assuming the constant value adopted by Wyatt et al. (2007) ($Q_D^* = 2 \times 10^6$ erg g^{-1}), and found that in this case our results (not shown) agreed more closely with equation (105).

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

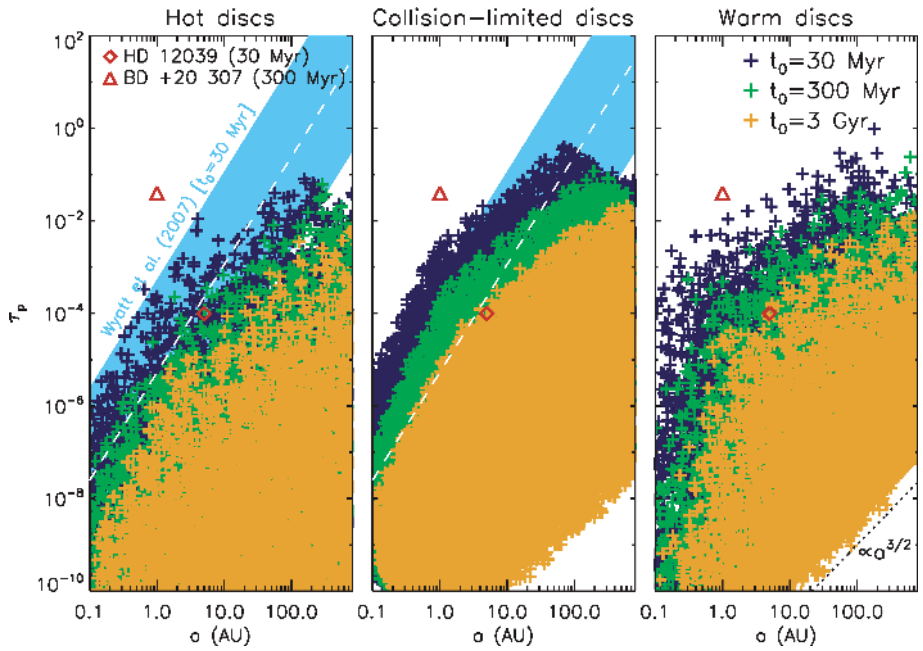

Figure 9. Optical depth associated with dust grains in hot (left) and collision-limited (middle) planetesimal discs, as well as for planetesimals in warm discs (right). Blue, green and yellow symbols are for planetesimal disc ages of 30 Myr, 300 Myr and 3 Gyr. The white dashed line and light blue band represent the maximum optical depth and its associated uncertainty, respectively, as estimated by Wyatt et al. (2007) – see equation (105). Also shown are the observed values for HD 12039 and BD +20 307.

Our simulations show that the maximum optical depth is roughly $\propto t_0^{-1}$, as predicted by equation (105), but the scaling with semimajor axis is quite different. The numerical simulations of Löhne et al. (2008) also show that the scaling of τ_p with a is generally more complicated than a power law (see top right-hand panel of their Fig. 11).

We also find that the maximum optical depth of warm discs can be almost an order of magnitude higher than that of hot and collision-limited discs of the same age and semimajor axis. The absence of warm discs in the bottom right corner of Fig. 9 is simply a consequence of the defining condition of warm discs, $t_c \lesssim t_0$ (equation 27), together with equation (64) relating the collision time to the optical depth. Discs exist below this line, but we label them 'hot' rather than 'warm'. For collision-limited discs, there is a similar cut-off caused by the thin-disc condition (equation 8) and equation (57) imposing a maximum value for m_{max} (e.g. see Fig. 5).

Wyatt et al. (2007) point out that a number of debris discs with $a \lesssim 10$ au have optical depths that exceed the limit (105) by factors of 10^3 or more (see also Moór et al. 2009). In Fig. 9, we show two debris discs with small semimajor axes, HD 12039 ($t_0 = 30$ Myr) and BD +20 307 ($t_0 = 300$ Myr), taken from table 1 of Wyatt et al. (2007). These stars have a range of spectral types, from F2 to K4, but our models based on a solar-type host star should still be reasonably accurate. We verify that the optical depth of BD +20 307 exceeds the maximum allowed for steady-state hot and collision-limited planetesimal discs with an age equal to the stellar age, while the optical depth of the disc around HD 12039 is consistent with steady-state models.

In Fig. 10, we show all seven debris discs listed in table 1 of Wyatt et al. (2007). For each system, we use the quoted values of the age t_0 and semimajor axis a to compute the maximum value of τ_p for hot, collision-limited and warm discs. The figure shows that two systems (HD 113766 and HD 12039) have optical depths consistent with a steady-state hot or collision-limited disc; one (BD +20 307) has an optical depth that is inconsistent with a steady-state hot, collision-limited or warm disc (by factors of 100, 50 and 10, respectively); and four (HD 72095, HD 69830, η Corvi, and HD 98800) are consistent with warm discs but not hot or collision-limited discs. However, of these last four, the first three have 10 μm silicate features in their spectra which imply that the IR emission comes from micron-sized grains, thus ruling out warm discs as well. All of our conclusions about hot and collision-limited discs are consistent with Wyatt et al. (2007), who suggest that the dust arises from planetesimals that have been scattered to small semimajor axes from a disc at much larger radii.

8.3 Summary

We have described a unified model of the evolution of gas-poor planetesimal discs, which is general enough to apply to all Keplerian discs of solid bodies, including debris discs, asteroid belts and planetary systems. Our model includes such processes as gravitational stability, evolution due to dynamical chaos, gravitational scattering, radiation and stellar wind pressure, Poynting–Robertson drag, and erosion or destruction by physical collisions. We characterize the discs by four parameters: disc mass (M_{disc}), disc semimajor axis (a), planetesimal size (r) and radial velocity dispersion or rms eccentricity (σ_r or e_0). The salient conclusions of our study include the following.

(i) Planetesimal discs can be categorized as dynamically 'hot', 'warm' or 'cold' depending on whether the planetesimal orbits cross and therefore collide and whether the collisions are erosive/disruptive. In cold discs the orbits do not cross and collisions do not occur; in hot discs the orbits cross but the collision time is longer than the disc age, and in warm discs the collisions are frequent but gentle enough that they do not substantially erode the particles within the age of the disc.

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

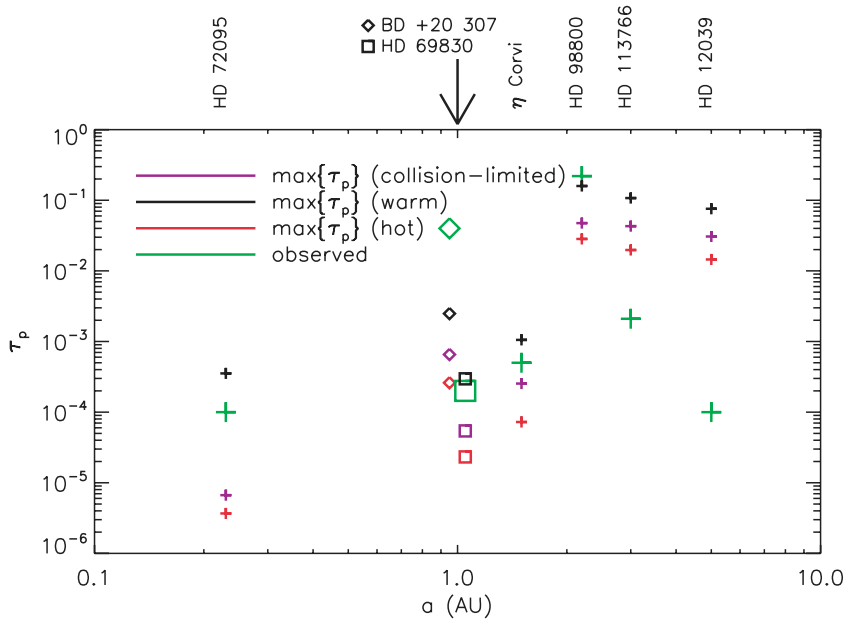

Figure 10. Maximum optical depths for hot and warm discs as calculated by our model, compared to the observed values, for the 7 systems listed in table 1 of Wyatt et al. (2007).

(ii) Massive discs with small semimajor axes can only survive for Gyr time-scales if they are cold. For example, after 3 Gyr hot discs at 1 or 10 au cannot exceed $1.3 \times 10^{-4} M_\oplus$ or $1.5 M_\oplus$ respectively (see Section 6.2). Gravitational stability imposes an upper limit on the number of planetesimals per octave that can be present in a cold disc of given surface density; for example, a cold disc of mass $100 M_\oplus$ cannot host more than 1–2 planetesimals per octave, while a disc of mass $1 M_\oplus$ can host $\sim$10 per octave (equation 67).

(iii) Warm discs can survive for Gyr time-scales over a wide range of semimajor axes and masses. At 1 au warm discs that survive for 3 Gyr must have mass $\lesssim 10^{-4} M_\oplus$; in this case the planetesimal radius is only 1 m, and warm discs composed of larger planetesimals must have even smaller masses (Fig. 8). At larger semimajor axes the allowed masses of warm discs and the planetesimals within them are much larger (Fig. 5). In some cases, warm discs may be detectable from the IR emission from the planetesimals themselves.

(iv) Planetesimal discs can be detected by a wide variety of observational techniques, including transits, gravitational microlensing, radial-velocity variations, and 'excess' IR emission ('debris discs'). With current technology the discs that can be detected by any of the first three methods are disjoint from those that can be detected in the IR (see Fig. 7). Many possible long-lived planetesimal discs cannot be detected by any method at present.

Despite the length of this paper, our analysis suffers from several shortcomings. The assumption of a monodisperse planetesimal disc is oversimplified, and probably incorrect given our limited understanding of disc formation. We suspect that our results are reasonably accurate provided that the total mass in the disc is dominated by planetesimals in a relatively small mass range, but this suspicion should be tested by analysis of discs with a range of planetesimal sizes. Our results also depend on a number of poorly determined parameters of order unity (Table 2) and do not incorporate a realistic model of the radial structure of the disc. In this paper, we have deliberately ignored all considerations of the formation process of planetesimal discs. It remains to be determined, by observations and theory, which of the wide variety of possible long-lived planetesimal discs are actually found in nature.

ACKNOWLEDGMENTS

We acknowledge support from the Institute for Advanced Study (IAS), NASA grant NNX08AH83G and NSF grant AST-0807444. KH is grateful for the Frank & Peggy Taplin Membership of the IAS. We thank the anonymous referee for many thoughtful comments that greatly improved the clarity and accuracy of our presentation.

REFERENCES

Agol E., 2002, ApJ, 579, 430
Aumann H. H. et al., 1984, ApJ, 278, L23
Backman D. et al., 2009, ApJ, 690, 1522
Batalha N. M., Jenkins J., Basri G. S., Borucki W. J., Koch D. G., 2002, in Favata F., Roxburgh I. W., Galadi D., eds, Stellar Structure and Habitable Planet Finding. ESA, Noordwijk. ESA SP-485, 35
Beer M. E., King A. R., Livio M., Pringle J. E., 2004, MNRAS, 354, 763
Beichman C. A. et al., 2005, ApJ, 622, 1160
Benz W., Asphaug E., 1999, Icarus, 142, 5
Bernstein G. M., Trilling D. E., Allen R. L., Brown M. E., Holman M., Malhotra R., 2004, ApJ, 128, 1364
Binney J., Tremaine S., 2008, Galactic Dynamics, 2nd edn. Princeton Univ. Press Princeton
Blum J., Wurm G., 2008, ARA&A, 46, 21
Brahic A., 1977, A&A, 54, 895
Bryden G., Beichman C. A., Rieke G. H., Stansberry J. A., Stapelfeldt K. R., Trilling D. E., Turner N. J., Wolszczan A., 2006, ApJ, 646, 1038
Burns J. A., Lamy P. L., Soter S., 1979, Icarus, 40, 1
Chambers J. E., Wetherill G. W., Boss A. P., 1996, Icarus, 119, 261
Chen C. H. et al., 2006, ApJS, 166, 351
Dent W. R. F., Walker H. J., Holland W. S., Greaves J. S., 2000, MNRAS, 314, 702
Dohnanyi J. W., 1969, J. Geophys. Res., 74, 2531
Dones L., Tremaine S., 1993, Icarus, 103, 67 (DT)
Farihi J., Jura M., Zuckerman B., 2009, ApJ, 694, 805

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022

Fernández J. A., 2005, Comets: Nature, Dynamics, Origin, and their Cosmogonical Relevance. Astrophysics and Space Science Library, 328. Springer, Dordrecht
Flynn G. J., Durda D. D., 2004, Planet. Space Sci., 52, 1129
Gledhill T. M., Scarrott S. M., Wolstencroft R. D., 1991, MNRAS, 252, 50P
Goldreich P., Tremaine S., 1980, ApJ, 241, 425
Goldreich P., Lithwick Y., Sari R., 2004a, ARA&A, 42, 549
Goldreich P., Lithwick Y., Sari R., 2004b, ApJ, 614, 497
Gould A., 1994, ApJ, 421, L71
Graham J. R., Kalas P. G., Matthews B. C., 2007, ApJ, 654, 595
Greaves J. S., Fischer D. A., Wyatt M. C., 2006, MNRAS, 366, 283
Greenzweig Y., Lissauer J. J., 1992, Icarus, 100, 440
Heng K., Keeton C. R., 2009, ApJ, in press (arXiv:0903.5304)
Hundertmark M., Hessman F. V., Dreizler S., 2009, A&A, 500, 929
Johansen A., Oishi J. S., Mac Low M-M., Klahr H., Henning T., Youdin A., 2007, Nat, 448, 1022
Jurić M., Tremaine S., 2008, ApJ, 686, 603
Kalas P. et al., 2008, Sci, 322, 1345
Kenyon S. J., Bromley B. C., 2004, AJ, 128, 1916
Kim D., Ryu Y.-H., Park B.-G., Chang H.-Y., Hwang K.-H., Chung S.-J., Lee C.-U., Han C., 2008, ApJ, 689, 1084
Koschn D., Grün E., 2001, Icarus, 154, 391
Kóspál Á., Ardila D. R., Moór A., Ábrahám P., 2009, ApJ, 700, L73
Krivov A. V., Löhne T., Sremčević M., 2006, A&A, 455, 509
Laskar J., 2008, Icarus, 196, 1
Löhne T., Krivov A. V., Rodmann J., 2008, ApJ, 673, 1123
Marois C., Macintosh B., Barman T., Zuckerman B., Song I., Patience J., Lafrenière D., Doyon R., 2008, Sci, 322, 1348
Moór A. et al., 2009, ApJ, 700, L25
Moro-Martín A. et al., 2007, ApJ, 658, 1312
O'Brien D. P., Greenberg R., 2003, Icarus, 164, 334
Paczyński B., 1996, ARA&A, 34, 419
Pan M., Sari R., 2005, Icarus, 173, 342
Papaloizou J. C. B., Terquem C., 2006, Rep. Prog. Phys., 69, 119
Parker A., Ivezić Ž., Jurić M., Lupton R., Sekora M. D., Kowalski A., 2008, Icarus, 198, 138
Reipurth B., Jewitt D., Keil K., 2007, Protostars and Planets V. University of Arizona Press, Tucson
Safronov V. S., 1972, Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets, NASA TTF-677, p. 206
Schneider P., Ehlers J., Falco E. E., 1992, Gravitational Lenses. Springer-Verlag, Berlin
Stewart G. R., Ida S., 2000, Icarus, 143, 28
Stewart S. T., Leinhardt Z. M., 2009, ApJ, 691, L133
Su K. Y. L. et al., 2006, ApJ, 653, 675
Thébault P., Augereau J.-C., 2007, A&A, 472, 169
Toomre A., 1964, ApJ, 139, 1217
Tremaine S., Zakamska N. L., 2004, AIP Conf. Proc. Vol. 713, The Search for Other Worlds, Am. Inst. Phys., Melville, NY, p. 243
Williams J. P., Andrews S. M., 2006, ApJ, 653, 1480
Witt H. J., Mao S., 1994, ApJ, 430, 505
Wyatt M. C., 2008, ARA&A, 46, 339
Wyatt M. C., Smith R., Greaves J. S., Beichman C. A., Bryden G., Lisse C. M., 2007, ApJ, 658, 569
Yoshinaga K., Kokubo E., Makino J., 1999, Icarus, 139, 328
Zheng Z., Ménard B., 2005, ApJ, 635, 599
Zuckerman B., 2001, ARA&A, 39, 549

APPENDIX A: MODIFYING VARIOUS FORMULAE IN DONES & TREMAINE (1993)

The rate of mass accretion in a rotating disc of planetesimals has been evaluated by Greenzweig & Lissauer (1992) and Dones & Tremaine (1993; hereafter DT). We need to modify their formulae, because their results are for the mass accretion rate of a large body (a planet) on a circular orbit in the mid-plane of the planetesimal disc, while we are interested in the rate for a typical planetesimal in a monodisperse planetesimal disc. Unless otherwise mentioned, the notation used in this Appendix is the same as in the main text.

In the dispersion-dominated regime, the rate of mass accretion is given by equations (72) and (90) of DT,

$$\dot{M} = \begin{cases} 2.7603\Sigma\Omega R_{\rm p}^2, & \Theta \ll 1, \\ 6.0828\Sigma\Omega^3 R_{\rm p} R_{\rm H}^3 \sigma_r^{-2}, & \Theta \gg 1, \end{cases} \tag{A1}$$

where $R_{\rm p}$ is the radius of the planet, $R_{\rm H} = a(M/{\rm M}_\odot)^{1/3}$ is its Hill radius (as defined by DT, which is different from the definition in the present paper) and σ_r is the planetesimal velocity dispersion in the radial direction. In the shear-dominated regime, we use equations (83) and (75) of DT,

$$\dot{M} = \begin{cases} 10.1\Sigma\Omega^2 R_{\rm p} R_{\rm H}^2 \sigma_r^{-1}, & \sigma_r \gtrsim \Omega\sqrt{R_{\rm p} R_{\rm H}}, \\ 6.47\Sigma\Omega R_{\rm p}^{1/2} R_{\rm H}^{3/2}, & \sigma_r \lesssim \Omega\sqrt{R_{\rm p} R_{\rm H}}. \end{cases} \tag{A2}$$

We first write these formulae in terms of the number density of planetesimals in the mid-plane, n_0, using $\Sigma = \sqrt{2\pi} n_0 m \sigma_z / \Omega$ (see text below equation 18 of DT),

$$\dot{M} = \begin{cases} 2.7603\sqrt{2\pi} n_0 m \sigma_z R_{\rm p}^2, \\ 6.0828\sqrt{2\pi} n_0 m \sigma_z \Omega^2 R_{\rm p} R_{\rm H}^3 \sigma_r^{-2}, \\ 10.1\sqrt{2\pi} n_0 m \sigma_z \Omega R_{\rm p} R_{\rm H}^2 \sigma_r^{-1}, \\ 6.47\sqrt{2\pi} n_0 m \sigma_z R_{\rm p}^{1/2} R_{\rm H}^{3/2}. \end{cases} \tag{A3}$$

The collision time as defined in the present paper is $t_{\rm c}^{-1} = \dot{M}/m$. The following modifications are made:

$$\begin{aligned} &\sigma_{r,z} \to \sqrt{2}\sigma_{r,z}, \\ &R_{\rm p} \to 2r, \\ &R_{\rm H} = a\left(M/{\rm M}_\odot\right)^{1/3} \to a\left(2m/{\rm M}_\odot\right)^{1/3}. \end{aligned} \tag{A4}$$

The first modification comes from assuming the colliding bodies have the same velocity dispersion, as opposed to one of them being on a circular orbit. The second and third modifications arise both colliding bodies have the same radius r and mass m, as opposed to one large body having radius R and mass M while the other has negligible mass and radius. We also note that $\sigma_z = \sigma_r(i_0/e_0)$, where we again choose $i_0/e_0 = 0.5$. Thus, the reciprocal of the collision time is

$$t_{\rm c}^{-1} = \begin{cases} 2.7603 \times 4 \times \sqrt{\pi} n_0 r^2 \sigma_r, \\ 6.0828 \times 2 \times \sqrt{\pi} n_0 \Omega^2 r a^3 \sigma_r^{-1} \left(m/{\rm M}_\odot\right), \\ 10.1 \times 2^{7/6} \times \sqrt{\pi} n_0 \Omega r a^2 \left(m/{\rm M}_\odot\right)^{2/3}, \\ 6.47 \times 2 \times \sqrt{\pi} n_0 r^{1/2} a^{3/2} \sigma_r \left(m/{\rm M}_\odot\right)^{1/2}. \end{cases} \tag{A5}$$

Finally, we decrease $t_{\rm c}^{-1}$ by $\sqrt{2}$ since the vertical motions of the particles imply that the mean density is reduced by this factor compared to the mid-plane density. We also replace n_0 by $\sqrt{2/\pi}\mathcal{N}\Omega/\sigma_r$,

$$t_{\rm c}^{-1} = \begin{cases} 2.7603 \times 4 \times \mathcal{N}\Omega r^2, \\ 6.0828 \times 4 \times \mathcal{N}\Omega r^2 \Theta, \\ 10.1 \times 2^{7/6} \times \mathcal{N}\Omega^2 r a^2 \sigma_r^{-1} \left(m/{\rm M}_\odot\right)^{2/3}, \\ 6.47 \times 2 \times \mathcal{N}\Omega r^{1/2} a^{3/2} \left(m/{\rm M}_\odot\right)^{1/2}. \end{cases} \tag{A6}$$

Comparing with equation (27), we get $f_1 = 4 \times 2.7603/16 = 0.690$ and $f_2 = 6.0828/4 = 1.521$. Similarly, by comparison with equation (30), we get $f_4 = 2^{7/6} \times 10.1 = 22.67$ and $f_5 = 2 \times 6.47 = 12.94$.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article/401/2/867/1148442 by guest on 20 August 2022