

Statistics of collision parameters computed from 2D simulations

Á. Süli^{1,2}★¹*Department of Astronomy, Eötvös University, H-1117 Budapest, Pázmány Péter sétány 1/A, Hungary*²*Konkoly Observatory, Research Centre for Astronomy and Earth Sciences, H-1121 Budapest, Konkoly Thege Miklós út 15-17, Hungary*

Accepted 2021 March 2. Received 2021 March 2; in original form 2020 March 11

ABSTRACT

There are two popular ways to speed up simulations of planet formation via increasing the collision probability: (i) confine motion to 2D, (ii) artificially enhance the physical radii of the bodies by an expansion factor. In this paper, I have performed 100 simulations each containing 10^4 interacting bodies and computed the collision parameters from the results of the runs. Each run was executed for a lower and a higher accuracy parameter. The main goal is to determine the probability distribution functions of the collision parameters and their dependence on the expansion factor. A simple method is devised to improve the determination of the collision parameters from the simulation data. It was shown that the distribution of the impact parameter is uniform and independent of the expansion factor. For real collisions, the impact velocity is greater than 1 mutual escape velocity, a finding that can be explained using the two-body problem. The results cast some doubts on simulations of the terrestrial planets' final accretion that have assumed merge. Collision outcome maps were created adopting the fragmentation model of recent studies to estimate the number of different types of collisions. A detailed comparison with earlier works indicates that there are similarities as well as significant differences between the different works. The results indicate that as the planetary disc matures and the masses of the bodies differ progressively than the majority of collisions lead to mass growth either via partial accretion or via graze-and-merge collision.

Key words: methods: analytical – methods: data analysis – methods: numerical – methods: statistical – protoplanetary discs.

1 INTRODUCTION

According to the most widely accepted model, the nebular hypothesis states that the formation process of the terrestrial planets is the result of a series of three consecutive but partly overlapping stages (Lissauer 1993; Chambers 2004; Morbidelli et al. 2012). In the first stage, the micrometre-sized dust grains coagulate into small aggregates through pairwise collisions (Dullemond & Dominik 2005). The collisions may proceed further to form 1–100 km-sized bodies, which are called planetesimals. It must be noted that this process making planetesimals can only be effective if the conditions are suitable (Weidenschilling 1977, 1997). Weidenschilling (1977) has described the dust motion in protoplanetary discs and showed that dust grains from micron to a few metres in size experience a radial motion towards the star. This radial drift strongly depends on the size and bodies with a critical size spiral into the star in a fraction of the disc lifetime. This loss of material quickly clears the disc preventing planetesimal formation. This depletion process is called the ‘radial-drift barrier’ of planet formation (Lai, Gonzalez & Maddison 2012) and was first studied in a minimum mass solar nebula in which the critical size corresponds to metre-sized bodies and thus later was inaccurately referred to as the ‘metre-size barrier’.

There exist an alternative hypothesis, where planetesimals or even Mars-size embryos may form directly from small dust grains concentrated by turbulence followed by gravitational collapse (Jo-

hansen, Youdin & Mac Low 2009; Cuzzi, Hogan & Bottke 2010), or in pressure maxima of protoplanetary discs (Lyra et al. 2008). Moreover, if planetesimals are concentrated in a pressure maximum, they can be quickly accreted by a growing embryo leading to the rapid formation of a solid core and a giant planet (Lyra et al. 2008; Sándor, Lyra & Dullemond 2011; Guilera & Sándor 2017). In this study, I focus on the collision theory.

Despite of the radial-drift barrier, the Solar system and the observed exoplanets prove the existence of an efficient and robust process that creates planetesimals in large numbers. Once planetesimals have formed their gravitational interactions control further growth, which determines their velocities and the parameters of occasional collisions. If turbulence is not too vehement, then dynamical friction ensures that the largest planetesimals have low relative velocities. This leads to runaway growth in which the largest objects grow more rapidly than smaller ones (Wetherill & Stewart 1989; Kokubo & Ida 1996).

The next stage is the oligarchic phase when the largest objects contain enough mass to dominate the velocity dispersion of smaller planetesimals (Ida & Makino 1993). In the oligarchic growth mode, the disc may partitioned into rings where each ring is dominated by a single planetary embryo that sweeps up planetesimals in its vicinity (Kokubo & Ida 1998). Nearby protoplanets grow at similar rates and their orbits are separated by more than 5 Hill radius.

Oligarchic growth ends when the number of planetesimal drops below such a threshold where their damping effect on protoplanet orbits becomes insufficient to prevent orbit crossing. This initiates the last phase of terrestrial planet formation, involving the collision

★ E-mail: a.suli@astro.elte.hu

and mutual accretion of protoplanets and embryos, the so-called giant impacts phase. During this phase, the number of embryos declines and the final planets forge, in stable non-crossing orbits. Most of the evolution took place after the gas disc had dissipated, so the processes mainly consisted of gravitational interactions and occasional collisions between embryos.

In the early stages of planet formation, the vast number of objects present in the dynamical system render the problem unfeasible for direct N -body methods. But at the beginning of the last stage, the number of significant bodies drops to a few 10^2 to 10^3 hence N -body integrations become applicable for calculating the evolution of the system. These methods are accurate and were used extensively to study the final assembly of the terrestrial planets.

According to previous studies (Chambers 2001, 2013; Liu, Zhou & Wang 2011), the time-scale of planetary systems formation is ~ 100 Myr. Integrating the motion of several 10^2 to 10^3 gravitationally interacting bodies for ~ 100 Myr still requires a huge computational effort consequently 2D model was frequently used. Another method to speed up N -body simulations of planet formation is to scale up the physical radii of the bodies by a factor. Both of these techniques increase the collision probability between objects and thus reducing the computational time.

The 2D integrations of planet formation (Cox & Lewis 1980; Lecar & Aarseth 1986; Beaugé & Aarseth 1990; Alexander & Agnor 1998) have notably contributed to our understanding of planetary accretion, and motivated scholars to investigate the problem in more detail or extend the model into 3D. On the other hand, there are findings indicating that 2D simulations may not provide a feasible model of a planet formation. According to Kokubo & Ida (1996), runaway growth takes place in 3D simulations but not in 2D, although this growth mode might have played a key role in planet formation. Another result of Chambers & Wetherill (1998) indicated that the time-scales of collision in 2D and 3D simulation differ significantly (a factor of ~ 10) thus the evolution of large bodies – the major concern in planet formation – follows different paths.

The numerical N -body simulations that assumed perfect accretion have been successful at reproducing the broad characteristics of the terrestrial planets in our Solar system (Wetherill 1994; Chambers 2001; Quintana et al. 2002, 2016; Raymond et al. 2004, 2006, 2009; O’Brien et al. 2006; Quintana & Lissauer 2006, 2014; Liu et al. 2011). These results were all based on a small (up to a dozen) number of realizations performed for each set of initial conditions. A larger set of 50 simulations of planet formation around the Sun was recently performed by Fischer & Ciesla (2014), who using the perfect-accretion model, demonstrated the need for a larger suite of simulations in order to infer results from a distribution of final planet configurations.

To date the late stage has been mainly investigated by accurate N -body simulations while collisions were modelled generally by perfect accretion. However, this oversimplified assumption breaks for collisions with higher velocity and/or larger impact parameter (Kokubo & Genda 2010). This simplification was necessary in order to keep the number of bodies below the initial number and the lack of detailed models on collisions between planetary mass bodies. Having a better concept of collisions and how it influences the evolution of the planets is a primary key in understanding how planets emerge from a swarm of planetesimals. For this very reason, Leinhardt & Stewart (2012, hereafter LS12) have conducted high-resolution simulations of collisions between planetesimals and the results were used to isolate the effects of different impact parameters on collision outcome. Their model predicts the size and velocity distribution of fragments as a function of the impact velocity, impact angle, and projectile to target mass ratio. The authors identified the

boundaries between different types of collision: (i) simple merger, (ii) merger with some mass escaping as fragments, and (iii) hit-and-run collisions. The analytical model developed by LS12 is a powerful tool that has two major advantages: it improves the physics of collisions in numerical simulations of planet formation and collision evolution, and is easy to adapt to an N -body code.

The collision model of LS12 made possible to implement more sophisticated, inter-particle gravity enabled N -body simulations of late-stage planet formation. Lines et al. (2014) used this model to simulate planet formation in the Kepler-34(AB) system’s circumbinary protoplanetary disc to examine whether planets can form in a hostile environment. The same collision model for gravity-dominated bodies has been implemented into an N -body tree code (Bonsor et al. 2015) and used to examine planet formation.

Chambers (2013) implemented this comprehensive collision model into the widely used *Mercury* integration package (Chambers 2001). Eight simulations of planet formation using the new collision model were presented and compared to eight simulations that were previously performed using the perfect-accretion collision model (Chambers 2001). The new simulations form 3 to 5 terrestrial planets moving on widely spaced orbits with growth complete by 400 My. The final planets that formed in each of these sets were shown to be comparable despite the significant difference in the number and frequency of collisions. In addition, the accretion time-scales were about twice as long when fragmentation was included.

Terrestrial planet formation is characterized by countless collisions among bodies highlighting that collisions are the core agent of planet formation. Impacts outcome span multiple regimes depending on the collision parameters. The collision history of the final body can largely influence the planet’s growth, stability, bulk composition, and habitability (Chambers 2001, 2013; Bonsor et al. 2015).

In this paper, I present the results of 100 simulations of planet formation in 2D started from the beginning of the stochastic stage. The main focus of the work is to present a detailed and comprehensive picture of the statistical distribution of the collision parameters. The dependence of these statistics on the expansion factor is also studied and a comparison of the results with others is presented. This is a first step towards examining the statistical distribution of the collision parameters. With a reliable statistics, one can assess the frequency of different collision outcome and posterior estimate the number of collisions lead to perfect merging. This is essential since in the majority of the N -body simulations colliding bodies were assumed to merge into a single new object; however, this basic premise has not been thoroughly and deliberately tested. This is a very strong assumption and can largely bias the result of planet formation, thus there is an impetus to assess the occurrence of perfect merging. The presented statistics can be also useful for N -body modellers to perform Monte Carlo simulations of planet formation including the model of LS12 (Stewart & Leinhardt 2012).

Using the scaling laws derived by LS12, Stewart & Leinhardt (2012) presented a retrospective analysis of previous simulations to estimate the range of true collision outcomes. In this work, a similar analysis is performed on the collision data adopting the LS12 model to estimate the occurrence of collision outcomes and the results are compared to previous works.

The rest of this paper is organized as follows. Section 2 contains a description of the simulations and initial conditions. Section 3 describes the results of these simulations, including the speed of the runs, the description of the collision geometry, the collision data and the distribution of the collision parameter. Section 4 contains a simple analytical model which is used to discuss the observed impact velocity distribution. Section 5 shows the collision outcome maps

Table 1. The mass, radius, and the density of the protoplanets. In the second column, the values are given in au and solar mass. In the third column, the values are given in Earth mass and radius units, while in the last column in SI units.

m_p	$5.1230212063 \times 10^{-10}$	1.71×10^{-4}	1.01903×10^{21}
R_p	$3.3120486271 \times 10^{-6}$	7.77×10^{-2}	4.95475×10^5
ρ_p	3.3662582520×10^6	–	2.0×10^3

derived by the model of [LS12](#) and the collisions of the simulations are plotted on the maps. A comparison with other works is also presented and the main results are summarized in Section 6.

2 THE SIMULATIONS

The equations of motion of 10^4 protoplanets around a star in the barycentric coordinate system O_{xy} was integrated. The protoplanets are confined to 2D and placed initially in the terrestrial region extending from 0.5 to 1.5 au. In this model, each body was interacting with all the other bodies and all bodies were modelled as a sphere.

In order to determine the total mass of solids in the ring, the parameters of the minimum mass solar nebula ([Hayashi 1981](#)) were used in which the surface density of solids is

$$\Sigma_s = \Sigma_1 r^{-\frac{3}{2}}, \quad (1)$$

where $\Sigma_1 = 7 \text{ g cm}^{-2}$ is the surface density of solids at 1 au and r is the distance from the star. Integrating equation (1) from 0.5 to 1.5 au results in $m_{\text{tot}} = 5.123 \times 10^{-6} M_{\odot} = 1.71 M_{\oplus}$ where M_{\odot} and M_{\oplus} denote the mass of the Sun and the Earth, respectively.

The simulations were started with $N_p = 10^4$ bodies around a star with $1 M_{\odot}$. Each protoplanet has the same physical properties: $m_p = m_{\text{tot}}/N_p = 5.123 \times 10^{-10} M_{\odot} \approx M_{\text{Ceres}}$ and the density is $\rho_p = 2.0 \text{ g cm}^{-3}$ which corresponds to the mean density of Ceres and consistent with silicon-rich rocky bodies. Consequently, the radius of the bodies R_p is approximately 500 km (see [Table 1](#)).

The semimajor axes a for this number of protoplanets are generated randomly with probabilities weighted in order to reproduce the disc surface density profile as defined by equation (1). Eccentricities e are randomized from a Rayleigh distribution with rms values of 0.02 with an upper limit of 0.2. The mean anomaly M and argument of pericentre ω are randomized uniformly from $[0, 2\pi]$.

In the simulations, the gas disc had dissipated by this stage and all collisions were assumed to be perfectly inelastic, forming a new body by conserving mass and linear momentum. Most of the previous work on planet formation used this oversimplified treatment of collisions.

The N -body systems are stochastic and so a large number of simulations of a given system, with changes in the initial conditions, are required in order to reach relevant conclusions. Most numerical simulations designed to explore the stochastic stage lack the large number of realizations needed to account for the stochastic nature of N -body systems. I improve on this limitation by performing 100 simulations of planet formation around a Sun-like star. To study the dependence on the expansion factor f used to artificially enhance the physical radii of the bodies, I have performed five different set of simulations for $f = 1, 2, 3, 5, 10$. In order to have statistically meaningful results 10 sets of initial conditions were generated, denoted by run id = 1, 2, ..., 10. For each value of f , all these 10 initial conditions were integrated for $T = 10^6 \text{ yr}$. The integrator used an adaptive step size with a tolerance of $\epsilon = 10^{-10}$ and $\epsilon = 10^{-13}$ for all 50 cases, resulting in total 100 simulations.

The applied integrator is the Runge–Kutta–Fehlberg 7(8) algorithm with adaptive step size, which has an acceptable speed and

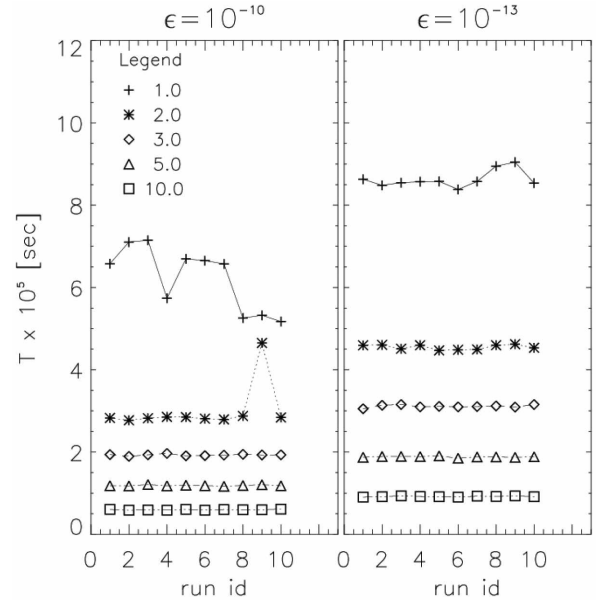


Figure 1. The run-time of all the 2×50 simulations for the two accuracy parameters. The data are plotted for $f = 1, 2, 3, 5, 10$ as plus sign, asterisk, diamond, triangle, and square symbols, respectively.

accurate in most situations ([Fehlberg 1968](#)). All simulations were performed with an open source GPU code `RED.CUDA`¹ designed to integrate planet and planetesimal dynamics in the framework of the core accretion planet formation. The `RED.CUDA` is written in CUDA C and runs on all NVIDIA GPUs with compute capability of at least 2.0. All simulations were performed on Tesla K20Xm device containing 2688 CUDA Cores and use CUDA driver version and run-time version 7.0 and 6.5, respectively.

3 RESULTS

3.1 Speed of simulation run

I show how the typical CPU run-time of a simulation denoted by T depends on f . [Fig. 1](#) displays T for each f value and for each of the 10 different runs. In the left-hand panel, the data correspond for the $\epsilon = 10^{-10}$ accuracy level, while the right one for $\epsilon = 10^{-13}$. The statistics of the data are given in [Table 2](#) where the minimum, maximum, mean, and standard deviation of T is calculated from the sample of the 10 runs belonging to a given f . The standard deviations, denoted by sd are usually more useful to describe the variability of the data. From [Fig. 1](#) and [Table 2](#), it is clear that the variation of the run-time is the most significant for $f = 1$ and it gradually diminishes as f gets larger.

The other observable fact is the decrease of T as f increases. Each mean run-time \bar{T} (fourth and eighth columns in [Table 2](#)) were divided by the maximum of the means:

$$Q(f) = \frac{\bar{T}(f)}{\max(\bar{T}(f))} = \frac{\bar{T}(f)}{\bar{T}(1)}, \quad (2)$$

and the ratios Q were plotted in [Fig. 2](#) as a function of f for both accuracy values. From the figure, it is clear that Q is approximately inversely proportional to f , and the $1/f$ function (dotted curve) is apparently a very good approximation of the measured data. In 2D,

¹<https://github.com/suliaron/red.cuda>

Table 2. The run-time in seconds for $\epsilon = 10^{-10}$ and $\epsilon = 10^{-13}$. In the second column the minimum, in the third the maximum, in the fourth the mean, and in the fifth the standard deviation of the T is given for $\epsilon = 10^{-10}$. The same values are given sixth–ninth for $\epsilon = 10^{-13}$. In the last column, q is the ratio between the means, see equation (3).

f	$\epsilon = 10^{-10}$				$\epsilon = 10^{-13}$				q
	min(T)	max(T)	\bar{T}	sd(T)	min(T)	max(T)	\bar{T}	sd(T)	
1	5.172×10^5	7.147×10^5	6.224×10^5	7.727×10^4	8.384×10^5	9.046×10^5	8.630×10^5	2.061×10^4	1.39
2	2.772×10^5	4.650×10^5	3.009×10^5	5.773×10^4	4.469×10^5	4.620×10^5	4.550×10^5	5.753×10^3	1.51
3	1.895×10^5	1.966×10^5	1.928×10^5	1.990×10^3	3.051×10^5	3.155×10^5	3.112×10^5	3.118×10^3	1.61
5	1.167×10^5	1.209×10^5	1.186×10^5	1.330×10^3	1.849×10^5	1.905×10^5	1.882×10^5	1.536×10^3	1.59
10	5.843×10^4	6.117×10^4	5.965×10^4	8.770×10^2	9.024×10^4	9.330×10^4	9.177×10^4	1.040×10^3	1.54

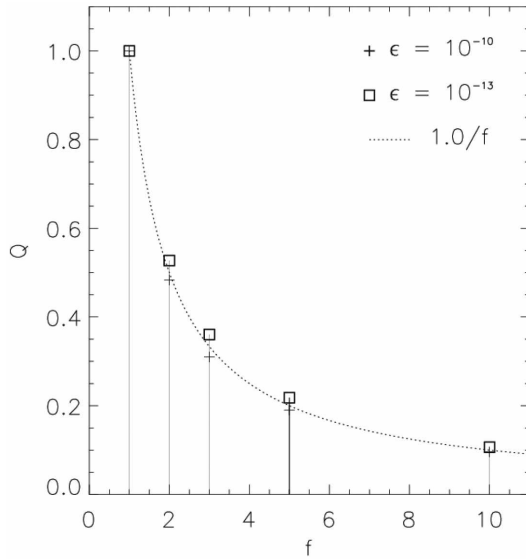


Figure 2. The variation of Q as a function of f for $\epsilon = 10^{-10}$ is plotted with plus sign and for $\epsilon = 10^{-13}$ plotted with square. The curve of $1/f$ is plotted by dotted line.

one can state that increasing the collision cross-section of the bodies by f will reduce the simulation time by f .

The q value in Table 2 is calculated for each f as

$$q(f) = \frac{\bar{T}(f; \epsilon = 10^{-13})}{\bar{T}(f; \epsilon = 10^{-10})}, \quad (3)$$

which has a mean of 1.527, indicating a 53 per cent increase in the mean run-time for a 3 orders of magnitude higher accuracy.

The run-time can be written as

$$T(f) = \sum_{i=0}^n T_i(f) = \eta \sum_{i=0}^n N^2(t_i, f) \delta t_i, \quad (4)$$

where n is the number of collisions, $N(t_i, f)$ is the number of bodies at time t_i , δt_i is the simulation time between the i th and $(i+1)$ th collision, and η is a parameter characterizing the hardware. The number of bodies versus time is depicted in Fig. 3 for different f on logarithmic scale. We have an almost identical figure for $\epsilon = 10^{-13}$. These curves are piecewise constant functions since each collision reduces the number by 1. It is visible that the curves in the $t \leq 1$ and $t \geq 10^5$ yr intervals are similar and in between their behaviour are alike, but the bigger f is the smaller the time when the fast decrease begins. These curves match very well with previous 2D simulations e.g. Lecar & Aarseth (1986) and Alexander & Agnor (1998). According to equation (4), the $t \leq 1$ part hardly contributes to the total run-time since it is only 1 yr (δt_i is small), while in the $t \geq 10^5$ interval N is a

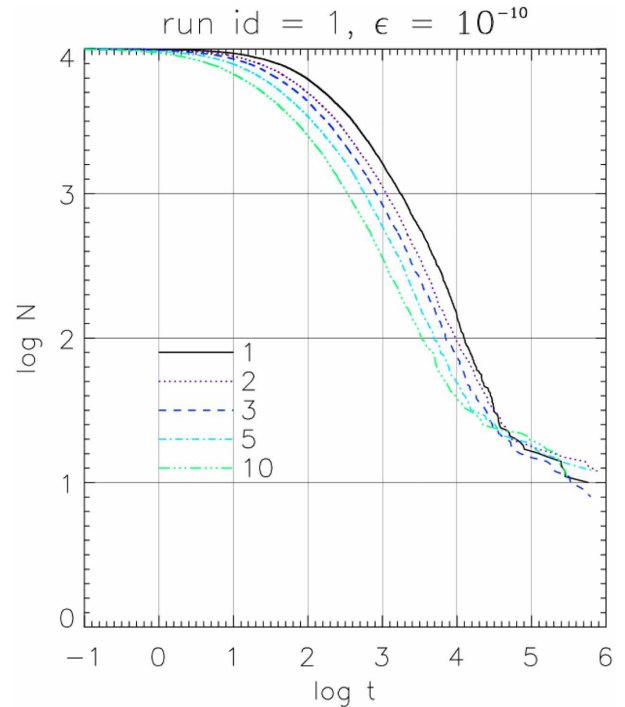


Figure 3. The logarithm of the number of protoplanets as a function of the logarithm of time. The $f = 1$ case is shown by black solid line, $f = 2$ dotted purple, $f = 3$ blue dashed, $f = 5$ light blue dot–dashed and the $f = 10$ green 3 dot–dashed line.

few dozens for all f values therefore it takes the same amount of time to complete.

In order to demonstrate the dependence of the run-time on f shown in Fig. 2, let's write the ratio of the instantaneous run-time belonging to $f = 1$ and f for specific simulation time epochs as follows

$$Q_k(f) = \frac{T(\tau_k, f)}{T(\tau_k, 1)} \approx \left(\frac{N(\tau_k, f)}{N(\tau_k, 1)} \right)^2, \quad k = 0, \dots, 4, \quad (5)$$

where $\tau_k = 10^k$. The results are depicted in Fig. 4 for the five different time epochs. We have an almost identical figure for $\epsilon = 10^{-13}$. It is clearly visible that with increasing t the Q_k ratios fit better and better to the $1/f$ function. The best fit is for $k = 2, 3,$ and 4 (the Q_3 and Q_4 curves overlap). According to equation (4) and Fig. 3, it is obvious that in these cases both N and δt are large, therefore principally this interval determines the run-time. This reasoning is not an exact proof why the run-time depends on the reciprocal of f but clearly demonstrates the underlying cause, i.e. the specific decrease of N as shown in Fig. 3.

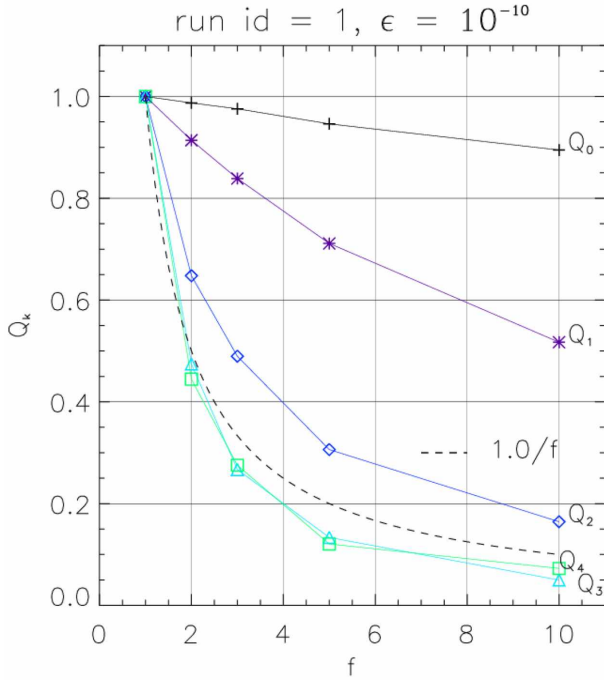


Figure 4. The variation of Q_k as a function of f . The curve of $1/f$ is plotted by dashed line.

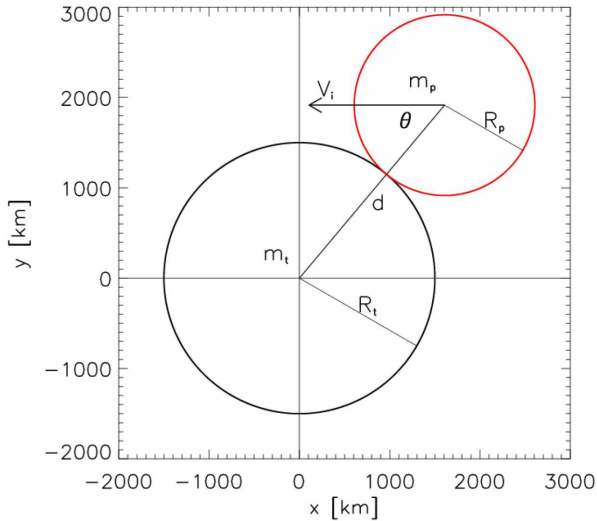


Figure 5. Schematic of the collision geometry.

3.2 Collision geometry and parameters

In the model initially all bodies have the same mass, but as collisions take place the masses will differ. Fig. 5 shows the geometry of a collision. In what follows the target is the body with the larger mass m_t and the other body is called the projectile with mass m_p . In the figure, the target is stationary and the projectile is moving from right to left with speed V_i , the units on the x and y axes are in km. The impact angle θ , is defined at the time of first contact as the angle between the line connecting the centres of the two bodies and the impact velocity vector (when $\theta = 0^\circ$ the collision is referred as head-on, while the $\theta = 90^\circ$ case corresponds to a grazing collision). The

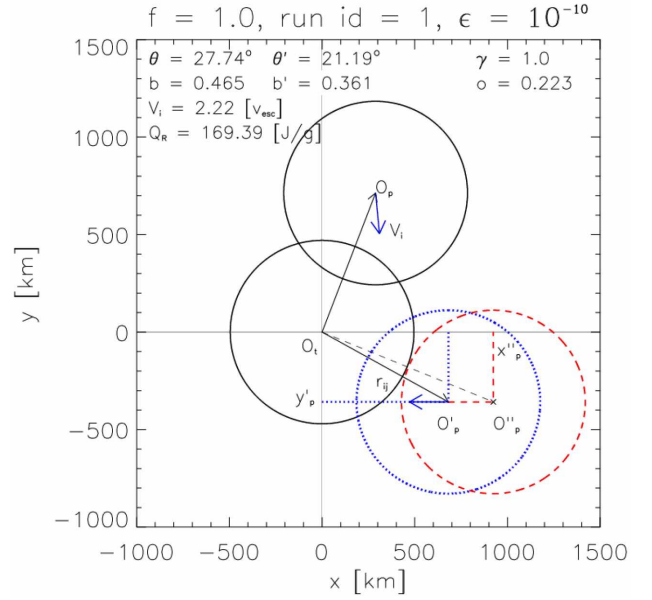


Figure 6. A collision of equally massive bodies from the simulation with $f = 1$ and run id = 1 and $\epsilon = 10^{-10}$ is depicted. For details see the text.

impact parameter b is

$$b = \sin \theta = B/d, \quad (6)$$

where B is the y coordinate of the projectile's centre and d is distance between the centres of the bodies.

The impact parameter has a notable influence on the collision outcome, because the kinetic energy of the projectile may only partially intersect the target for oblique impact, e.g. in the collision geometry shown in Fig. 5, the top of the projectile does not directly hit the target. Consequently, a portion of the projectile may shear off and only the kinetic energy of the lower part of the projectile will be involved in disrupting the target. According to LS12, the outcome of a collision is dependent on the kinetic energy of the interacting mass that strongly depends on b .

In order to describe the dependence of catastrophic disruption on impact angle, two geometrical collision groups were introduced by LS12: non-grazing, where most of the projectile interacts with the target and grazing where less than half the projectile interacts with the target. Following Asphaug (2010), the critical impact parameter

$$b_{\text{crit}} = \frac{R_t}{R_t + R_p}, \quad (7)$$

is reached when the centre of the projectile is tangent to the surface of the target. Grazing impacts are defined to occur when $b > b_{\text{crit}}$.

In Fig. 6, a non-grazing collision from the simulation with $f = 1$, run id = 1 and $\epsilon = 10^{-10}$ is plotted. The reference frame is fixed at the target whose centre is marked by O_t , the $O_t x$ and $O_t y$ axes are parallel with the Ox and Oy axes, respectively. In the figure, O_p marks the centre of the projectile. The solid circles denote the colliding bodies. The origin of the impact velocity vector \mathbf{V}_i is at O_p . To better visualize the impact geometry, the system is rotated until the impact velocity vector became parallel with the $O_t x$ axis and points towards left (as in Fig. 5). The rotated projectile is plotted by a dashed blue circle and its centre is denoted by $O'_p = (x'_p, y'_p)$. The impact parameter b and angle θ are

$$b = \frac{|y'_p|}{r_{ij}}, \quad \theta = \arcsin(b), \quad (8)$$

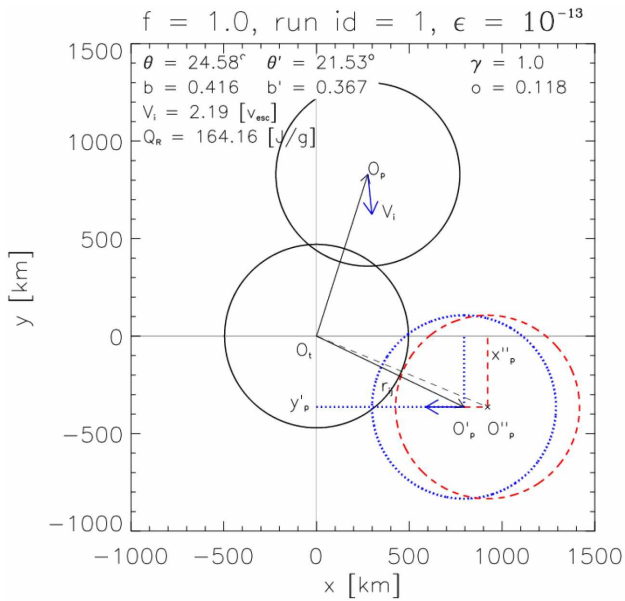


Figure 7. Same as Fig. 6 but for $\epsilon = 10^{-13}$.

where $r_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$ is the mutual distance of the point masses, and \mathbf{r}_i is the particle's barycentric position vector.

The projectile and the target are overlapping which is a consequence of the collision detection method applied in the numerical code. After each integration step, the mutual distances of the bodies are calculated and compared against the sum of the bodies' radii multiplied by f . A collision happens whenever the

$$r_{ij} \leq f(R_i + R_j) \quad (9)$$

criterion fulfilled.

When solving the equations of motion, the particles are treated as point masses therefore two bodies can get arbitrary close. Since the collision detection is done after the integration step, the collision can be detected only after the real physical contact hence overlapping is inevitable. The extent of the overlap can be defined as

$$o = 1 - \frac{r_{ij}}{f(R_i + R_j)}, \quad (10)$$

when $o = 0$ the two spheres just touch each other. A consequence of the overlap is that the parameters defining the collisions cannot exactly determined from the data produced by the numerical simulation. An improvement to this problem is described in Section 3.3.

The impact velocity is given in mutual escape velocity unit which is defined as follows (Genda et al. 2012):

$$V_{\text{esc}} = k_G \sqrt{2 \frac{m_t + m_p}{R_t + R_p}}, \quad (11)$$

where k_G is the Gaussian constant of gravity. In the literature on planetary collisions, Q_R denotes the specific energy of impact and it is given by

$$Q_R = \frac{1}{2} \frac{m_t v_t^2 + m_p v_p^2}{m_t + m_p}, \quad (12)$$

where v_t and v_p are the speed of the target and projectile with respect to the centre of mass, respectively. The specific impact energy is shown in the upper left corner of Figs 6 and 7. The projectile-to-target mass ratio, $\gamma = m_p/m_t$ and the overlap o are displayed in the upper right corner.

Table 3. Summary of parameters with their values from Figs 6 and 7. In the last two columns, the absolute Δ and relative change Δ_r is given where the reference values are taken from the $\epsilon = 10^{-10}$ case.

	$\epsilon = 10^{-10}$	$\epsilon = 10^{-13}$	unit	Δ	Δ_r (per cent)
o	0.223	0.118	–	–0.105	47.1
b	0.465	0.416	–	–0.049	10.5
b'	0.361	0.367	–	0.006	1.7
V_i	2.222	2.187	$[V_{\text{esc}}]$	–0.035	1.6
V_i'	2.206	2.180	$[V_{\text{esc}}]$	–0.026	1.2
Q_R	169.39	164.16	$[J g^{-1}]$	–5.24	3.1
Q_R'	166.92	163.01	$[J g^{-1}]$	–3.91	2.3
γ	1.0	1.0	–	0	0

3.3 Improve the parameters

The difficulty caused by the overlap of the bodies can be overcome by the following method. Let us assume that the direction of the relative velocity vector \mathbf{v}_{ij} does not change during the last integration step (after which the collision was detected). This assumption was checked with additional simulation where the step size was 10 s and \mathbf{v}_{ij} was monitored. It was found that \mathbf{v}_{ij} hardly changed after the first contact. In this case, it is possible to shift the rotated projectile (dotted blue circle on Fig. 6) along the x -axis (i.e. parallel with \mathbf{V}_i) until the first contact of projectile and target. The x_p'' -coordinate of the shifted projectile is (see Figs 6 and 7)

$$x_p'' = \sqrt{f^2 (R_t + R_p)^2 - y_p'^2}, \quad (13)$$

and the improved collision parameter b' and the associated collision angle θ' can be computed as

$$b' = \frac{|y_p'|}{f(R_t + R_p)}, \quad \theta' = \arcsin(b'). \quad (14)$$

In Fig. 6, one can see that $b = 0.465$ and $\theta = 27.74^\circ$, while the numeric values of the improved parameter is $b' = 0.361$ and $\theta' = 21.19^\circ$. Since $r_{ij} \leq f(R_t + R_p)$ therefore it follows that $b' \leq b \leq 1$ and $\theta' \leq \theta$ which means that the real impacts tend to be more head-on collisions than the ones detected from the raw numerical data.

To check this method, the results from the $\epsilon = 10^{-13}$ simulation are used and the same collision is displayed in Fig. 7. As expected by using higher accuracy, the collision detection happens with a significantly smaller overlap of 0.118. The computed $b = 0.416$ is smaller than in the $\epsilon = 10^{-10}$ case. From Fig. 7, the improved parameter $b' = 0.367$, which agrees very well with the value computed from the $\epsilon = 10^{-10}$ case using equation (14) proving the efficiency of the method.

Using the above procedure and applying the law of energy conservation, the impact velocity and specific impact energy can also improve. The total mechanical energy E_1 when the collision is detected (blue circles on Figs 6 and 7) and the energy of the shifted projectile (red circles) E_2 are equal and it allows one to compute the improved impact speed of the shifted projectile:

$$V_i' = \sqrt{V_i^2 - k_G^2 m_t \left(\frac{1}{r_{ij}} - \frac{1}{f(R_t + R_j)} \right)}. \quad (15)$$

Table 3 presents the parameters discussed in detail above for the two accuracy parameters. The value of o decreased by 47 per cent, b by 10 per cent, and b' increased by 2 per cent for the higher precision. The V_i , V_i' , Q_R , and Q_R' are almost the same for the two precisions, the relative changes are less than 3 per cent. From equation (15), it

follows that the real impact velocity is slightly less than the measured one.

3.4 Description of the collision data

During the simulations, the \mathbf{r}_i and \mathbf{v}_i vectors and the physical properties of the colliding bodies were recorded in a data base. Hereafter, I refer to a data set as *single run data* when I select from the data base records with a specific f and a single run id, and to a data set as *multiple run data* when I select from the data base records with a specific f and multiple run ids. For the multiple run, data I always selected the run ids from 1 to 10.

In general, at the end of each simulation ~ 10 bodies remained so for each single run the number of collisions n , i.e. the sample size is $n \approx 9990$ therefore a multiple run data set contains approximately 99 900 collisions. In total for the 50 runs, there were 499 507 and 499 455 collisions for $\epsilon = 10^{-10}$ and $\epsilon = 10^{-13}$, respectively. In this paper, I used this data base to study the distribution of the collision parameters. The details of the samples are given in Table A1 for $\epsilon = 10^{-10}$ and $\epsilon = 10^{-13}$, where the rows with run = 1–10 correspond to multiple run data, while the others correspond to single run data. In the table, the minimum and maximum values of V_i and Q_R are also listed. For $\epsilon = 10^{-10}$, the values are on the left side, while for $\epsilon = 10^{-13}$ on the right side.

According to the results, the minimum value of b' is 9.22×10^{-7} and 2.99×10^{-8} for $\epsilon = 10^{-10}$ and $\epsilon = 10^{-13}$, respectively, while the maximum is 1 for both ϵ . This parameter has a theoretical range of $[0, 1]$. From the data, it is clear that it does not depend on f . To save space b' is not listed.

In Table A1 with $f = 1$, the smallest value of V_i' is 0.978 while the biggest is 19.517. The variation of $\min(V_i')$ is minor for a fixed f and its value decreases with increasing f from 0.99 to 0.2. The values of $\max(V_i')$ do not decrease with increasing f , they are $8.356^{+2.001}_{-1.146}$ except for one case for $f = 1$ and run id = 6 when $\max(V_i') = 19.517$. In Table A1, the data for this specific run without the 19.517 value are also presented (empty cells have the same value as above). It turns out, that the maximum of V_i' drops down to 8.299, which is well in the other runs' range. This is an extreme value and in this single run data all other impacts are within the above range.

The minimum of Q'_R is 1.23×10^4 J kg $^{-1}$ while the maximum is 1.31×10^7 J kg $^{-1}$ for $f = 1$. The minimum values fall in a rather narrow range for a specific f that decreases with increasing f . The maximum values fall in a limited range for a specific f that shows a decreasing trend with increasing f .

For each f , the table shows the range for the multiple run data, designated by 1–10 in the run column. For $f = 1$, the data are listed containing the extreme impact speed and without it.

As reported by Table A1 in the case of $\epsilon = 10^{-13}$, the sample sizes are in almost perfect agreement with the lower accuracy runs, the ranges of the parameters are very similar and the data show analogous behaviour. In this case, there is no extreme value. Comparing the results for the two accuracy parameters, one can conclude that the size of the populations, the ranges of the impact speed, and specific impact energy are very similar.

In Table 4, the mean of the minimum and maximum values of b' are calculated and listed along with their standard deviations. It is clear that these values practically do not depend on f : the minimum is very close to 0 while the maximum to 1, the standard deviation is very limited for both quantities, the deviations around the mean is small. The figures are alike for $\epsilon = 10^{-13}$, therefore they are not shown.

In Table 5, the mean of the minimum and maximum values of V_i' are calculated from the values given in Table A1 and listed along

Table 4. The mean of the minimum (second column) and maximum (fourth column) of b' for multiple run data and their standard deviations in the third and fifth columns, respectively.

f	$\overline{\min(b')}$	$\text{sd}(\min(b'))$	$\overline{\max(b')}$	$\text{sd}(\max(b'))$
1	1.56×10^{-4}	1.28×10^{-4}	0.999985	4.74×10^{-5}
2	6.80×10^{-5}	5.84×10^{-5}	0.999686	3.88×10^{-4}
3	1.22×10^{-4}	9.77×10^{-5}	0.999404	6.96×10^{-4}
5	5.94×10^{-5}	6.32×10^{-5}	0.999093	7.41×10^{-4}
10	1.40×10^{-4}	8.73×10^{-5}	0.999274	4.04×10^{-4}

Table 5. The same for V_i' as in Table 4.

f	$\overline{\min(V_i')}$	$\text{sd}(\min(V_i'))$	$\overline{\max(V_i')}$	$\text{sd}(\max(V_i'))$
$\epsilon = 10^{-10}$				
1	0.991	0.005	8.712	0.805
2	0.689	0.016	8.268	0.587
3	0.559	0.008	8.119	0.517
5	0.418	0.016	8.401	0.743
10	0.275	0.029	8.278	0.484
$\epsilon = 10^{-13}$				
1	0.990	0.004	8.322	0.641
2	0.686	0.020	8.260	0.789
3	0.554	0.015	8.062	0.779
5	0.423	0.009	8.134	0.643
10	0.279	0.015	8.105	0.520

Table 6. The same for Q'_R as in Table 4.

f	$\overline{\min(Q'_R)}$	$\text{sd}(\min(Q'_R))$	$\overline{\max(Q'_R)}$	$\text{sd}(\max(Q'_R))$
$\epsilon = 10^{-10}$				
1	1.312×10^4	1.485×10^3	4.415×10^6	7.681×10^5
2	7.670×10^3	7.084×10^2	2.665×10^6	5.158×10^5
3	5.473×10^3	7.611×10^2	2.378×10^6	3.368×10^5
5	3.489×10^3	3.670×10^2	2.454×10^6	4.300×10^5
10	1.789×10^3	2.761×10^2	2.358×10^6	2.778×10^5
$\epsilon = 10^{-13}$				
1	1.279×10^4	1.756×10^3	4.498×10^6	1.125×10^6
2	7.127×10^3	5.111×10^2	2.957×10^6	8.200×10^5
3	5.123×10^3	6.030×10^2	2.459×10^6	6.417×10^5
5	3.030×10^3	4.610×10^2	2.282×10^6	3.684×10^5
10	1.776×10^3	3.311×10^2	2.262×10^6	2.803×10^5

with their standard deviations for $\epsilon = 10^{-10}$ and $\epsilon = 10^{-13}$. The mean of the minimum value depends strongly on f , it decreases from 0.99 to 0.28 in both cases, while the maximum essentially does not depend on it, but stays around 8.3. The standard deviation is confined to a narrow interval for both the minimum and maximum, indicating modest variability of the values.

In Table 6, the extreme values and the mean of Q'_R derived from the data given in Table A1 are shown along with their standard deviations for the two accuracy parameters. As expected, the mean of the minimum value depends strongly on f , it decreases from $\sim 13 \times 10^3$ to $\sim 1.8 \times 10^3$ in both cases, while the maximum drops from $\sim 4.4 \times 10^6$ to $\sim 2.7 \times 10^6$ and then stays constant around this value as f increases.

As a final note the mean, median, and sd for all multiple run data were calculated for the two accuracy parameters. These results are presented in Table 7. The median is the value separating the higher half from the lower half of a data sample. The impact parameters'

Table 7. The mean, median, and standard deviation of b' , V'_i , and Q'_R . The results are given for all multiple run data.

f	\bar{b}	Med(b')	sd(b')	\bar{V}'_i	Med(V'_i)	sd(V'_i)	\bar{Q}'_R $\times 10^5$	Med(Q'_R) $\times 10^5$	sd(Q'_R) $\times 10^5$
$\epsilon = 10^{-10}$									
1	0.486	0.482	0.283	1.836	1.517	0.903	1.964	1.311	2.184
2	0.487	0.485	0.284	1.627	1.309	0.961	1.541	0.941	1.777
3	0.489	0.488	0.284	1.550	1.234	0.991	1.406	0.804	1.707
5	0.486	0.483	0.282	1.479	1.173	1.023	1.298	0.686	1.686
10	0.485	0.483	0.282	1.419	1.129	1.051	1.216	0.600	1.673
$\epsilon = 10^{-13}$									
1	0.494	0.492	0.286	1.826	1.508	0.907	1.947	1.286	2.159
2	0.494	0.491	0.286	1.624	1.302	0.969	1.535	0.927	1.782
3	0.495	0.494	0.287	1.544	1.224	0.998	1.402	0.793	1.721
5	0.495	0.495	0.287	1.480	1.173	1.033	1.304	0.685	1.710
10	0.493	0.491	0.287	1.424	1.131	1.062	1.229	0.599	1.699

mean and median values are close to 0.5, which is the middle of its range. In the case of V'_i , the mean is definitely larger than the median for all data sets. This implies that there are some large values that affect the mean and most of the values are closer to the median than to the mean. This fact also reflects in the relative large sd values. Since the specific impact energy is a strong function of the impact speed, the median of Q'_R is significantly smaller than the mean. Again this is visible from the larger spread of the data, given the larger sd values.

3.5 The bin size

To construct the histograms, the first step is to bin the range of values and then count how many values fall into each interval. In this paper, the bins are consecutive, adjacent, and non-overlapping intervals of a variable. The bins have equal size. Since there is no ultimate method to determine the number of bins, after some experimentation I found that the Rice rule gives the best number of bins:

$$k = [2n^{\frac{1}{3}}], \quad bs = \frac{\max - \min}{k}, \quad (16)$$

where bs is the bin size, \min , \max denotes the minimum and maximum of the underlying variable and the braces indicate the ceiling function. Throughout this paper, I use normalized histograms to display relative frequencies which is the proportion of cases that fall into each of several categories, with the sum of the heights equalling 1. Histograms give an approximation of the probability density function, pdf, of the underlying variable.

A cumulative histogram refers to the running total of the values. That is, the cumulative histogram H_i of a histogram h_j is defined as:

$$H_i = \sum_{j=1}^i h_j. \quad (17)$$

Since the collision parameters have continuous probability distribution hereafter I will use the terms pdf and cdf.

3.6 Distribution of collision parameters

3.6.1 Analysis of the correction for the overlap

In Section 3.3, a method was presented to improve the impact parameter, speed, and specific impact energy (see equations 14 and 15). Here, I show how it influences their distribution. In Fig. 8, I have summarized the result for $f = 1$ and run id = 1 with $\epsilon =$

10^{-10} . Panel (a) shows the pdf of θ (solid black curve) and θ' (red dotted curve), panel (b) displays the pdf of b (solid black curve) and b' (red dotted curve). The number of collision is $n_1 = 9991$ and the bin size is $bs(\theta') = 2.045^\circ$ and $bs(b) = bs(b') = 0.023$. The bin size for θ is $\approx 2.4^\circ$ which is larger than for θ' . The reason for this difference originates from that $\max(\theta) = 105^\circ$ which is larger than the theoretically allowed 90° ; another reason to improve the parameters.

In Fig. 8 from panel (a) it is clearly visible that the distribution of θ' is different from that of θ . For $\theta < 40^\circ$, the relative frequency of θ is similar to that of θ' , while for $\theta > 40^\circ$ the pdf curves diverge more and more, the relative frequency of θ is higher than θ' . Our 2D results do not support the observation of Shoemaker (1962), who concludes that a 45° impact angle ($b = 0.707$) is most probable. Preliminary results of 3D runs show evidence of a peak around 45° therefore it seems that there is a significant difference between 2D and 3D model when collision parameter is an important aspect. The causes of this discrepancy should be discussed in a separate paper. From Table 8, cdf(θ') reaches 90 per cent at $\theta' \approx 61^\circ$ and 99 per cent at $\theta' \approx 77^\circ$. These values are shown with dotted and solid black vertical lines denoted by θ'_{90} and θ'_{99} , respectively.

In Fig. 8 panel (b) the distribution of b and b' are depicted. These are very similar for $b \leq 0.8$ ($\approx 53^\circ$): the frequency of b' is about 5–10 per cent higher than b . For $b > 0.9$, the difference becomes prominent. For the last bin, the relative frequency of b is 0.06 while it is 0.01 for b' . From Table 8, the cdf(b') reaches 0.9 at $b' \approx 0.864$ and 0.99 at $b' \approx 0.977$. These values are shown with dotted and solid black vertical lines denoted by b'_{90} and b'_{99} , respectively. By eyeballing the data, it is reasonable to state that the distribution of b' is uniform in the range $[0, 0.95]$, and has a sharp decrease in the last bin.

In Fig. 8 panel (c) the distributions of V_i and V'_i are shown. The pdf and cdf functions for these quantities are very similar. According to Table 8, one gets that 90 per cent of V'_i are less than 2.908 and 99 per cent are less than 4.824. These values are shown with dotted and solid black vertical lines denoted by $V'_{i,90}$ and $V'_{i,99}$, respectively. There are no data with $V'_i < 0.993$, the maximum of the pdf is around 1 than it decreases fast as V'_i gets larger. For $V_i \geq 5$, the probability of a collision is less than 1 per cent.

Panel (d) shows the distribution of Q_R and Q'_R . The pdf and cdf functions are almost identical. From Table 8, cdf reaches 90 per cent at $Q'_R \approx 3.07 \times 10^5$ and 99 per cent at $Q'_R \approx 10^6$ J kg $^{-1}$. These values are shown with dotted and solid black vertical lines denoted by $Q'_{R,90}$ and $Q'_{R,99}$, respectively.

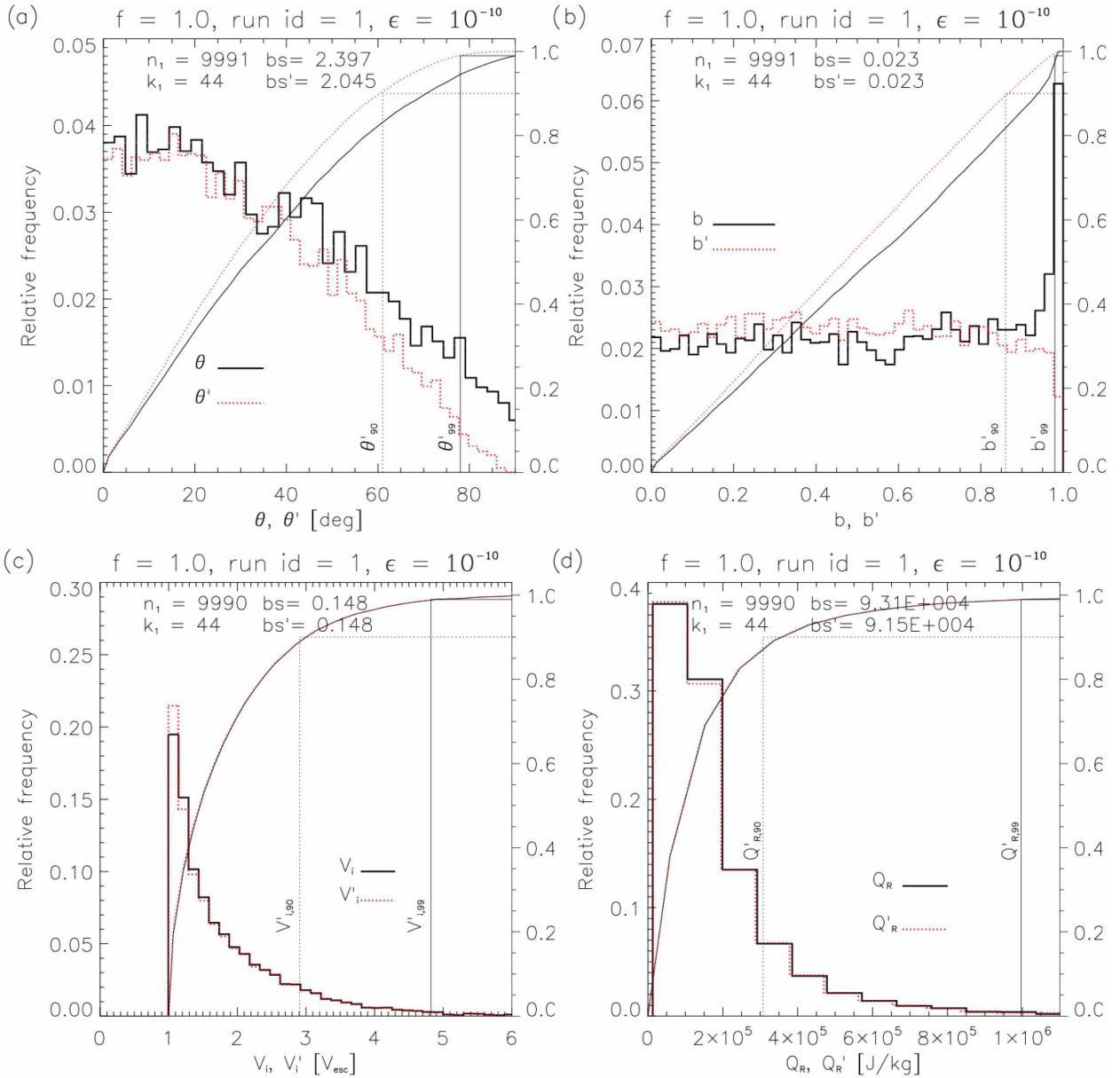


Figure 8. Comparison of the distribution of the impact parameters with their improved version for $\epsilon = 10^{-10}$ and run id = 1. On the left vertical axis, the relative frequency is shown and the corresponding pdf curves are plotted with solid black and dotted red line segments for the raw and the improved quantity, respectively. The cdf curves computed from equation (17) for the raw and improved quantities are also plotted with thin black and dotted red lines. The scale for the cdf is given on the right vertical axis. On panel (a) θ and θ' are compared. The two vertical lines denote the location where cdf reaches 90 per cent and 99 per cent, respectively. On panel (b) the b and b' , on panel (c) V_i and V_i' while on panel (d) Q_R and Q_R' are plotted. The legends are the same for each panel.

Table 8. Summary of the values for $f = 1$ and run id = 1 at which the cdf reaches 90 per cent and 99 per cent for $\epsilon = 10^{-10}$ and $\epsilon = 10^{-13}$.

cdf	$\epsilon = 10^{-10}$		$\epsilon = 10^{-13}$	
	90 per cent	99 per cent	90 per cent	99 per cent
θ' (deg)	61.364	77.727	61.364	79.773
b'	0.864	0.977	0.886	0.977
V_i' (V_{esc})	2.908	4.824	3.084	5.176
Q_R' (J kg^{-1})	307 734	996 513	335 928	1066 795

As it is apparent from Fig. 8 the difference between the raw and improved quantities is not negligible, primarily in the case of the impact angle, parameter (panels a and b) and speed (panel c). This finding provides a further argu-

ment to use the correction for the overlap method described in Section 3.3.

In Fig. 9, the same quantities for $\epsilon = 10^{-13}$ are shown. The overall behaviour of the curves is very similar to the $\epsilon = 10^{-10}$ case but, as expected, the difference between the raw and improved quantities is smaller.

To compare directly the results produced by the runs with $\epsilon = 10^{-10}$ and $\epsilon = 10^{-13}$ in Fig. 10 the distribution of b' are presented on the left-hand panel for both accuracy values. Apart from a random variation, the pdf curves are akin and fluctuate around the same mean while the cdf curves practically cannot be distinguished from each other. From the right-hand panel of Fig. 10, there is apparently a notable difference between the pdf curves of V_i' around 1, where the red dotted curve is about 0.05 higher than the solid black one indicating more frequent collision with lower velocity when the

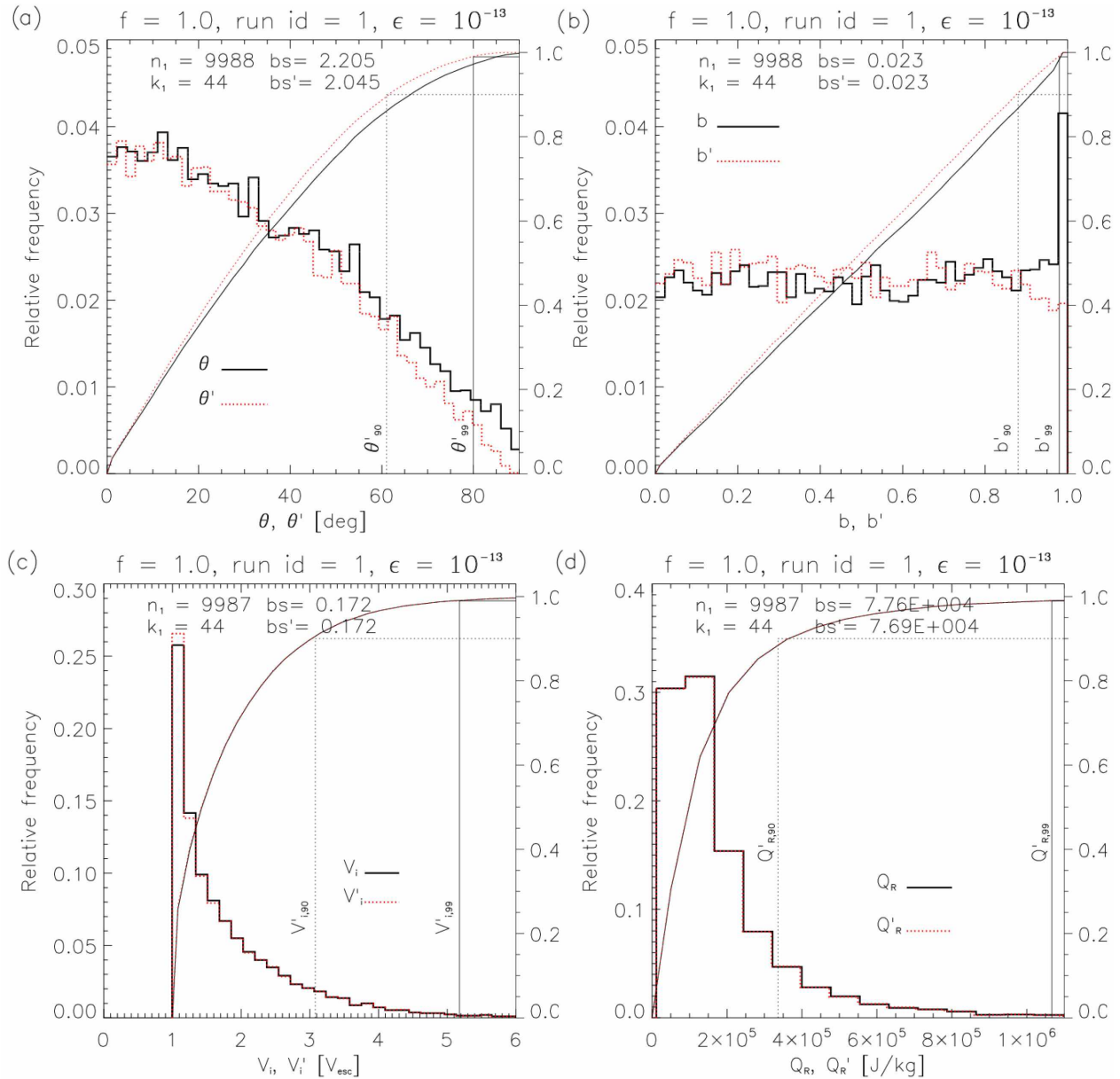


Figure 9. The same as Fig. 8 but for $\epsilon = 10^{-13}$.

accuracy is higher. Beyond 1 the shape of the two curves is very similar, although the dotted curve is slightly above the black curve, consequently there are more impact with lower velocity for the higher accuracy. The cdf curves are almost identical.

From these comparisons and analysis, it is evident that one should always correct the impact parameters as it was described in Section 3.3 and from a statistical point of view it is enough to follow the evolution of the system with a lower accuracy parameter if one is interested only in the collision statistical properties of the system. The only exception is the frequency of collisions around 1, where the lower accuracy simulations provide lower frequency.

3.6.2 Comparison of statistics derived from single runs

It is instructive to compare the results of the 10 different runs for a given f . In Fig. 11, the left-hand panel shows the pdfs and cdfs of b' for $f = 1$ for all the 10 runs. The pdf curves are akin and all

have a slight decrease at $b \geq 0.95$. The mean of the pdfs is 0.0222 which is shown by the thick horizontal line. Lets assume a uniform distribution between $[0, 1]$ with $k = 44$ bins then the mean of it is $1/k = 0.0227$ which is very close to the observed value. Except for a random variation around the mean all the pdfs are similar to each other. Furthermore, the cdf curves are essentially identical and have a slope of 1. These properties strongly indicate that the impact parameter has a uniform distribution within $[0, 0.95]$, for $b > 0.95$ the pdf drops off slightly.

As noted earlier the impact parameter has a notable influence on the collision outcome. For equal size bodies $\gamma = 1$ and $b_{\text{crit}} = 0.5$ which is shown by a solid thick black vertical line on the left-hand panel of Fig. 11, denoted by $\gamma_{1:1}$. According to the figure and make use of the result that b has a uniform distribution, the probability of a grazing impact is

$$P(\text{grazing}) = 1 - b_{\text{crit}}. \quad (18)$$

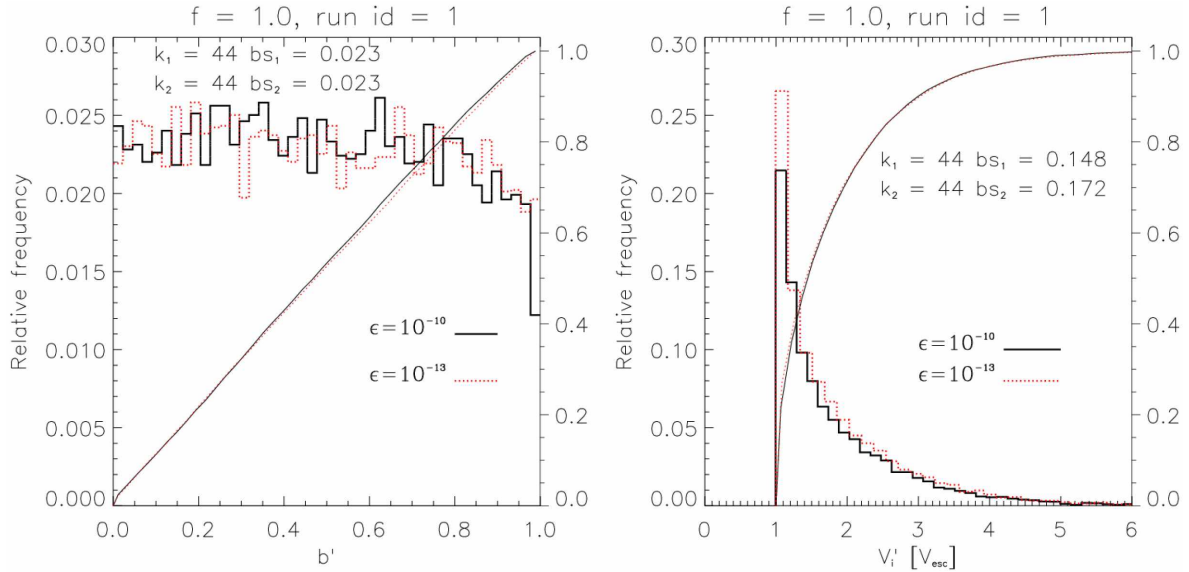


Figure 10. On the left-hand panel the distribution of b' , on the right-hand panel V_i' is compared for $\epsilon = 10^{-10}$ and $\epsilon = 10^{-13}$.

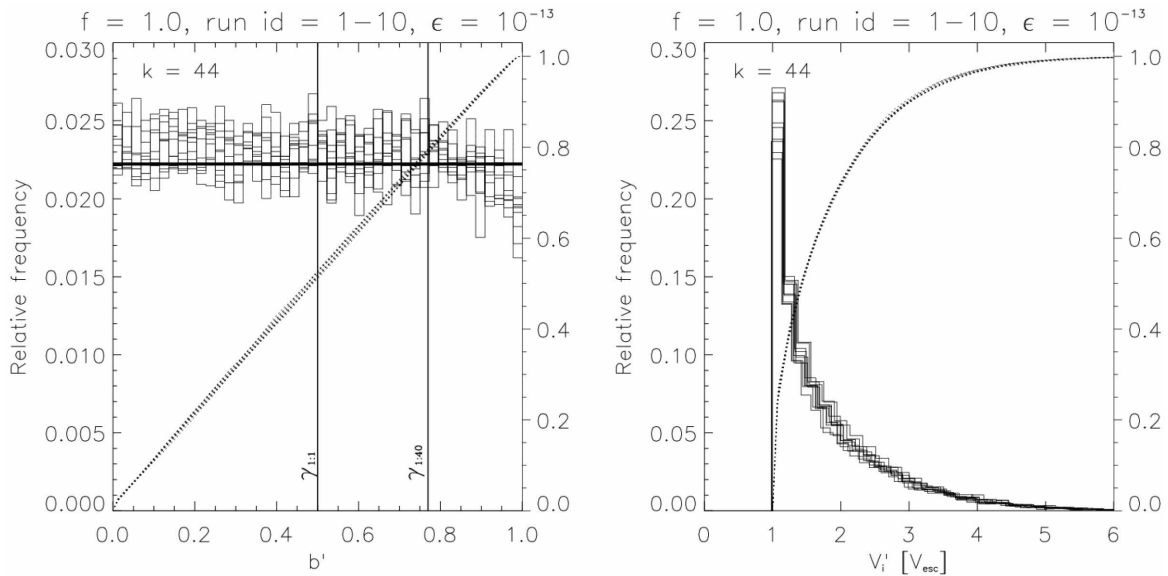


Figure 11. Summary of the results for $\epsilon = 10^{-13}$ for run = 1, ..., 10. On the left-hand panel, the distribution of b' is displayed. The thick black horizontal line is the mean of all the 10 pdfs. The vertical lines denote the threshold of grazing impacts between equal-sized bodies ($\gamma_{1:1}$) and between bodies with $\gamma = 1:40$. On the right-hand panel, the pdfs of V_i' are shown. For details see text.

Assuming equal density, the critical parameter can be expressed with γ as

$$b_{\text{crit}} = \frac{1}{1 + \gamma^{1/3}}, \quad (19)$$

so the probability of a grazing impact as a function of γ is

$$P(\text{grazing}) = \frac{\gamma^{1/3}}{1 + \gamma^{1/3}}, \quad (20)$$

which yields 50 per cent for equal size bodies. For e.g. $\gamma = 1:40$ the $b_{\text{crit}} = 0.77$ which is denoted by $\gamma_{1:40}$ on Fig. 11. This implies that less than 23 per cent of the collisions lead to grazing impact for bodies with a $\gamma \leq 0.025$. As the protoplanetary disc matures and larger and larger bodies emerge from the swarm γ may reach smaller values and the chance of a grazing impact reduces.

The right-hand panel of Fig. 11 shows the pdfs and cdfs of V_i' for the same 10 runs. There are no collisions with impact speed less than 0.993, the maxima of the pdfs are at ≈ 1.07 , beyond it the pdfs drop off quite steeply at first, but then more slowly and finally the trend levels off around 5. Evidently, there is a strong negative correlation, as V_i' gets larger than ≈ 1.07 the relative frequency of collisions drop off quite steeply. From the figure, it is obvious that the pdfs are very similar, while the cdfs look identical.

3.6.3 Dependence on f

In this section, the distribution of the collision parameters is compared for all the f values. The distribution of the impact parameters is displayed in Fig. 12 for $f = 1, 2, 3, 5, 10$ with $\epsilon = 10^{-13}$. Again, the pdf curves show a very similar behaviour

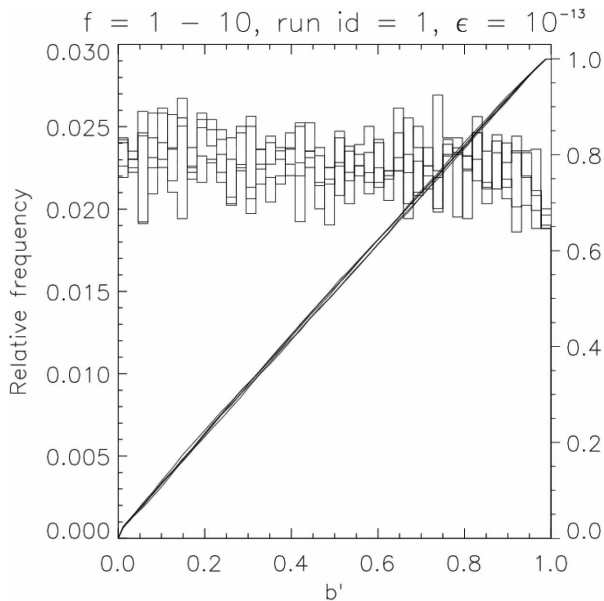


Figure 12. The pdf and cdf of b' for $\epsilon = 10^{-13}$ with run id = 1 for $f = 1, 2, 3, 5, 10$.

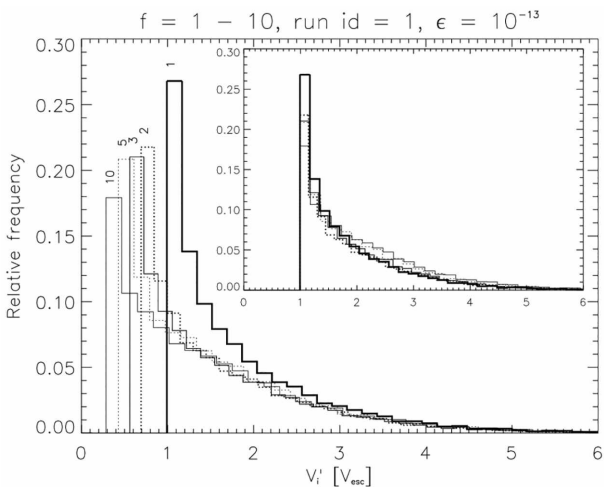


Figure 13. The same as Fig. 12 for V_i . The thick solid black curve shows the pdf of $f = 1$. The inset plot shows the shifted pdfs; see text for details.

while the cdf curves are practically identical. It is evident that the impact parameter does not depend on f , it has a uniform distribution.

The distribution of the impact velocity for $f = 1, \dots, 10$ is shown in Fig. 13 for $\epsilon = 10^{-13}$. Apparently the pdf of the impact velocity strongly depends on f . The domain of the pdf is equal to $[\min(V_i), \max(V_i)]$ (see Table A1) and $\min(V_i)$ shift towards lower values with increasing f , but the shape remains very similar as it is apparent from the inset plot, which was created by shifting all the pdfs right to the location of minimum belonging to $f = 1$ which is approximately 1. The lower boundary of the domain, $\min(V_i)$ and location of the pdfs' maxima which is denoted by $V_{i,m}$ are listed in Table 9.

Table 9. For each f , the minimum value of the impact speed (second column), the location of the pdfs' maxima (third column), the calculated impact speed V'_{rel} from equation (28; fourth column), and the minimum distance f_0 and d_0 from equation (30) are listed for $\epsilon = 10^{-13}$.

f	$\min(V_i)$	$V_{i,m}$	V'_{rel}	$\min(f_0)$	$\min(d_0)$ [au]
1	0.993	1.0801	1	71.68	2.40×10^{-4}
2	0.693	0.7686	0.707	50.63	1.69×10^{-4}
3	0.561	0.6435	0.577	53.73	1.80×10^{-4}
5	0.427	0.5181	0.447	56.59	1.89×10^{-4}
10	0.284	0.3754	0.316	51.69	1.73×10^{-4}

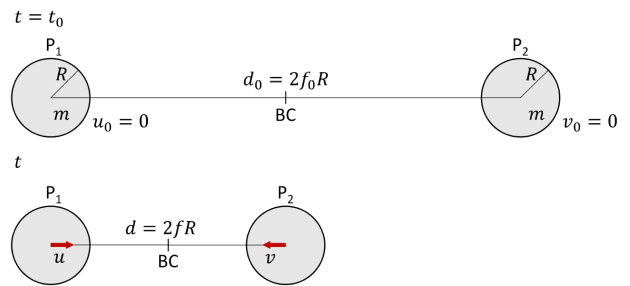


Figure 14. The two-body problem in the centre of mass reference frame. BC denotes the barycentre.

4 TWO-BODY APPROXIMATION TO MODEL THE IMPACT VELOCITY

The observed behaviour of the lower boundary of the domain and the location of pdfs' peak of the impact velocity can be explained by a simple model described below.

I have applied a straightforward physical model based on the two-body problem and used conservation laws to estimate the impact velocity as a function of the mutual distance d . In this model only the two colliding bodies P_1 and P_2 are considered and the effects of all others are neglected. In order to simplify the calculations, I have assumed that the colliding bodies have the same mass m and radius R . The model is depicted in Fig. 14 where the coordinate system is fixed to the barycentre of the system denoted by BC in the figure.

At time $t = t_0$ bodies P_1 and P_2 are d_0 apart from each other, and both bodies are in rest, $u_0 = v_0 = 0$. Let us calculate the velocity u and v of P_1 and P_2 at time t when their distance is d . I note that the relative velocity plays the role of V_i . If P_1 is the target then the impact speed is $u - v$.

From the conservation of the linear momentum, it follows that $u = -v$ and from the conservation of energy:

$$-k_G^2 \frac{m^2}{2f_0 R} = \frac{1}{2} m u^2 + \frac{1}{2} m v^2 - k_G^2 \frac{m^2}{2f R}, \quad (21)$$

where I have parametrized the distance with $f \geq 1$ such that $d = 2fR$. Substituting $v = -u$ into equation (21) one gets

$$u = k \sqrt{\frac{m}{2R}} \sqrt{\frac{f_0 - f}{f_0 f}}. \quad (22)$$

The escape velocity of P_2 with respect to P_1 is

$$V_e = k \sqrt{\frac{m}{R}}, \quad (23)$$

therefore u can be written as

$$u = \frac{1}{2} V_e \sqrt{\frac{f_0 - f}{f_0 f}}. \quad (24)$$

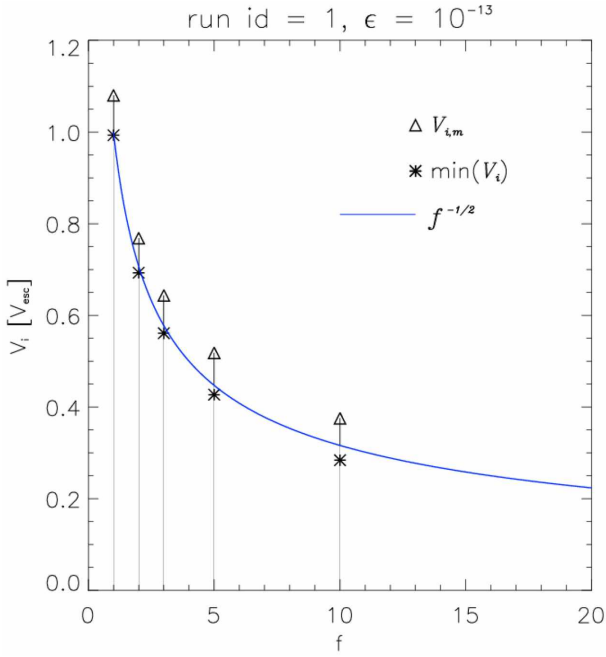


Figure 15. The lower boundary of the domain $\min(V_i)$ and the location of the maximum $V_{i,m}$ versus f shown by asterisk and triangle symbols, respectively. The solid blue curve is the graph of the function in equation (28).

The relative velocity V_{rel} of P_1 with respect to P_2 is

$$V_{\text{rel}} = u - v = V_e \sqrt{\frac{f_0 - f}{f_0 f}}, \quad (25)$$

and introducing $V'_{\text{rel}} = \frac{V_{\text{rel}}}{V_e}$ than equation (25) can be written as

$$V'_{\text{rel}} = \sqrt{\frac{f_0 - f}{f_0 f}}. \quad (26)$$

I remark that V'_{rel} is measured in escape velocity unit and plays the role of the impact speed in the preceding text. If one knows the initial distance d_0 and thus f_0 than the velocity V'_{rel} at a distance of d can be computed from equation (26). If $f_0 \gg f$ then $f_0 - f \approx f_0$, therefore

$$\frac{f_0 - f}{f_0 f} \approx \frac{1}{f}, \quad (27)$$

and then equation (26) becomes

$$V'_{\text{rel}} \approx f^{-1/2}. \quad (28)$$

This result gives an approximation of the minimum of the impact speed. Real physical impacts happen when $f=1$, substituting this into equation (28) results $V'_{\text{rel}} \approx 1$, i.e. the impact speed is approximately 1 escape velocity unit. This is exactly in line with the distribution of the impact velocities (see e.g. Figs 8 and 9, and the right-hand panel of Fig. 11 and explains the lack of impact velocity less than about 1 for $f=1$.

In Fig. 15, the blue curve is the graph of equation (28). The $\min(V_i)$ data points match very well to this curve so this model explains nicely the shift observed on Fig. 13. From Fig. 15, one can see that the minimum values are just below the blue curve, which is a natural consequence of that some of the bodies are close to each other, i.e. the assumption $f_0 \gg f$ breaks down. Substituting $f_0 > f = 1$ into

equation (26), one gets

$$V'_{\text{rel}} = \sqrt{\frac{f_0 - 1}{f_0}} < 1. \quad (29)$$

Using the formula of equation (28) for the different f values, the results are listed in the fourth column of Table 9.

The initial distance characterized by f_0 can be expressed from equation (26)

$$f_0 = \frac{f}{1 - f V_{\text{rel}}^2}. \quad (30)$$

Substituting the minimum of the impact speed into V'_{rel} from the second column of Table 9 into equation (30) then the initial distance of the two bodies can be estimated and the resulting f_0 is displayed in the fifth column of Table 9.

5 COLLISION OUTCOME MAPS

Recent works (LS12; Genda et al. 2012) based on the combination of hydrocode and N -body gravity code computations studied the impacts between planetary mass bodies and the outcome of planetary collisions have been parameterized in terms of the masses and velocities of the colliding bodies. According to these calculations, the authors devised formulae for the mass of the largest remnant denoted by M_{lr} produced in a collision as a function of the masses of the bodies involved, the impact velocity and impact angle. The authors also identified the boundaries between different types of collision:

- (i) perfect merging ($M_{\text{lr}} = M_{\text{tot}}$)
- (ii) partial accretion with some mass escaping as fragments ($M_{\text{lr}} < M_{\text{tot}}$)
- (iii) partial erosion of the target ($M_{\text{lr}} < m_t$)
- (iv) pure hit-and-run ($M_{\text{lr}} = m_t$) and erosive hit-and-run (lead to some erosion of the target and more significant damage of the projectile.)

where $M_{\text{tot}} = m_t + m_p$.

Using the analytical model of LS12 with adopted values of $\bar{\mu} = 0.36$ and $c^* = 1.9$, I derived example collision outcome maps. I present an analysis of the simulations to assess the range of true collision outcomes. For this purpose, two colour-coded collision outcome maps were calculated that are shown in Fig. 16 for mass ratios of $\gamma = 1$ (left-hand panel) and $\gamma = 1:2$ (right-hand panel). In the figure, dark blue denotes perfect merging, light blue partial accretion, white colour net erosion to the target, and finally green hit-and-run region. In every case, the simulation data were used to calculate the collision parameters and to determine the outcome based on the LS12 model. The vertical red line denotes the onset of hit-and-run events at b_{crit} .

The details of the calculation of the collision regimes are given in the Appendix of LS12. The model assumes a sudden transition between grazing and non-grazing impacts, which is of course, artificial. In Fig. 16, the thick red curve corresponds to the critical velocity for catastrophic disruption, where the largest remnant contains half the total mass, $M_{\text{lr}} = 0.5M_{\text{tot}}$. Note that this curve corresponds to the target erosion boundary for $\gamma = 1$ cases (the transition from partial accretion or hit-and-run to the erosion region). The lower red dot-dashed curves correspond to the impact velocity needed to disperse 10 per cent and the upper red dashed curve to 90 per cent of m_t .

Fig. 16 also shows details of all the collisions that occurred in the simulations for $f = 1$ and $\epsilon = 10^{-13}$ for a single run data set,

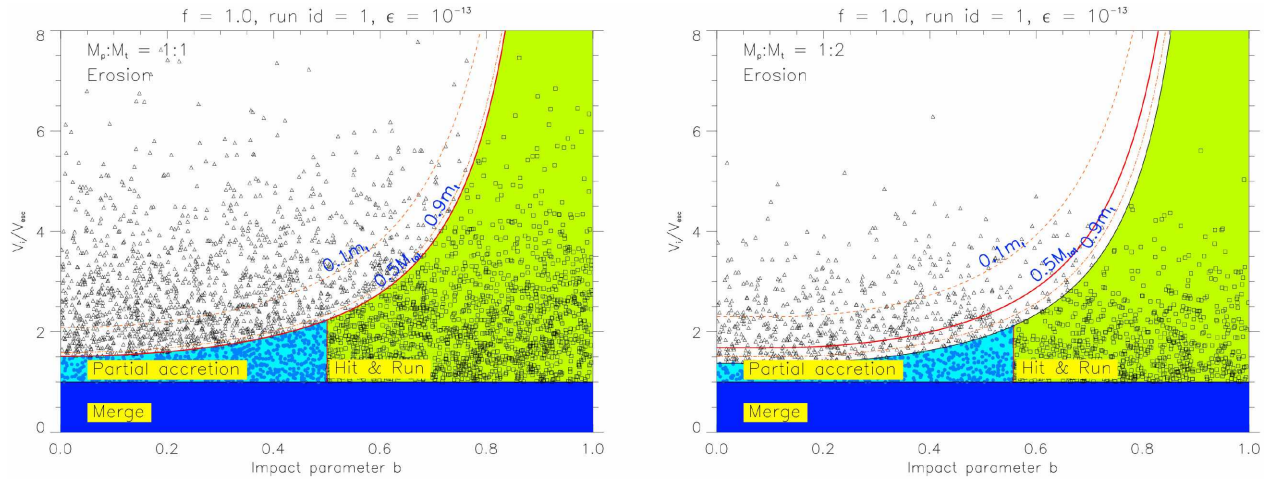


Figure 16. Predicted collision outcome maps using the analytical model of LS12 for strengthless planets ($\bar{\mu} = 0.36$ and $c^* = 1.9$). Impact velocity is normalized by the mutual surface escape velocity assuming a bulk density of 2000 kg m^{-3} . Coloured regions denote perfect merging (dark blue), partial accretion (light blue), net erosion to the target (white), and hit-and-run (green). The vertical red line denotes the onset of hit-and-run events at b_{crit} . Thick black red denotes the critical disruption velocity for half the total mass ($0.5M_{\text{tot}}$) remaining; red dashed curve denotes 90 per cent ($0.9m_t$) and 10 per cent ($0.1m_t$) of target mass in largest remnant. The different symbols indicate the outcome of each collision: diamonds for mergers, squares for hit-and-run collisions, filled circles for collisions in which the target gained mass, and triangles for collisions in which the target lost mass.

Table 10. Predicted collision outcome ratios in percentages for different γ with $f = 1$ and $\epsilon = 10^{-13}$. The total number of listed cases is 67 716 representing 68 per cent of the sample.

	$\gamma = 1$	$\gamma = \frac{1}{2}$	$\gamma = \frac{1}{3}$	$\gamma = \frac{1}{4}$	$\gamma = \frac{1}{10}$
n	36 656	18 564	7866	4088	542
n_m/n	0.05	0.11	0.22	0.24	0.55
n_{pa}/n	17.16	24.21	32.43	37.77	63.47
n_e/n	39.79	35.35	27.77	24.00	6.27
n_{hr}/n	43.00	40.33	39.59	37.99	29.70
n_{e1}/n	39.79	23.43	10.39	5.04	0.37
n_{e2}/n	37.39	29.83	20.21	14.36	0.92
n_{e3}/n	23.29	9.87	2.81	0.76	0.00

run id = 1 with each symbol representing a single collision with $\gamma = 1$ (left) and $\gamma = \frac{1}{2}$ (right). The symbols indicate the type of collision involved: diamonds for mergers, squares for hit-and-run collisions where the target survives intact, circles for collisions that increased the mass of the target body (possibly with some mass from the projectile escaping as fragments), and triangles for collisions that eroded mass from the target. From the figure, it is apparent that there is no correlation between V_i and b . The number of the different type collisions are summarized Table 10.

In Table 10, the total number of collisions n and the ratios of the different types of outcomes are given for five γ values. The number of perfect merging, partial accretion, erosion, and hit-and-run events are denoted by n_m , n_{pa} , n_e , and n_{hr} , respectively. In the table, the subscript e1 denotes collisions where $M_{\text{lr}} \leq 0.5M_{\text{tot}}$, i.e. the catastrophic disruption, e2 denotes collisions where $M_{\text{lr}} \leq 0.9m_t$ and e3 denotes collisions where $M_{\text{lr}} \leq 0.1m_t$, i.e. the supercatastrophic collision. The number of merging collisions is very low, less than 1 per cent for all listed γ values.

According to the table, there is a clear trend which is visualized in Fig. 17, where I have plotted the number of the different type collisions against γ .

The initial single mass distribution relaxes into a continuous power-law mass distribution in ~ 100 yr which is in line with the

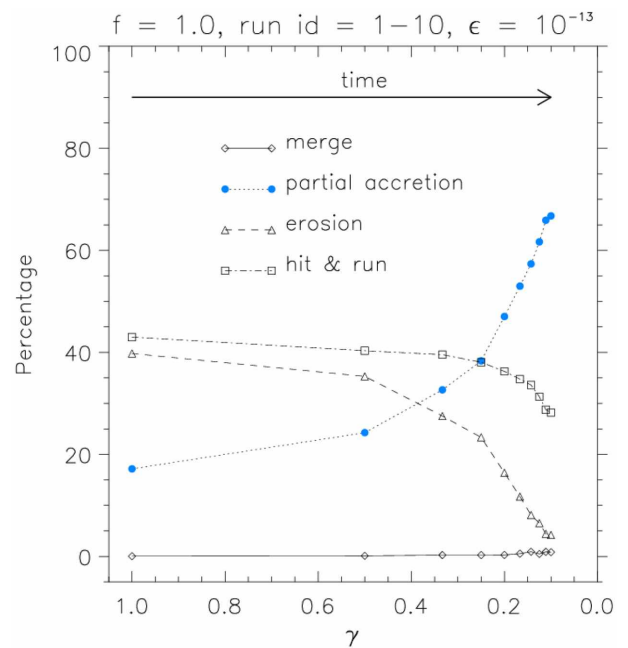


Figure 17. The frequency of different types of collision as a function of γ . Note that the horizontal axis is reversed in order to reflect the progress of time, as it is shown by the upper arrow. The symbols indicate the type of collision: diamonds for mergers, blue circles for partial accretion, triangles for collisions that eroded mass from the target, and squares for hit-and-run. The collisions were collected from multiple run data with $f = 1$ and $\epsilon = 10^{-13}$.

results of Kokubo & Ida (1996). As the simulation time proceeds the largest body of the continuous mass distribution separates from it. In the beginning of the simulation $\gamma = 1$ and as the continuous power-law mass distribution develops the decreasing minimum of γ reflects the time evolution of the system, at least early in the simulation. Later on this relationship breaks and all we can say is that the minimum value of γ is indicative of the minimum age of the

system. In order to reflect this initial relationship between time and $\min(\gamma)$, the horizontal axis γ is reversed and the time is shown by the black arrow in the top of the figure.

It is clearly shown in Fig. 17 that the number of merging collisions remains very low, below 1 per cent and shows a very modest increase as γ decreases. As time proceeds and the bodies grow via collisions γ may reach smaller values therefore in later times the chance of a merging collision slightly increases. The number of partial accretion depends strongly on γ . As $\min(\gamma)$ decreases the chance of partial accretion events increases and for $\gamma \leq 0.5$ the increase speeds up such that the proportion of partial accretion reaches ≈ 70 per cent for $\gamma = 0.1$. On the other hand, the number of erosion declines with decreasing γ just as n_{hr} which follows a similar pattern but it decreases with a smaller pace. If the collision rate is maintained by gravitational focusing and/or additional material is coming from outer space via migration than γ may reach smaller values and the chance of a grazing impact reduces, see equation (20). For $\gamma < 0.2$ partial accretion dominates and consequently as the disc matures and time increases the larger bodies grow faster, i.e. the rich get richer.

The importance of hit-and-run collisions was recognized by Asphaug, Agnor & Williams (2006), namely in these events, a pair of objects collides at an oblique angle. Momentum causes the bodies to graze each other while exchanging some mantel material and afterwards they separate again. Generally, the final bodies have similar masses to the original pair but smaller relative velocities. Hit-and-run collisions can be further divided into two subclasses. When the first collision speed is only slightly larger than the mutual escape velocity, the bodies collide, separate but in a second collision they merge. These are the so-called graze-and-merge collisions. On the other hand, when the initial impact velocity is higher, the two bodies separate with a relative speed larger than the escape velocity, which is not followed by a second impact. These are true hit-and-run collisions. The boundary between these two regimes has been clearly identified in a second series of impact simulations carried out by Genda et al. (2012).

According to the definition of Kokubo & Ida (1996), runaway growth means that the largest body in its feeding zone grows more rapidly than the second largest one and the ratio of the mass of the largest body M_{max} and the mean mass $\langle m \rangle$ grows monotonically with time. In their 3D calculations, they presented evidence of runaway growth and the results in this work summarized in Fig. 17 also support this observation. As the single mass distribution relaxes into the continuous power-law mass distribution more and more collisions may involve a larger and smaller body, i.e. $\gamma \leq 0.2$ hence partial accretion dominates therefore the larger the body is the faster it grows. The numerical simulations of Aarseth, Lin & Palmer (1993) showed that runaway coagulation occurs and can lead to the formation of protoplanetary cores. In their 2D simulations in the case of larger eccentricities (hot accretion), the degree of runaway growth can be the same as in 3D. In the 3D simulations of Kokubo & Ida (1996), the ratio $M_{\text{max}}/\langle m \rangle \approx 140$ at $t = 20\,000$ yr, in their 2D one the ratio ≈ 15 , while in the work of Aarseth et al. (1993) it was ≈ 13 . Using the data of this work, preliminary calculations show that $M_{\text{max}}/\langle m \rangle \approx 10$ at $t = 20\,000$ yr. It must be noted that the above mentioned works cannot be directly compared because on one hand of the significant differences between the treatment of the gravitational forces and the collisions and on the other hand the different initial conditions. However, the general conclusions and trends are similar. In this work, the main focus is on the statistics of the collision parameters and hence growth is studied from the collision point of view, therefore these results are valid for both 2D and 3D cases. The issue about the orderly and runaway growth as

well as the mass distribution of the bodies and its time evolution will be investigated in a next paper.

5.1 Comparison with previous works

A direct comparison to previous works is hard to present because of the different simulation setups. The most important is that the present simulations were performed in 2D while the others in 3D. The initial conditions, the f expansion factor of the radii differs from work to work and the values in the tables or figures are given for different γ or γ is not exactly specified in the papers.

To estimate the effects of including the collision model of LS12 on planet formation, Stewart & Leinhardt (2012) analysed the impact parameters of previous N -body simulations that used only perfect accretion (O'Brien et al. 2006; Raymond et al. 2009). The authors divided the collision outcomes from the simulations into three groups: group 1 consists of all collisions from eight different simulations in O'Brien et al. (2006), group 2 includes all collisions from 40 simulations by Raymond et al. (2009) while group 3 contains only giant impacts from the latter simulations. Groups 1 and 2 were subdivided into two subclass containing collisions between embryos and planetesimals ($\gamma \leq \frac{1}{40}$) and between only embryos ($\gamma \geq \frac{1}{10}$). The predicted collision outcome statistics from table 1 of Stewart & Leinhardt (2012) is summarized in the columns from 5 to 9 in Table 11.

The simulations of Chambers (2013) also include the LS12 model to take care of fragmentation and hit-and-run collisions. The collision statistics of these calculations using table 1 and figs 5 and 6 of Chambers (2013) are reported in the tenth and eleventh columns of Table 11.

In a recent work, Bonsor et al. (2015) performed N -body simulations to track the change in the bulk Mg/Fe and Si/Fe ratios of the protoplanets. The N -body code used the LS12 model to realistically model collisions. However, the paper did not present the collision statistics in a tabular form, this kind of data could be extracted from fig. 4 of the paper. The approximate values are presented in the last 2 columns of Table 11. Similar runs were performed by Carter et al. (2015) and the collision statistics derived from fig. 5 of the paper is qualitatively the same as in Bonsor et al. (2015).

The impacts for multiple run data with $f = 1$ and $\epsilon = 10^{-13}$ of this work is presented in the first three columns of Table 11, where the second column list data for all γ , third column for giant impacts with $\gamma \geq \frac{1}{10}$ and the third column shows for planetesimal impacts with $\gamma \leq \frac{1}{40}$. These specific values were selected since the initial γ between the planetesimals and embryos were similar in Stewart & Leinhardt (2012).

Comparing the giant impacts of this work (third column of Table 11) with that of Stewart & Leinhardt (2012; fifth–seventh columns), one sees similarities and significant differences. The perfect merge ratios roughly agree in all cases except in group 1 where 0 merge event was detected, but the sample size in that group is at least an order of magnitude smaller than in the other ones. The hit-and-run case represents 29.9 per cent of all events in this work and this value agrees well with that of group 1 (31.3 per cent), 2 (27.7 per cent), and 3 (28.2 per cent). The partial accretion measured in this work is typically larger by a factor of ≈ 1.5 , the best agreement is with group 2: 44.7 per cent versus 39.2 per cent. Significant differences were found in the case of erosion: this study found 24.9 per cent while this quantity is less than 3 per cent in all the groups. The reason for this large difference is unclear since the impact velocity distribution measured in this study (see Figs 10, 11, 12, and 15) and those presented in Stewart & Leinhardt (2012; figs 2 and

Table 11. Predicted collision outcome ratios calculated from the work of Stewart & Leinhardt (2012), Chambers (2013), Bonsor et al. (2015), and from the present simulations. The number of graze-and-merge events is denoted by n_{gm} .

	This work			Stewart & Leinhardt (2012) [table 1]				Chambers (2013) [table 1, figs 5 and 6]		Bonsor et al. (2015) [figs 3 and 4]		
	All $\gamma \in (0, 1]$	Giant $\gamma \geq \frac{1}{10}$	Planet. $\gamma \leq \frac{1}{40}$	Giant $\gamma \geq \frac{1}{10}$	Planetesimal $\gamma \leq \frac{1}{40}$		all coll. (1 sim)	all coll. (em - em)	$t = 0$ all coll.	$t = 6 \times 10^5$ all coll.		
n	99899	93330	1530	67	544	1165	1140	3142	≈ 386	92	–	–
n_m/n	0.32	0.3	0.9	0.0	0.7	0.6	0.0	0.6	3.6	25	6	5
n_{pa}/n	29.82	44.7	76.8	26.9	39.2	36.1	72.6	69.4	32.4	37	41	43
n_e/n	24.14	24.9	0.1	3.0	0.6	1.3	0.0	1.9	19.7	0	3	16
n_{hr}/n	45.72	29.9	22.2	31.3	27.7	28.2	23.6	25.4	44.3	38	50	36
n_{gm}/n	–	–	–	38.8	31.8	33.8	3.8	2.7	–	–	–	–
n_{e3}/n	12.07	0.2	0	1.5	0	0.3	0.0	0.0	–	–	1	10

Note. In the work of Chambers (2013), the n_m/n contains also the graze-and-merge events.

10) show good agreement. This large discrepancy deserves careful further studies since erosion plays a crucial role in the formation and the final bulk composition of the terrestrial planets. Notwithstanding, the special case of erosion the supercatastrophic events ratio of this work (0.2 per cent) matches quite well with the value in group 3 (0.3 per cent).

Analysing the planetesimal collisions detected in this work (see the fourth column) and that of Stewart & Leinhardt (2012; eighth, ninth columns), a very good agreement can be observed: partial accretion 76.8 per cent versus 72.6 per cent and 69.4 per cent, hit-and-run events 22.2 per cent versus 23.6 per cent and 25.4 per cent and there was no supercatastrophic event in all three simulations.

Comparing all impacts of this work (second column) with that of Chambers (2013; tenth column), one sees that in overall the probabilities of different collision outcomes are similar. The results for partial accretion and erosion should be highlighted, 29.8 per cent versus 32.4 per cent and 24.1 per cent versus 19.7 per cent, respectively. These present a far better agreement than in the previous comparison. Also the hit-and-run events match closely: 45.7 per cent versus 44.3 per cent. On the other hand, the comparison of giant impacts (third column versus eleventh) shows significant differences: the perfect accretion calculated in this work is 0.3 per cent much less than the 25 per cent reported by Chambers (2013), the erosion in this work is 24.9 per cent versus 0 per cent. The values for partial accretion and hit-and-run broadly agree. It must be noted that in the work of Chambers (2013) γ was not specified, these values were derived from the embryo–embryo collisions taken from the last row of table 1 in Chambers (2013). Since the collision outcome depends strongly on γ , these values are only approximate and therefore they are only indicative.

The last 2 columns contain the collision statistics compiled from figs 4 and 5 of Bonsor et al. (2015) both containing collision events with $\gamma \in (0, 1]$. The $t = 0$ means the beginning of the simulation while $t = 6 \times 10^5$ is the end of it. These values should be compared with the second column and one sees that in this work, perfect merging is about 20 times less than in Bonsor et al. (2015). The partial accretion is almost equal, while the hit-and-run ratio of this work is 45.7 per cent which is in between the values of 50 per cent and 36 per cent at $t = 0$ and $t = 6 \times 10^5$, respectively. Again the erosion is largely different: at $t = 0$ it is 24 per cent versus 3 per cent, but this difference decreases and the erosion reaches 16 per cent at the end of the simulation.

Comparing the works enumerated in Table 11 with each other, the differences are large and the results vary in large intervals. On one hand, this is a natural consequence of the different simulation

setup, applied numerical integrator method and expansion factor but on the other hand, this issue has to study further in detail; especially the cause behind the large difference in the measured erosion and perfect accretion.

I note that the model of LS12 applies a different definition for the escape velocity:

$$V'_{\text{esc}} = k_G \sqrt{2M'/R'}, \quad (31)$$

where $M' = m_t + m_{\text{interact}}$. The m_{interact} denotes the interacting mass of the projectile estimated to be involved in the collision. Since $m_{\text{interact}} \leq m_p$ therefore it may be a contributing factor in the difference of perfect mergers in Bonsor et al. (2015) and Chambers (2013).

6 CONCLUSIONS

I have performed 2D N -body simulations of planetary accretion. The statistics of the collision parameters is investigated with 10 000 equal-mass protoplanets under perfect and gas-free accretion. In this paper, the detailed statistics of the collision parameters are presented along with a simple method to improve the collision parameters. Using the two-body problem and the conservation laws, I explained the main features of the observed impact velocity distributions. The collision outcome maps are shown for specific projectile-to-target mass ratios. Combining the results of the simulations with the model developed by LS12 estimates for the different type of collisions are given as well as detailed comparisons with previous works on collision outcomes and frequencies.

The main conclusions of this study are:

(i) Using the two-body approximation for the colliding bodies a simple method was presented to improve the impact parameter b to b' by equation (14) and the impact velocity V_i to V'_i by equation (15). Using V'_i the improved specific energy Q'_R can be computed.

(ii) It is apparent from Figs 8 and 9 panel (b) that the pdf curves of b and b' are similar and from a statistical point of view they do not deviate, i.e. the improvement of the impact parameter is not too remarkable in their pdfs. Similarly the pdfs of the improved impact velocity from panel (c) of the same figures show little difference, although the largest discrepancy is for the most important interval. These findings are also supported by Fig. 10, where the results are directly compared for $\epsilon = 10^{-10}$ and $\epsilon = 10^{-13}$. I emphasize that the correction is important for the individual cases and it is obvious that the improvement of b and V_i is essential and all future simulations that use the model of LS12 (or similar) should incorporate it to provide more reliable results.

(iii) It was shown that in 2D the impact parameter has a uniform distribution within $[0, 0.95]$, and for $b > 0.95$ the pdf drops off slightly.

(iv) According to Fig. 11, the different runs produce the same distribution of b , V_i , and Q_R therefore it is enough to do only one run to obtain creditable statistical data.

(v) Making use of the two-body approximation it was possible to explain the hiatus of impact speed less than $1 V_{\text{esc}}$ for $f = 1$ which is a consequence of the conservation laws. The derived formula gives an excellent approximation of the minimum impact speed for different f .

(vi) It was shown that the impact parameter b does not depend on f as it is presented by Fig. 12. On the other hand, the distribution of V_i is a strong function of f , the pdf of the impact speed shifts towards zero as f increases, see Fig. 13. This behaviour was explained by the two-body approximation shown in Fig. 15.

(vii) The model of LS12 was used to determine the number of different types of collisions, see Fig. 16 and Table 10. For equal-mass bodies, the majority of collisions (43 per cent) are hit-and-run events and erosion (40 per cent). Partial accretion constitutes 17 per cent of all cases and 0.05 per cent mergers were measured. The frequencies of different events are summarized in Fig 17.

(viii) A thorough comparison with previous works was presented and promising similarities along with significant differences were found. The most significant difference is in the case of erosion, where this work reports 25 per cent, while all the others are much smaller, less than 3 per cent except for one case of Chambers (2013) where it reaches approximately 20 per cent. The ratio of perfect merge roughly agree with those of Stewart & Leinhardt (2012) but differs largely from the results of Chambers (2013) and Bonsor et al. (2015). The hit-and-run values match reasonably, while the partial accretion ratios broadly agree with other values. It was also shown that serious differences are present among the previous works.

(ix) The proportion of partial accretion increasing more and more steeply as $\min(\gamma)$ decreases, i.e. as the time increases, see Fig. 17. For $\gamma \leq 0.1$, the majority of collisions (≈ 70 per cent) is partial accretion and 25 per cent is hit-and-run events (comparable values were found by Stewart & Leinhardt 2012). A significant fraction of the latter events are graze-and-merge collisions that also leads to accretion. These results provide a independent evidence for the runaway growth mode described in detail by Kokubo & Ida (1996). The implications for planet formation are that the larger the difference between the masses of the objects (γ is small) the larger the probability for collisions in which M_{ir} grows. This can further decrease $\min(\gamma)$ and thus induce a positive feedback favouring more intensive mass growth.

Below I summarize some technically important findings:

(i) The run-times for the two accuracy parameters were compared and it was shown that the simulations for $\epsilon = 10^{-13}$ takes about 40 per cent more time to complete than those with $\epsilon = 10^{-10}$.

(ii) It was shown that the scaled mean run-time is inversely proportional to f for both accuracy parameters. The basic cause of this is the specific decrease of the number of bodies.

(iii) Using the model described in Section 3.3, the parameters of the collision computed from the simulations can be improved. Comparing the improved values, it turned out that there are not too significant difference between the distributions, i.e. from a statistical point of view the lower accuracy simulations already provide useful statistical data.

ACKNOWLEDGEMENTS

This work was partly supported by the ÚNKP-19-4 New National Excellence Program of the Ministry for Innovation and Technology, partly by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. I acknowledge the support of the Hungarian National Research, Development, and Innovation Office (NKFIH), under the grant K- 119993 and the joint Stiftung Aktion Österreich-Ungarn program through project 95öu10. I would like to thank the support of NVIDIA Corporation with the donation of a Tesla C2075 and K40 GPUs. The N -body calculations were run using the National Information Infrastructure Development supercomputer facility.

DATA AVAILABILITY

The data underlying this paper will be shared on reasonable request to the authors.

REFERENCES

- Aarseth S. J., Lin D. N. C., Palmer P. L., 1993, *ApJ*, 403, 351
 Alexander S. G., Agnor C. B., 1998, *Icarus*, 132, 113
 Asphaug E., 2010, *Chem. Erde/Geochem.*, 70, 199
 Asphaug E., Agnor C. B., Williams Q., 2006, *Nature*, 439, 155
 Beaugé C., Aarseth S. J., 1990, *MNRAS*, 245, 30
 Bonsor A., Leinhardt Z. M., Carter P. J., Elliott T., Walter M. J., Stewart S. T., 2015, *Icarus*, 247, 291
 Carter P. J., Leinhardt Z. M., Elliott T., Walter M. J., Stewart S. T., 2015, *ApJ*, 813, 72
 Chambers J. E., 2001, *Icarus*, 152, 205
 Chambers J. E., 2004, *Earth Planet. Sci. Lett.*, 223, 241
 Chambers J. E., 2013, *Icarus*, 224, 43
 Chambers J. E., Wetherill G. W., 1998, *Icarus*, 136, 304
 Cox L. P., Lewis J. S., 1980, *Icarus*, 44, 706
 Cuzzi J. N., Hogan R. C., Bottke W. F., 2010, *Icarus*, 208, 518
 Dullemond C. P., Dominik C., 2005, *A&A*, 434, 971
 Fehlbeg E., 1968, NASA Technical Report, 287
 Fischer R. A., Ciesla F. J., 2014, *Earth Planet. Sci. Lett.*, 392, 28
 Genda H., Kokubo E., Ida S., 2012, *ApJ*, 744, 8
 Guilera O. M., Sándor Zs., 2017, *A&A*, 604, A10
 Hayashi C., 1981, *Prog. Theor. Phys. Suppl.*, 70, 35
 Ida S., Makino J., 1993, *Icarus*, 106, 210
 Johansen A., Youdin A., Mac Low M.-M., 2009, *ApJ*, 704, L75
 Kokubo E., Genda H., 2010, *ApJ*, 714, L21
 Kokubo E., Ida S., 1996, *Icarus*, 123, 180
 Kokubo E., Ida S., 1998, *Icarus*, 131, 171
 Laibe G., Gonzalez J. F., Maddison S. T., 2012, *A&A*, 537, A61
 Lecar M., Aarseth S. J., 1986, *ApJ*, 305, 564
 Leinhardt Z. M., Stewart S. T., 2012, *ApJ*, 745, 79 (LS12)
 Lines S., Leinhardt Z. M., Paardekooper S., Baruteau C., Thebault P., 2014, *ApJ*, 782, L11
 Lissauer J. J., 1993, *ARA&A*, 31, 129
 Liu H., Zhou J.-L., Wang S., 2011, *ApJ*, 732, 66
 Lyra W., Johansen A., Klahr H., Piskunov N., 2008, *A&A*, 491, L41
 Morbidelli A., Lunine J. I., O'Brien D. P., Raymond S. N., Walsh K. J., 2012, *Annu. Rev. Earth Planet. Sci.*, 40, 251
 O'Brien D. P., Morbidelli A., Levison H. F., 2006, *Icarus*, 184, 39
 Quintana E. V., Lissauer J. J., 2006, *Icarus*, 185, 1
 Quintana E. V., Lissauer J. J., 2014, *ApJ*, 786, 33
 Quintana E. V., Lissauer J. J., Chambers J. E., Duncan M. J., 2002, *ApJ*, 576, 982
 Quintana E. V., Barclay T., Borucki W. J., Rowe J. F., Chambers J. E., 2016, *ApJ*, 821, 126
 Raymond S. N., Quinn T., Lunine J. I., 2004, *Icarus*, 168, 1
 Raymond S. N., Quinn T., Lunine J. I., 2006, *Icarus*, 183, 265

Raymond S. N., O'Brien D. P., Morbidelli A., Kaib N. A., 2009, *Icarus*, 203, 644
 Sándor, Zs., Lyra W., Dullemond C. P., 2011, *ApJL*, 728, L9
 Shoemaker E. M., 1962, in Kopal Z., ed., *Physics and Astronomy of the Moon*. Academic Press, New York, p. 283
 Stewart S. T., Leinhardt Z. M., 2012, *ApJ*, 751, 32
 Weidenschilling S. J., 1977, *MNRAS*, 180, 57

Weidenschilling S. J., 1997, *Icarus*, 127, 290
 Wetherill G. W., 1994, *GeCoA*, 58, 4513
 Wetherill G. W., Stewart G. R., 1989, *Icarus*, 77, 330

APPENDIX A: SIMULATION SUMMARY

Table A1. Summary of simulation outcomes and results. Each row corresponds to one single run data, while the run with 1–10 to multiple run data and gives the range within which the sample values fall for $\epsilon = 10^{-10}$ and $\epsilon = 10^{-13}$, n is the number of collision.

f	run	n	$\epsilon = 10^{-10}$				n	$\epsilon = 10^{-13}$			
			$\min(V'_i)$ [V_{esc}]	$\max(V'_i)$ [V_{esc}]	$\min(Q'_R)$ [J kg^{-1}]	$\max(Q'_R)$ [J kg^{-1}]		$\min(V'_i)$ [V_{esc}]	$\max(V'_i)$ [V_{esc}]	$\min(Q'_R)$ [J kg^{-1}]	$\max(Q'_R)$ [J kg^{-1}]
1	1	9991	0.993	10.357	1.25×10^4	4.34×10^6	9988	0.993	8.661	1.11×10^4	3.58×10^6
1	2	9993	0.995	9.803	1.24×10^4	5.02×10^6	9988	0.991	7.950	1.25×10^4	3.72×10^6
1	3	9993	0.993	8.313	1.52×10^4	5.13×10^6	9991	0.991	7.552	1.35×10^4	4.18×10^6
1	4	9990	0.978	8.624	1.28×10^4	4.50×10^6	9992	0.978	8.417	1.24×10^4	5.69×10^6
1	5	9991	0.993	8.184	1.23×10^4	4.05×10^6	9990	0.991	9.030	1.09×10^4	3.52×10^6
1	6	9991	0.994	19.517	1.64×10^4	1.31×10^7	9991	0.993	8.018	1.48×10^4	5.23×10^6
1	6	9990		8.299		4.11×10^6					
1	7	9988	0.988	8.222	1.23×10^4	5.99×10^6	9990	0.991	8.374	1.66×10^4	3.46×10^6
1	8	9990	0.992	8.017	1.22×10^4	3.67×10^6	9990	0.990	7.264	1.26×10^4	4.83×10^6
1	9	9989	0.985	9.190	1.18×10^4	3.82×10^6	9989	0.992	8.609	1.19×10^4	3.93×10^6
1	10	9990	0.995	8.111	1.33×10^4	3.51×10^6	9990	0.992	9.346	1.17×10^4	6.83×10^6
1	1–10	99906	0.978	19.517	1.18×10^4	1.31×10^7	99899	0.978	9.346	1.09×10^4	6.83×10^6
1	1–10	99905		10.357		5.99×10^6					
2	1	9989	0.700	7.627	7.75×10^3	2.07×10^6	9991	0.693	7.338	7.44×10^3	4.33×10^6
2	2	9990	0.697	8.921	7.96×10^3	3.28×10^6	9989	0.697	9.138	6.44×10^3	2.86×10^6
2	3	9991	0.677	9.099	8.90×10^3	2.84×10^6	9987	0.696	7.492	7.43×10^3	2.03×10^6
2	4	9990	0.691	7.795	8.37×10^3	2.38×10^6	9989	0.674	9.653	6.71×10^3	3.20×10^6
2	5	9991	0.693	7.953	6.28×10^3	2.62×10^6	9991	0.693	8.710	7.64×10^3	2.60×10^6
2	6	9992	0.696	7.412	7.61×10^3	2.01×10^6	9985	0.695	7.502	7.06×10^3	2.02×10^6
2	7	9988	0.696	8.028	7.53×10^3	2.63×10^6	9992	0.692	8.411	6.20×10^3	2.43×10^6
2	8	9992	0.696	8.410	7.19×10^3	2.43×10^6	9989	0.696	8.519	7.66×10^3	3.95×10^6
2	9	9991	0.648	8.856	7.23×10^3	2.69×10^6	9989	0.632	8.341	7.42×10^3	2.39×10^6
2	10	9992	0.694	8.577	7.88×10^3	3.70×10^6	9988	0.692	7.500	7.27×10^3	3.76×10^6
2	1–10	99906	0.648	9.099	6.28×10^3	3.70×10^6	99890	0.632	9.653	6.20×10^3	4.33×10^6
3	1	9993	0.565	7.856	5.61×10^3	2.12×10^6	9990	0.561	7.791	4.21×10^3	2.08×10^6
3	2	9990	0.565	8.372	5.43×10^3	2.40×10^6	9991	0.559	7.646	5.15×10^3	2.01×10^6
3	3	9991	0.560	7.478	4.69×10^3	3.00×10^6	9989	0.559	8.105	5.20×10^3	2.25×10^6
3	4	9989	0.560	7.793	6.08×10^3	2.08×10^6	9987	0.541	8.416	5.39×10^3	2.43×10^6
3	5	9990	0.564	8.490	4.60×10^3	2.47×10^6	9987	0.563	7.462	4.76×10^3	3.82×10^6
3	6	9988	0.565	8.362	5.09×10^3	2.40×10^6	9990	0.563	8.362	4.54×10^3	2.40×10^6
3	7	9989	0.563	8.900	6.00×10^3	2.72×10^6	9988	0.563	9.958	5.49×10^3	3.40×10^6
3	8	9991	0.561	7.210	6.26×10^3	1.78×10^6	9991	0.558	7.403	6.40×10^3	2.08×10^6
3	9	9989	0.540	8.358	4.40×10^3	2.40×10^6	9988	0.514	8.185	4.83×10^3	2.30×10^6
3	10	9990	0.548	8.371	6.59×10^3	2.40×10^6	9991	0.557	7.290	5.25×10^3	1.82×10^6
3	1–10	99900	0.540	8.900	4.40×10^3	3.00×10^6	99892	0.514	9.958	4.21×10^3	3.82×10^6
5	1	9989	0.426	7.915	2.80×10^3	2.15×10^6	9986	0.427	8.430	2.72×10^3	2.44×10^6
5	2	9990	0.422	8.549	3.08×10^3	2.51×10^6	9987	0.429	9.260	2.68×10^3	2.94×10^6
5	3	9988	0.428	7.527	3.16×10^3	1.94×10^6	9985	0.429	9.100	2.76×10^3	2.84×10^6
5	4	9990	0.391	8.837	3.48×10^3	2.68×10^6	9991	0.397	7.775	3.60×10^3	2.07×10^6
5	5	9991	0.429	8.513	3.73×10^3	2.49×10^6	9989	0.424	7.451	3.14×10^3	1.90×10^6
5	6	9988	0.423	8.992	3.96×10^3	2.77×10^6	9989	0.423	8.356	2.42×10^3	2.40×10^6
5	7	9989	0.384	7.643	3.51×10^3	2.00×10^6	9990	0.422	7.637	3.69×10^3	2.00×10^6
5	8	9990	0.425	7.762	3.82×10^3	2.22×10^6	9991	0.424	7.757	3.63×10^3	2.06×10^6
5	9	9989	0.429	9.954	3.62×10^3	3.40×10^6	9988	0.425	8.065	2.70×10^3	2.23×10^6
5	10	9990	0.425	8.322	3.72×10^3	2.38×10^6	9989	0.428	7.507	2.96×10^3	1.93×10^6
5	1–10	99894	0.384	9.954	2.80×10^3	3.40×10^6	99885	0.397	9.260	2.42×10^3	2.94×10^6
10	1	9990	0.287	9.232	1.54×10^3	2.92×10^6	9988	0.284	8.362	2.29×10^3	2.40×10^6
10	2	9990	0.288	8.667	2.06×10^3	2.58×10^6	9989	0.287	8.667	1.73×10^3	2.58×10^6

Table A1 – *continued*

f	run	n	$\epsilon = 10^{-10}$				$\epsilon = 10^{-13}$				
			$\min(V'_i)$ [V_{esc}]	$\max(V'_i)$ [V_{esc}]	$\min(Q'_R)$ [J kg^{-1}]	$\max(Q'_R)$ [J kg^{-1}]	n	$\min(V'_i)$ [V_{esc}]	$\max(V'_i)$ [V_{esc}]	$\min(Q'_R)$ [J kg^{-1}]	$\max(Q'_R)$ [J kg^{-1}]
10	3	9990	0.284	7.525	1.98×10^3	1.94×10^6	9988	0.272	7.520	2.25×10^3	1.94×10^6
10	4	9990	0.247	7.763	2.09×10^3	2.07×10^6	9989	0.238	7.804	1.33×10^3	2.09×10^6
10	5	9990	0.289	8.246	1.96×10^3	2.33×10^6	9991	0.286	8.373	1.49×10^3	2.40×10^6
10	6	9992	0.288	8.358	1.60×10^3	2.40×10^6	9990	0.282	8.355	1.44×10^3	2.39×10^6
10	7	9990	0.291	8.580	1.59×10^3	2.53×10^6	9989	0.282	8.576	1.87×10^3	2.52×10^6
10	8	9990	0.288	8.020	1.30×10^3	2.21×10^6	9989	0.286	7.004	1.91×10^3	1.68×10^6
10	9	9989	0.286	8.064	1.71×10^3	2.23×10^6	9988	0.290	8.065	1.54×10^3	2.23×10^6
10	10	9990	0.200	8.326	2.05×10^3	2.38×10^6	9988	0.283	8.325	1.90×10^3	2.38×10^6
10	1–10	99901	0.200	9.232	1.30×10^3	2.92×10^6	99889	0.238	8.667	1.33×10^3	2.58×10^6

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.