

# Constraints on the pre-impact orbits of Solar system giant impactors

Alan P. Jackson,<sup>1,2★</sup> Travis S. J. Gabriel<sup>2</sup> and Erik I. Asphaug<sup>2,3</sup>

<sup>1</sup>Centre for Planetary Sciences, University of Toronto, 1265 Military Trail, Toronto, Ontario, M1C 1A4, Canada

<sup>2</sup>School of Earth and Space Exploration, Arizona State University, 781 E. Terrace Mall, Tempe, AZ 85287, USA

<sup>3</sup>Lunar and Planetary Laboratory, University of Arizona, 1269 E. University Blvd, Tucson, AZ 85721, USA

Accepted 2017 November 6. Received 2017 October 24; in original form 2017 June 27

## ABSTRACT

We provide a fast method for computing constraints on impactor pre-impact orbits, applying this to the late giant impacts in the Solar system. These constraints can be used to make quick, broad comparisons of different collision scenarios, identifying some immediately as low-probability events, and narrowing the parameter space in which to target follow-up studies with expensive  $N$ -body simulations. We benchmark our parameter space predictions, finding good agreement with existing  $N$ -body studies for the Moon. We suggest that high-velocity impact scenarios in the inner Solar system, including all currently proposed single impact scenarios for the formation of Mercury, should be disfavoured. This leaves a multiple hit-and-run scenario as the most probable currently proposed for the formation of Mercury.

**Key words:** methods: numerical – celestial mechanics – Moon – planets and satellites: formation – planets and satellites: terrestrial planets.

## 1 INTRODUCTION

Giant, planetary scale impacts are a key component of the final phase of terrestrial planet formation (e.g. Kenyon & Bromley 2006; Raymond et al. 2009; Kokubo & Genda 2010). Within our own Solar system evidence for giant impacts is ubiquitous. A giant impact event is the leading theory for the formation of Earth’s Moon (e.g. Cameron & Ward 1976; Canup 2004a), for which the impactor has been named ‘Theia’ by Halliday (2000). On Mars, a massive impact appears to be the best explanation for the hemispheric dichotomy (e.g. Wilhelms & Squyres 1984; Marinova, Aharonson & Asphaug 2008; Nimmo et al. 2008; Marinova, Aharonson & Asphaug 2011), such that the northern lowlands are then rather the hemisphere-scale Borealis impact basin. The large core fraction of Mercury has also been suggested to have resulted from formation in a giant impact event (Anic 2006; Benz et al. 2007; Asphaug & Reufer 2014). Indeed, the only terrestrial planet for which there is no clear evidence of a giant impact event is Venus, and it is likely no coincidence that this is also the terrestrial planet whose surface and geological history are least well understood. Giant impacts are also not limited to the inner Solar system; giant impact scenarios are proposed, for example, for the Pluto–Charon system (e.g. Stern et al. 2006; Canup 2011) and Haumea (Leinhardt, Marcus & Stewart 2010).

Each of the giant impact events identified above is postulated as being the final one experienced by the body in question – there may have been earlier events, and indeed for Earth it is likely there

were, but direct evidence for any earlier giant impact events was erased by the last. As these events played such an important role in determining the final basic characteristics (e.g. composition, angular momentum and state of differentiation) of the target planet, and the formation of satellites and dynamical relationships with remnants, if we want to understand the history of the Solar system it is important to examine where the impactors might have originated.

For the Moon in particular there are concerns about matching the isotopic similarity with Earth given the dynamical constraints and the physics of collisions (e.g. Asphaug 2014). Meteorites display a diversity of isotopic signatures widely believed to be the result of chemical gradients and inhomogeneities in the Solar nebula (e.g. Clayton 2003). While there is evidence of a decrease in isotopic diversity for larger bodies (e.g. Ozima et al. 2007), presumably due to accretional mixing across reservoirs, differences are quite substantial between Mars and other achondrite parent bodies. It is therefore expected that giant impactors and their targets will have readily discernible isotopic signatures (Kaib & Cowan 2015) that will be evident in the compositions of the final planets and their satellites. Outlining the origin of giant impactors can thus provide testable geochemical constraints and help validate models for interpreting isotope geochemistry (e.g. Young et al. 2016).

Because giant impacts mix and disrupt the compositions of their contributing materials, a consistent and correct geochemical interpretation requires understanding the provenance of the projectile. In some models for the impact origin of the Moon, for example, Theia is a fast interloping proto-planet from the outer Solar system that collided at several times the escape velocity  $v_{\text{esc}}$ , while in other models it is a Mars-like planet that accreted nearby and collided just barely over  $v_{\text{esc}}$ . Dynamical analysis is therefore a required

\* E-mail: [ajackson@cita.utoronto.ca](mailto:ajackson@cita.utoronto.ca)

compliment to geochemical analysis, and the parameters of a giant impact event can provide unique insight into the orbit of the impactor immediately prior to the collision, allowing one to assess the likelihood of the projectile deriving from a given source region in the Solar system.

Rivera (2002) and Quarles & Lissauer (2015) explored the possible origins of Theia using suites of  $N$ -body integrations to find locations in which a potential Theia can remain quasi-stable for several tens of Myr before colliding with Earth to produce a system reminiscent of the present terrestrial planets. While  $N$ -body simulations afford perhaps the most detailed method of examining the origins and timing of an impact event, they are computationally expensive and thus one is inevitably forced to limit the parameter space under investigation. The motivation for this work is to obtain a more general analytical basis for examining giant impact projectile origins, which can be applied to the several competing hypotheses for Moon formation, as well as to giant impact scenarios involving Mercury, Mars, Pluto and other planets.

The impact velocity  $v_{\text{imp}}$  is one of the most important parameters in determining the outcome of a planet-forming collision (e.g. Asphaug 2010; Leinhardt & Stewart 2012). In terms of the pre-impact orbits,  $v_{\text{imp}}$  is equal to the relative velocity between the orbits of the target and impactor at the point of impact. If we can constrain the orbit of the target, then we can use this information to place constraints on the pre-impact orbit of the impactor, and hence, its provenance and possible composition. In most giant impacts, the projectile is much less massive than the target, in which case we expect the orbit of the target to be close to the orbit of the final planet as we see it today. In making this assumption, the allowable projectile orbits can be computed analytically much faster than using  $N$ -body simulations, enabling the exploration of much larger parameter spaces and a wider range of giant impact scenarios.

In this work, we analyse a variety of scenarios for the last several giant impacts believed to have occurred in the Solar system. We use the velocities of those collisions as interpreted by various published models to constrain the pre-impact orbits of the impactor bodies, e.g. Theia. First, in Section 2, we outline our methods for relating the pre-impact orbits to the orbit of the target body. In Section 3, we then apply this to the Moon-forming giant impact for several different impact scenarios: the ‘Canonical’ model (Hartmann & Davis 1975; Cameron & Ward 1976, as reviewed by e.g. Canup 2004a), the hit-and-run model of Reufer et al. (2012), the equal-mass impactors model of Canup (2012), and the small, fast impactor model of Čuk & Stewart (2012), before comparing to the results of Quarles & Lissauer (2015) as a benchmark. We then apply our pre-impact orbit determination schema to the formation of Mercury (Benz et al. 2007; Asphaug & Reufer 2014) in Section 4. In Section 5, we briefly discuss the Borealis basin impact on Mars (Marinova et al. 2008, 2011), as an example of a scenario where the target has an eccentric orbit, and other Solar system giant impacts including Pluto–Charon in Section 6. Finally, we bring together our conclusions in Section 7.

## 2 METHOD

Jackson et al. (2014) (hereafter J14) presented equations relating the orbit of a particle before and after it receives a ‘kick’ to its velocity. They used these in the context of relating the orbits of debris produced in an impact event to the orbit of the progenitor body. We can however reverse the causal relationship and use the same equations to relate the orbit of an impactor to that of the target body and the relative velocity between the two at the time of impact.

We direct the reader to section 2 of J14 for the full derivation of these equations, and a detailed discussion of their behaviour; however, for reference, we reproduce the equations we shall use here for circular target orbits,

$$\frac{a_T}{a_I} = 1 - 2 \left( \frac{v_{\text{rel}}}{v_k} \right) S_\theta S_\phi - \left( \frac{v_{\text{rel}}}{v_k} \right)^2, \quad (1)$$

$$e_I^2 = 1 - \left( \frac{h_I^2}{h_T^2} \frac{a_T}{a_I} \right), \quad (2)$$

$$\cos I_I = \left[ 1 + \left( \frac{v_{\text{rel}}}{v_k} \right) S_\theta S_\phi \right] \left( \frac{h_I^2}{h_T^2} \right)^{-\frac{1}{2}}, \quad (3)$$

$$\frac{h_I^2}{h_T^2} = 1 + 2 \left( \frac{v_{\text{rel}}}{v_k} \right) S_\theta S_\phi + \left( \frac{v_{\text{rel}}}{v_k} \right)^2 (C_\theta^2 + S_\theta^2 S_\phi^2), \quad (4)$$

and for eccentric target orbits,

$$\frac{a_T}{a_I} = 1 - \left( \frac{v_{\text{rel}}}{v_k} \right)^2 - \frac{2}{\sqrt{1 - e_T^2}} \left( \frac{v_{\text{rel}}}{v_k} \right) S_\theta (S_{\phi-f_T} + e_T S_\phi), \quad (5)$$

$$e_I^2 = 1 - (1 - e_T^2) \left( \frac{h_I^2}{h_T^2} \frac{a_T}{a_I} \right), \quad (6)$$

$$\cos I_I = \left[ 1 + \frac{(1 - e_T^2)^{\frac{1}{2}}}{1 + e_T C_{f_T}} \left( \frac{v_{\text{rel}}}{v_k} \right) S_\theta S_{(\phi-f_T)} \right] \left( \frac{h_I^2}{h_T^2} \right)^{-\frac{1}{2}}, \quad (7)$$

$$\begin{aligned} \frac{h_I^2}{h_T^2} = & 1 + 2 \frac{(1 - e_T^2)^{\frac{1}{2}}}{1 + e_T C_{f_T}} \left( \frac{v_{\text{rel}}}{v_k} \right) S_\theta S_{(\phi-f_T)} \\ & + \frac{1 - e_T^2}{(1 + e_T C_{f_T})^2} \left( \frac{v_{\text{rel}}}{v_k} \right)^2 (C_\theta^2 + S_\theta^2 S_{(\phi-f_T)}^2). \end{aligned} \quad (8)$$

We make some minor alterations to the notation of J14 to emphasize the purpose for which we are using the equations here. The subscripts T and I refer to orbital elements and quantities of the target and impactor, respectively. The velocity kick between the orbits is denoted as  $v_{\text{rel}}$ , since it is the relative velocity at the point of impact in this case. Otherwise we follow the same notation as J14:

|                     |  |
|---------------------|--|
| $a$                 | semimajor axis,  |
| $e$                 | eccentricity,  |
| $I$                 | inclination,   |
| $f$                 | true anomaly,  |
| $\Omega$            | longitude of ascending node,   |
| $\omega$            | argument of pericentre,  |
| $v_k$               | circular speed at orbital distance $a$ ,   |
| $S_x$               | $\sin(x)$ ,  |
| $C_x$               | $\cos(x)$ ,  |
| $\theta$ and $\phi$ | spherical polar coordinate angles defining the orientation of the relative velocity. |

Since orbits precess over relatively short time-scales, the orientation of the orbits,  $\Omega$  and  $\omega$  are not of interest here and the equations for  $\Omega_I$  and  $\omega_I$  are omitted. Similarly, the true anomalies at the time of impact are not of interest other than in the true anomaly of the target defining the range of possible impactor orbits. Note also that since  $v_{\text{rel}}$  is defined at the instant of impact the impulsive assumption of J14 is always satisfied.

For a given target, orbit and relative velocity there is a unique mapping between  $\theta$  and  $\phi$  and the impactor orbit; however,  $\theta$  and  $\phi$  are not known so we assume them to be isotropically distributed.

As such each target orbit and relative velocity maps to a range of possible impactor orbits.

### 3 THE MOON

The formation of the Moon is the most well-studied giant impact event in the Solar system. The modern theory was first developed in the 1970s, following the Apollo missions, which revealed a completely igneous planet with very little metallic iron (e.g. Hartmann & Davis 1975; Cameron & Ward 1976), and over decades of refinement produced what has come to be known as the ‘Canonical’ model for Moon formation (e.g. Canup 2004a). In this model, a roughly Mars-sized Theia strikes the nearly fully formed proto-Earth at around  $45^\circ$  (the most probable angle) at just above the mutual escape velocity, in what can be described as a ‘graze-and-merge’ collision, with most of Theia ultimately accreted by Earth.

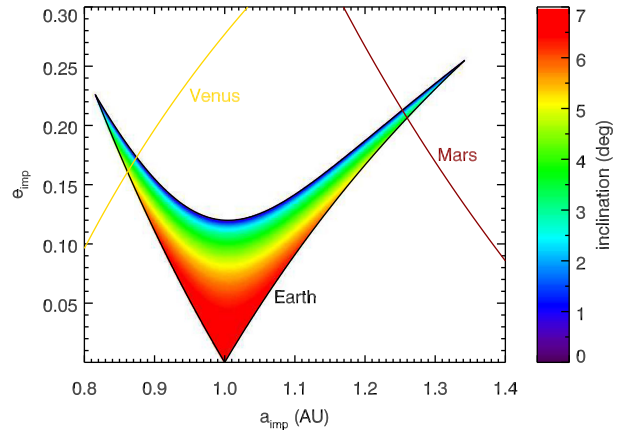
The Canonical model can well re-produce many aspects of the Earth–Moon system, including mass, angular momentum and the depletion of iron in the Moon relative to Earth. However, a large fraction of the material that forms the Moon in the Canonical model is derived from Theia, whereas the growing body of isotopic evidence (especially oxygen) suggests that the Moon is nearly identical to Earth, leading to a conflict with the Canonical model (e.g. Asphaug 2014). This conflict has initiated many new studies to investigate how it might be resolved, including a renewed interest in alternative impact scenarios. As such we will analyse a total of four models for the Moon-forming giant impact: the Canonical model and those of Reufer et al. (2012), Canup (2012) and Ćuk & Stewart (2012). Throughout, we assume that the target body occupies a circular orbit at 1 au and inclinations are measured relative to the plane of this orbit.

#### 3.1 Canonical and Canup (2012) models

The Canonical model involves an impact between a proto-Earth of around  $0.9 M_\oplus$  and a Mars-sized Theia of around  $0.1 M_\oplus$  colliding at an oblique angle of around  $45^\circ$  at a low speed of around  $1.05 v_{\text{esc}}$ , where  $v_{\text{esc}}$  is the mutual escape velocity (e.g. Canup & Asphaug 2001; Canup 2004a,b). The model is Canonical in that it takes the angular momentum of the final Earth–Moon system as a conserved quantity, which has to be close to the value observed today. According to Ćuk & Stewart (2012), the original angular momentum could have been more than twice as great, leading to diverse new models. Canup (2012) suggests an impact between two nearly equal mass semi-Earths at a lower angle, but with a very similar velocity. Both the Canonical and the semi-Earth models use very similar relative velocities, and will therefore produce similar results for the potential pre-impact orbits of Theia, albeit that the assumption that the orbit of the target is that of the present-day Earth is probably less accurate for the model of Canup (2012).

In Fig. 1, we show the range of pre-impact orbital parameters allowed for Theia under the Canonical and Canup (2012) models. In the three dimensions of semimajor axis, eccentricity and inclination, this is a thin surface which we represent here in 2D by applying a colour map to represent the inclination. The impact velocity of  $1.05 v_{\text{esc}}$  corresponds to  $v_{\text{rel}}/v_k = 0.12$ , or  $v_{\text{rel}} = 3.5 \text{ km s}^{-1}$ . This is a very low impact velocity, only just above the  $1 v_{\text{esc}}$  of a pure free-fall impact and hence the range of pre-impact orbits is also quite small.

The 3D surface of allowed pre-impact orbits of Theia occupies a distinctive ‘V’ shape when projected into semimajor axis and eccentricity which is characteristic of this type of problem. The inner



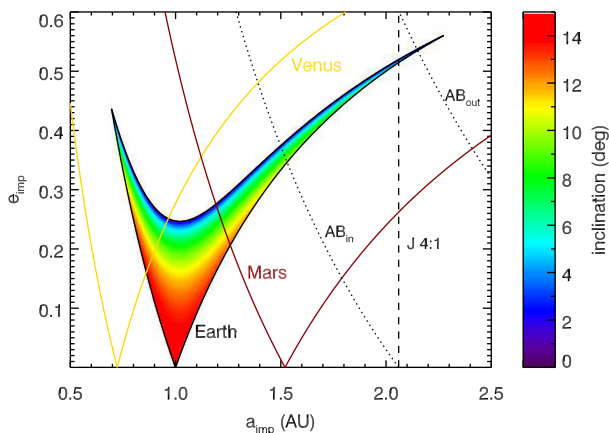
**Figure 1.** The allowed semimajor axis – eccentricity – inclination space for Theia in the Canonical and Canup (2012) models for an impact velocity of  $1.05 v_{\text{esc}}$ . Orbits above the yellow line are Venus crossing, while those above the dark red line are Mars crossing.

(low semimajor axis) boundary corresponds to potential Theian orbits that impact Earth when Theia is at apocentre, while the outer (high semimajor axis) boundary corresponds to potential Theian orbits that impact Earth when Theia is at pericentre. The upper boundary is the line along which  $I_{\text{imp}} = 0^\circ$ . In yellow and dark red, we show the lines above which Theia will cross the orbits of Venus and Mars, respectively. As seen, in these very low relative velocity scenarios, the ranges of allowed orbits in the Venus- and Mars-crossing regions are small and most of the allowed pre-impact orbits for Theia only cross the orbit of Earth.

While we only directly constrain the orbit of Theia immediately before the impact, the allowed orbital parameters suggest that in these two scenarios Theia likely originated from near the proto-Earth. It is relatively easy to excite the eccentricity of an early Solar system body over millions of years, and so we would expect that Theia could start off somewhat below the coloured region in Fig. 1 before being excited up into the allowed pre-impact region later. Indeed this would fit well with the Moon-forming impact occurring somewhat late in the process of Solar system formation.

Dynamical pathways that involve bringing Theia from a more distant birthplace (in semimajor axis) on to the pre-impact orbit are more difficult to construct. Such pathways necessarily involve scattering by other Solar system planets, and as we can see in Fig. 1, the regions of overlap with Venus- or Mars-crossing orbits are small, so that any such multiple scattering pathway would have to be rather fine-tuned to produce the correct relative velocity at impact. This expectation that Theia originates from within roughly the semimajor axis range allowed for the pre-impact orbit is supported by published detailed dynamical models, as we discuss in Section 3.4.

If, by the time of final accretion, there was a constant gradient in isotopic composition with distance from the Sun, this would imply that the difference in compositions between proto-Earth and a nearby Theia were less than that between Earth and Mars or Earth and Venus. However, the isotopic dissimilarity between the Earth and Mars is pronounced, and the Moon predicted by the Canonical model, being made mostly out of Theia material, would still be impossible to explain without Theia forming out of almost identical materials as the Earth, or else a vigorous, whole-mantle mixing mechanism to fully or partially homogenise isotopic abundances between the post-impact Earth and proto-lunar disc (e.g. Pahlavan & Stevenson 2007), or some form of post-formation processing



**Figure 2.** The allowed semimajor axis – eccentricity – inclination space for Theia in the Reufer et al. (2012) models for an impact velocity of  $1.2 v_{\text{esc}}$ . Orbits above the yellow line are Venus crossing, while those above the dark red line are Mars crossing. The dashed line indicates the 4:1 mean-motion resonance with Jupiter at 2.06 au, while the dotted lines indicate orbits that have apocentres at 2.06 and 3.3 au, the inner and outer edges of the main asteroid belt.

(Salmon & Canup 2012). The near-identity of Earth–Moon isotopic abundances (e.g. Young et al. 2016) is a critical point to bear in mind when comparing the different scenarios.

### 3.2 Reufer et al. (2012) model

Reufer et al. (2012) proposed an alternative giant impact scenario with a ‘hit-and-run’ impact in which a substantial silicate-dominated fraction of Theia continues downrange after the impact and is not incorporated into the Earth–Moon system. Hit and run requires a slightly higher impact velocity than the Canonical scenario, of  $1.2 v_{\text{esc}}$ , which corresponds to  $v_{\text{rel}}/v_k = 0.25$ , or  $v_{\text{rel}} = 7.4 \text{ km s}^{-1}$ , about twice the relative velocity of the Canonical scenario. Thus, it brings more impact energy, but leaves behind less Theia-derived material, than the Canonical scenario. Because only a fraction of the proto-lunar disc derives from Theia in the hit-and-run scenario, it remains isotopically more similar to Earth. Forming from a hotter disc, the protolunar materials then have potential for further isotopic equilibration.

As we show in Fig. 2, owing to the increased velocity, the allowed range of orbital parameters for the pre-impact orbit of Theia is significantly expanded in the hit-and-run scenario as compared with the Canonical or Canup (2012) scenarios. As a result, a significant portion of the parameter space is now in regions that will cross the orbits of Venus or Mars, in addition to the proto-Earth. In particular, there is a tail that extends somewhat past the 4:1 mean-motion resonance with Jupiter at 2.06 au that marks the approximate inner edge of the asteroid belt. For considering potential origins of Theia in the region of the present asteroid belt, a more relevant comparison is the apocentre of the pre-impact orbit, since Theia will have been perturbed from its initially quasi-stable orbit before impact. The region between the two dotted lines thus indicates those orbits that have apocentres between 2.06 and 3.3 au, which encompasses the Main Belt.

The possibility that Theia originated in the Main Belt, indeed around the orbit of Ceres, is particularly relevant to composition, because the region  $\sim 2\text{--}3$  au and beyond was considerably more volatile-rich than around 1 au during planet formation. One of the

collisional scenarios that Reufer et al. (2012) identify as a good fit to producing the Earth–Moon system is one in which the impactor is water rich, as might be expected if Theia originated from the outer Main Belt. Ceres, the asteroid at  $a = 2.8$  au that contains almost half the mass of the present Main Belt, has a substantial mass fraction of water (Thomas et al. 2005; Russell et al. 2016), and more massive planetesimals, since lost from the region between Mars and Jupiter, might have had even greater water fractions (e.g. Asphaug 2017). At  $0.15\text{--}0.2 M_{\oplus}$ , Theia might have represented an anomalously large Main Belt planetesimal by the time of the presumed Moon-forming giant impact  $\sim 50\text{--}150$  Myr after the first solids. Still, simulations have shown that individual large bodies can survive in the vicinity of the asteroid belt for up to  $\sim 1$  Gyr before colliding with a terrestrial planet (e.g. Chambers 2007).

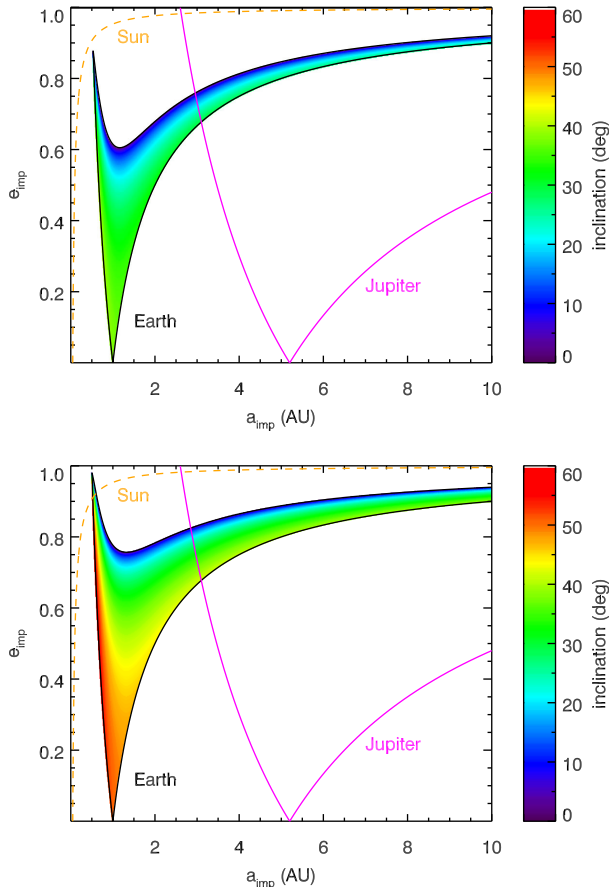
### 3.3 Čuk & Stewart (2012) model

Whereas the scenarios of Reufer et al. (2012) represent only a relatively minor departure from the Canonical model, at least in terms of the relative velocity required, the scenarios proposed by Čuk & Stewart (2012) are a much more dramatic deviation. In their scenario, the proto-Earth is already a rapidly rotating object, with spin period  $< 3$  h, and Theia is a smaller body that strikes at much higher velocity, retrograde to the Earth’s spin. Arguing that final Earth–Moon angular momentum could be twice that of the present day, Čuk & Stewart (2012) introduce models that successfully reproduce other aspects of the Earth–Moon system. Required impact velocities are  $2\text{--}2.5 v_{\text{esc}}$ , corresponding to  $v_{\text{rel}}/v_k = 0.65\text{--}0.86$ , or  $v_{\text{rel}} = 19.3\text{--}25.6 \text{ km s}^{-1}$ , almost a sixfold increase from the relative velocity of the Canonical scenario.

As would be expected given the much larger relative velocities required by this scenario, the parameter space of pre-impact orbits shown in Fig. 3 is much larger than that in Figs 1 and 2. These very high relative velocities actually allow for hyperbolic orbits, such that it is not possible to show the entire parameter space in Fig. 3. While the allowed parameter space based solely on the relative velocity is very large, we can use additional constraints to reduce the plausible range. Bodies that pass extremely close to the Sun will experience extreme heating and strong tides that can alter the structure and orbit of the body. As such impactors on Sun-grazing orbits are implausible. We also note that there is today a complete lack of material interior to Mercury and that terrestrial planet formation models also do not generally include material so close to the Sun (e.g. Raymond et al. 2009; Chambers 2013). The Sun-grazing region does not have a sharp edge, but as a representative boundary in Fig. 3 we show in orange the line above which Theia comes within  $10 R_{\odot}$  at pericentre.

We also acknowledge the model of Rufu, Aharonson & Perets (2017), where the Moon is generated through the accretion of several smaller moonlets, each moonlet generated in a series of lesser (and individually more probable) giant impacts into a rotating proto-Earth. Each subsequent impact would either diminish or enhance the overall angular momentum of the system. Across their simulations of giant impacts with the target rotating at  $\omega = 0.1, 0.25, 0.5\omega_{\text{max}}$ , where  $\omega_{\text{max}}$  is the spin limit of the target, the scenarios that generate debris discs composed predominantly of Earth material require impact velocities of  $v_{\text{imp}} \geq 2.0 v_{\text{esc}}$ . So in terms of required velocity, we view the model of Rufu et al. (2017) in the same light as that of Čuk & Stewart (2012), but instead of requiring a single, high-velocity impact into a fast-rotating Earth, Rufu et al. (2017) requires a series of impactors with high relative velocity.





**Figure 3.** Semimajor axis – eccentricity – inclination space for the pre-impact orbit of Theia in the Čuk & Stewart (2012) models for an impact velocity of  $2.0v_{\text{esc}}$  (top) and  $2.5v_{\text{esc}}$  (bottom). Orbits above the magenta line are Jupiter crossing, while those above the dashed orange line are Sun grazing (having pericentres within  $10 R_{\odot}$  of the Solar surface).

### 3.4 Comparison to Quarles & Lissauer (2015)

The analytical approach developed here can be compared directly with numerical simulations in the case of Earth–Moon system formation. Quarles & Lissauer (2015) (hereafter QL15) conducted a detailed set of  $N$ -body integrations of a model early Solar system with five terrestrial planets, Mercury, Venus, Mars, the proto-Earth and Theia, looking for successful cases in which Theia collides with the proto-Earth between 8 and 200 Myr after the start of the simulation, the widest possible range of ages for Moon formation. The majority of their simulations (their table 3) begin with the semimajor axis of Theia lying between 0.76 and 1.55 au, covering the range allowed for both the Canonical and semi-Earth (Canup 2012) scenarios. They conduct a smaller set of simulations extending the range of initial semimajor axes for Theia to 0.44 and 2.18 au, thereby also largely covering the range allowed for the hit-and-run (Reufer et al. 2012) scenario.

In most successful cases of Moon formation, QL15 find that Theia originates from inside 1.3 au, and that from 0.76 to 1.3 au the distribution of initial semimajor axes for Theia is fairly flat (their fig. 5). Origins for Theia outside 1.3 au are much more likely to result in an outcome that does not look like the Solar system. In Fig. 4, we overlay our allowed pre-impact semimajor axis – eccentricity – inclination distribution for the Canonical and Canup (2012) scenarios (as shown in Fig. 1) on the distribution of simula-

tion outcomes for the Nice4 data set of QL15 (after their fig. 2d). We can see that the initial conditions for which they identify success or pseudo-success cluster just outside our allowed pre-impact orbit distribution. This is expected as it allows Theia to have a substantial period of quasi-stability before moving on to a crossing orbit as defined by the bounded area in Fig. 4 and impacting the proto-Earth.

For their broader set of simulations that allows initial semimajor axes for Theia of 0.44–2.18 au, QL15 divide the results into three broad bins by initial semimajor axis,  $0.44 \leq a \leq 0.75$ ,  $0.75 < a \leq 1.55$  and  $1.55 < a \leq 2.18$ . Within the inner bin they find that 4.7 per cent of collisions are successful, in the middle bin 14 per cent are successful, and in the outer bin 6.3 per cent are successful. A further 11.4 per cent, 22 per cent and 20.9 per cent, respectively, fall into the pseudo-success category in which between 8 and 200 Myr after the start of the simulation there is a collision between any small body (Mercury, Mars, Theia) and any large body (Earth, Venus) other than Theia–Earth.

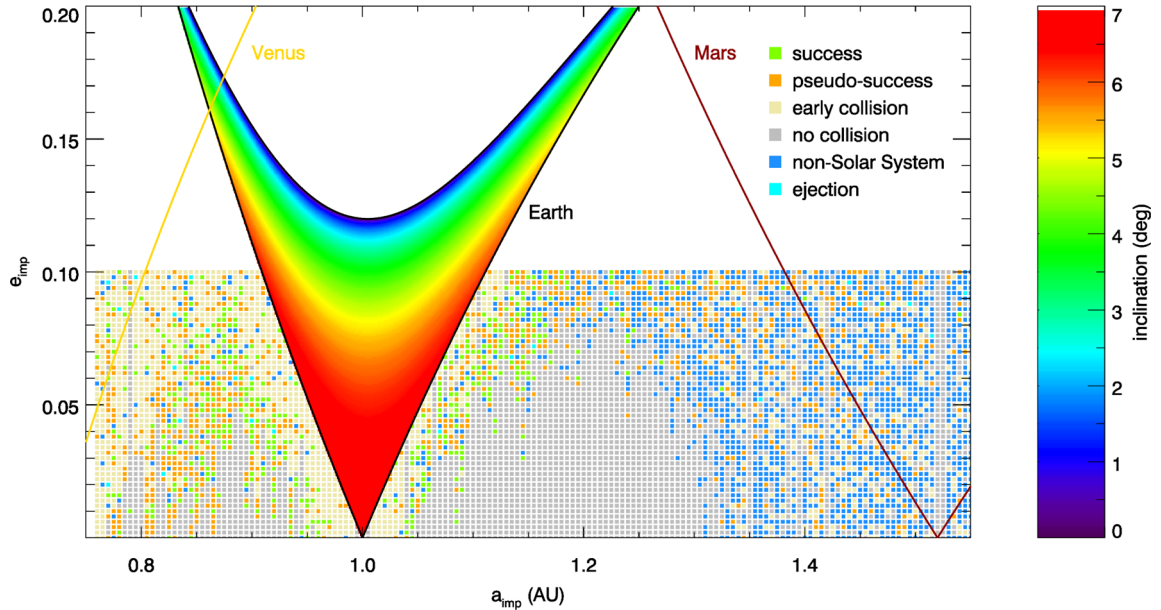
The middle bin corresponds roughly to the allowed semimajor axis range for the Canonical and Canup (2012) models, while the inner bin is largely interior to the allowed ranges for all three low-velocity models [Canonical, Canup (2012) and Reufer et al. (2012)] and the outer bin covers the outer part of the allowed range in the Reufer et al. (2012) model. The inner bin has the lowest rate of successful impacts and thus matches with our expectations. While it is more difficult to examine the distribution in any single bin due to the lower number of simulations, all of the successful cases of the inner bin are concentrated at the larger semimajor axes. The results of QL15 are thus consistent with our suggestion that  $N$ -body simulations to conduct detailed analysis of proposed giant impact scenarios should be concentrated in the regions indicated by the distributions for the pre-impact orbits.

As we mentioned in Section 3.1, we could envisage dynamical pathways involving multiple scattering events with the other terrestrial planets, and/or secular orbital evolution to substantially change the semimajor axis of Theia prior to it being put on to the final impact trajectory. In the particular cases of the Canonical and Canup (2012) scenarios, this seems rather unlikely due to the small size of the overlap between the allowed pre-impact orbits and the Mars- and Venus-crossing regions. The other scenarios have larger overlaps, but nonetheless any pathway involving multiple scattering events or additional orbital evolution is inherently more complex than one involving only a single scattering event, and so we would expect it to be less likely, unless the single scattering event is extreme such that it has a much lower probability. This is supported by the  $N$ -body studies of QL15, since they do indeed find that in successful impacts Theia originates from close to the distribution of pre-impact orbit parameters.

### 3.5 Additional constraints on the pre-impact orbit

In addition to the relative velocity of the impact, it is valuable to consider other quantities to further constrain the parameter space of allowed pre-impact orbits. One such additional constraint comes from the obliquity of Earth and the inclination of the lunar orbit. Together these define the orientation of the total angular momentum vector of the Earth–Moon system.

In all of the scenarios, we have discussed above, with the exception of Čuk & Stewart (2012), the majority of the angular momentum of the post-impact Earth–Moon system is imparted by the giant impact (see Canup 2008 for an explicit examination of the effect of pre-impact rotation for the Canonical scenario). Although works by Čuk & Stewart (2012) and Wisdom & Tian (2015) have suggested



**Figure 4.** Our semimajor axis – eccentricity – inclination distribution for the Canonical (and Canup 2012) scenario overlaid on the results of QL15 as shown in their fig. 2d for the Nice4 data set. The key at top right indicates the different collision outcomes in the background plot (see QL15 for the full definitions). Data from QL15 provided by Quarles.

that Solar tides may be able to change the magnitude of the total angular momentum of the Earth–Moon system more than previously thought, the orientation of the total angular momentum vector is nonetheless still an important constraint. Even for the scenarios of Čuk & Stewart (2012), though much of the final angular momentum of the system is derived from the rapid spin of the pre-impact proto-Earth, the orientation of the impact relative to the spin axis is still important in determining the outcome.

A constraint on the orientation of the post-impact angular momentum vector translates into a constraint on the inclination of the pre-impact orbit. Since a planet is a small target on the scale of the Solar system, we can consider the plane perpendicular to the relative velocity to be analogous to a dart board or a clock face. Even for an impactor with a very small pre-impact inclination it is possible for the strike to occur at the 12 o’clock or 6 o’clock positions at the top or bottom of the board (Fig. 5b), and we expect the distribution of strike locations to be roughly uniform around the circle. Clearly if all of the post-impact angular momentum is derived from the impact then an impact that strikes at the top or bottom will result in the post-impact angular momentum being oriented at  $90^\circ$  to the target orbit. The smallest obliquity will occur if the impact strikes exactly in the plane of the target orbit, at the 9 o’clock or 3 o’clock positions of our planetary dart board. In this case, the obliquity will be equal to the inclination of the pre-impact orbit of the impactor (Fig. 5c). The obliquity will only be zero if the impactor and target orbits are co-planar (Fig. 5a). Note that in the extremely unlikely case of a head-on impact ( $\theta = 0$ ) the post-impact spin will be zero and thus the obliquity undefined.

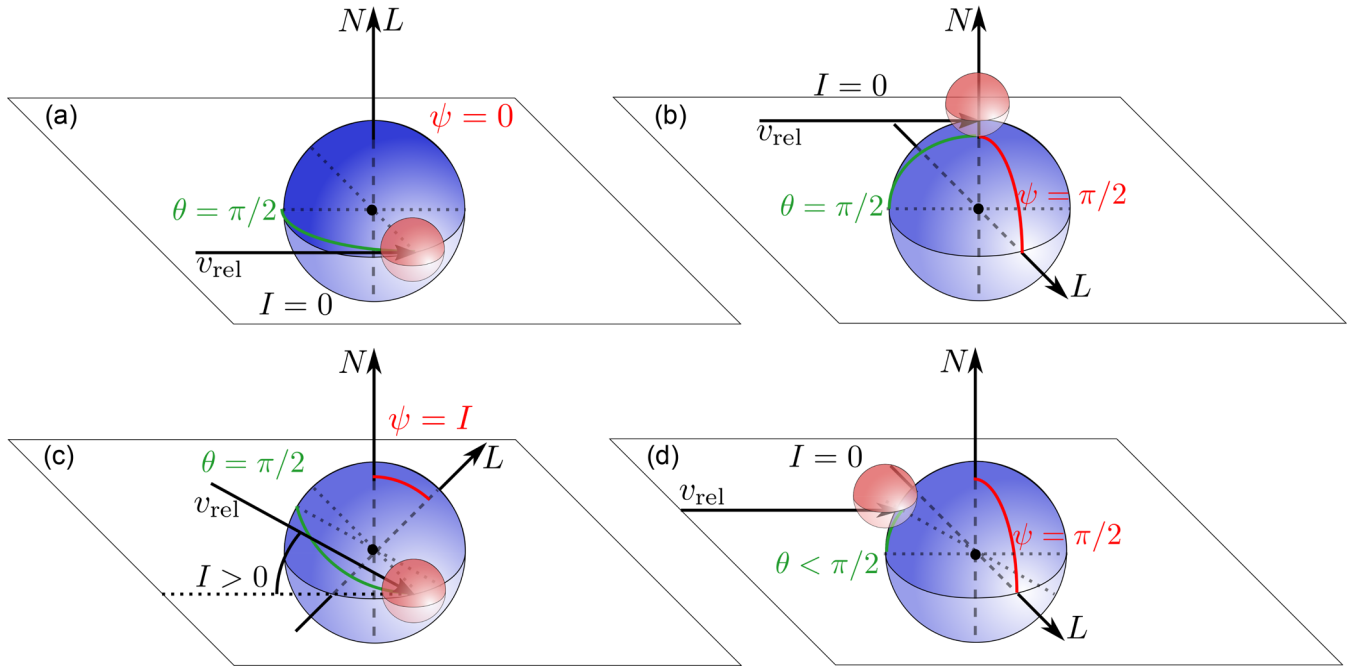
Today the orbit of the Moon is inclined at  $5.1^\circ$  to the ecliptic, while Earth has an obliquity of  $23.4^\circ$  and the mutual obliquity between the spin axis of Earth and the orbit of the Moon varies between  $18^\circ$  and  $28^\circ$  due to the effects of precession and nutation. The orbit of the Moon contains the majority ( $>80$  per cent) of the total angular momentum of the Earth–Moon system. Over time tidal evolution has resulted in a transfer of angular momentum from the spin of Earth to the lunar orbit. Earlier in this evolution when the Moon

was closer to Earth the inclination of the lunar orbit was larger, while the obliquity of Earth and the mutual obliquity were smaller, with all three being probably around  $10^\circ$  initially (e.g. Touma & Wisdom 1994).

If we assume that both the target and the impactor are non-rotating initially, such that all of the angular momentum of the post-impact system is imparted in the impact, then we can use the initial obliquity of  $\sim 10^\circ$  of the Earth–Moon system to constrain the inclination of the impactor pre-impact orbit to be less than  $10^\circ$ . This does not result in any reduction in the parameter space for the Canonical and Canup (2012) models, but does exclude some low-eccentricity pre-impact orbits for the Reufer et al. (2012) models.

The Čuk & Stewart (2012) scenario is more complicated, because here the pre-impact spin of the target is very fast and contributes the majority of the final angular momentum of the system. Although they only examined impacts in which the impactor strikes in the equatorial plane of the target, their results show a distinct preference for retrograde impacts over prograde or head-on impacts. The study of Canup (2008) also showed that impacts in which the equatorial plane of the target is orthogonal to the impact lie roughly halfway between prograde and retrograde impacts in outcome. So we can expect that the preference for retrograde impacts over prograde impacts of Čuk & Stewart (2012) will extend to a preference over orthogonal impacts. As such, while the constraints we can place on the pre-impact inclination are not as strong as for the other scenarios, we can still exclude significant regions of the large range of inclinations allowed in Fig. 3. For example, excluding inclinations over  $20^\circ$  (thus cutting Fig. 3 along the cyan band) allows us to exclude  $\sim 91$  per cent of the orbits in the upper panel of Fig. 3 and  $\sim 97$  per cent of the orbits in the lower panel of Fig. 3. Any exclusion of higher inclination orbits significantly increases the minimum eccentricity of the pre-impact orbit and increases the proportion of the solution space that is encompassed by the tail of Jupiter-crossing orbits.

For the case of the Čuk & Stewart (2012) scenario, it is also worth looking again at the very high relative velocities required.



**Figure 5.** Diagram to illustrate the relationship between the orientation of the impact, the inclination and the obliquity. The white plane represents the orbital plane of the target (the larger blue sphere), with the vector  $N$  being the normal to the orbital plane. The vector  $L$  is the (post-impact) spin angular momentum of the target, with the pre-impact spin assumed to be zero. The angle  $\theta$  is the impact angle, measured between the centres of the bodies at the point of impact and a line through the centre of the target parallel to the impact velocity vector,  $\theta = \pi/2$  for a grazing impact and  $\theta = 0$  for a head-on impact. The angle  $\psi$  is the obliquity, the angle between the spin axis and the orbit normal. The angle  $I$  is the inclination, the angle between the orbital planes of the target and the impactor. In panels (a) and (b), we illustrate how the same impact angle and inclination can result in very different obliquities if the target is struck at the ‘side’ as opposed to the ‘top’ or the ‘bottom’. In panel (c), we then show how for non-zero inclinations the minimum obliquity is greater than zero even if the target is struck at the side. Finally, panel (d) shows an impact with the same orientation as (b) but with a smaller impact angle. Although the magnitude of the spin angular momentum imparted may be smaller in this case, the obliquity is the same.

In a population of self-stirred bodies, the velocity dispersion is roughly equal to the escape velocity, and so a random velocity of twice the escape velocity is unlikely. This is supported by QL15, who consider the distribution of the impacts in their simulations as a function of the impact velocity and find that there are very few impacts with velocities as high as  $2 v_{\text{esc}}$ , and almost no successful cases with impact velocities above  $1.6 v_{\text{esc}}$ . Broader studies of terrestrial planet formation (e.g. O’Brien, Morbidelli & Levison 2006; Quintana et al. 2016) similarly find that impacts at high relative velocities are unlikely, especially for embryo–embryo impacts.

This low likelihood of high-velocity impacts only applies to projectiles originating from within the inner Solar system, however. A Theia originating in the outer Solar system would necessarily have a higher typical impact velocity, due to the high eccentricity required to cross the Earth. Dynamical studies of terrestrial planet formation do not typically include sources in the outer Solar system, and so it is difficult to judge the relative probabilities of a high-velocity giant impact from within the inner Solar system versus an impact by an outer Solar system body during terrestrial planet formation. If a high-velocity impact scenario like that of Čuk & Stewart (2012) becomes the favoured scenario for the formation of the Moon, then the possibility of an outer Solar system origin is of particular interest, as it is likely that Theia would have been rich in volatiles including water.

### 3.6 Discussion of Moon-formation scenarios

From our analysis of the allowed distributions of pre-impact orbits we can divide the Moon-formation scenarios into two broad

categories. The Canonical, Canup (2012) and Reufer et al. (2012) scenarios are all relatively low impact velocity events. The Čuk & Stewart (2012) scenario and the preferred impacts of Rufu et al. (2017) occupy a second category as high-velocity events. The allowed pre-impact orbits for low-velocity events place the origin of Theia firmly within the inner Solar system, while for the high impact velocity events, outer Solar system origins are allowed.

Quarles & Lissauer (2015) show that the number of Earth-impactors originating from within the inner Solar system at the high velocities required for the Čuk & Stewart (2012) scenario is very low, and the number of successful impacts is even lower. They also find that within the inner Solar system an impactor that satisfied the impact conditions of the Čuk & Stewart (2012) scenario would have more likely originated beyond the orbit of Mars, where the influence of Jupiter is stronger. This is in accordance with our suggestion that the impactor in this high-velocity scenario could have originated in the outer Solar system, for which a high impact velocity is guaranteed.

Although it is easier to obtain such a high impact velocity through a scattering encounter with Jupiter than through fortuitous multiple scattering within the inner Solar system, this is still an unlikely scenario. Due to the high escape velocity of Jupiter relative to the Keplerian orbital speed, objects that undergo encounters with Jupiter are likely to be ejected from the Solar system (e.g. Wyatt et al. 2017). This applies whether the object being scattered originates from the inner or outer Solar system. Scattering by Jupiter is a necessary but not sufficient condition; the potential impactor must also be scattered on to a trajectory that crosses the inner planets, and it must then impact the desired terrestrial planet before it

is ejected from the Solar system by a subsequent encounter with Jupiter.

A dynamical pathway that requires scattering with Jupiter prior to impact with the proto-Earth also requires tighter constraints on the inclination of the pre-impact orbit than from obliquity considerations, since the orbits of Earth and Jupiter are well aligned ( $1.3^\circ$  today). The inclination of the pre-impact orbit must be comparably low for it to be able to intersect both planets.

In their simulations of the debris ejected by the Moon-forming impact, which we can take as a rough proxy for the behaviour of bodies in the inner Solar system, Jackson & Wyatt (2012) find that 2804 particles out of a total of 36 000 are ejected from the Solar system over 10 Myr of simulated time, compared with 27 that hit the Sun, both outcomes that are dominated by scattering with Jupiter. As such we might expect to need tens of potential impactors to scatter off Jupiter and be ejected for every one that is scattered on to an impact trajectory, and that is without consideration for whether that impact trajectory would have the correct relative velocity. By comparison, all of the terrestrial planets lie in the ‘accreted’ regime in Wyatt et al. (2017), indicating the ultimate fate of a body scattered by the terrestrial planets is most likely to be accretion.

We thus conclude that the Ćuk & Stewart (2012) scenario is a considerably less likely event than any of the lower impact velocity scenarios. Similarly, this conclusion applies to the Rufu et al. (2017) scenario, in which each individual impact that generates a disc primarily composed of Earth material is a less likely event than any single lower-velocity Moon-formation scenario. Furthermore, of the moderate-to-high velocity scenarios of Rufu et al. (2017),  $V_{\text{imp}} \sim 2V_{\text{esc}}$ , which allow for predominantly Earth material to be distributed into the disc, these collisions occur primarily in head-on configurations. This impact configuration is at the tail end of the impact angle distribution for a giant impact and is not as likely to occur as more probable configurations, at impact angles near  $45^\circ$ , which contribute upwards of 30–60 per cent of impactor material to the proto-Lunar disc in their models, causing it to become isotopically distinct from Earth.

Our prediction that Theia is unlikely to have originated from inside around 0.7 au is in agreement with the detailed investigation of QL15. But more distant bodies are allowed; in particular, the water-rich impactor suggested by Reufer et al. (2012) is consistent with the range of pre-impact orbits allowed, given the relative velocity specified in their scenario. Determining the relative likelihoods of the three low impact velocity scenarios, including hit and run, requires more detailed investigation with  $N$ -body simulations, like that of QL15, who suggest that the Canonical and Canup (2012) scenarios are the most intrinsically likely scenarios due to their impact velocity and impact angle.

One notable factor that we predict but that is missing in the  $N$ -body study of QL15 is the pre-impact inclination. Aside from a small subset with initial inclinations of  $0.66^\circ$  for Theia the simulations of QL15 all begin with Theia co-planar to Earth. Given the non-zero inclinations, we predict for the pre-impact orbit and the close match between the edge of the region of successes and pseudo-successes found by QL15 and our pre-impact orbit distribution, the effect of non-zero initial inclinations would be an interesting avenue to investigate in future studies. A non-zero initial inclination for Theia would likely affect the final value of the angular momentum deficit for the terrestrial planet system. Similarly, on the basis of our pre-impact orbit prediction and the trend in the results of QL15 there are likely initial conditions with semimajor axes in the range of 1.1–1.2 au and eccentricities greater than 0.1 that would result in success or pseudo-success.

## 4 MERCURY FORMATION

We now turn to other giant impact scenarios. Of the four terrestrial planets, Mercury stands out as the most peculiar. Unlike the rest of the terrestrial planets, the composition of Mercury is dominated by its iron core, which comprises around 70 per cent of the mass of the planet (Hauck et al. 2013). This is far higher than the close to chondritic iron content of the other terrestrial planets, which are also (aside from the Moon) significantly more volatile rich than Mercury. This strongly suggests that Mercury underwent a unique formation process that caused it to deviate from the path followed by the rest of the terrestrial planets.

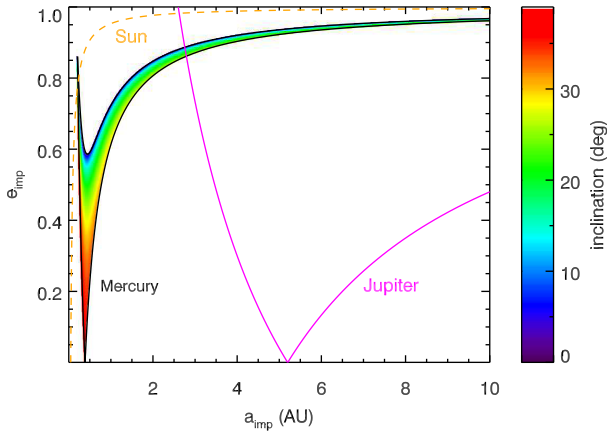
There are two principle classes of proposed methods of achieving the iron enrichment of Mercury: processes that alter the structure of the proto-planetary disc such that Mercury forms from a population of iron-rich dust or planetesimals, and processes that alter the structure of Mercury after it has formed to remove the bulk of the silicate mantle. In the class of adjustments to the proto-planetary disc are suggestions that variation in the condensation of minerals in the inner regions of solar nebula would lead to iron-rich conditions (e.g. Ebel & Alexander 2011) and mechanisms to induce radial separation between iron and silicate dust grains/planetesimals (Weidenschilling 1978), including the recent suggestion of photophoresis (e.g. Wurm, Tieloff & Rauer 2012). Removal of silicate mantle from a planetary mass proto-Mercury is mediated by collisions, either sand-blasting by many small impacts (e.g. Svetsov 2011) or by one or more giant impacts (e.g. Benz et al. 2007; Asphaug & Reufer 2014). Here, we will examine the giant-impact Mercury formation scenarios of Benz et al. (2007) and Asphaug & Reufer (2014).

### 4.1 Proto-Mercury as the target

In Benz et al. (2007), the collisional removal of mantle material from proto-Mercury is achieved by impacting it at high speed with a smaller body, resulting in a net loss of mass from the proto-Mercury, primarily by the ejection of mantle silicate materials. The target proto-Mercury has a mass 2.25 times the present mass of Mercury, and in their successful scenarios, the impactor has a mass of 0.375–0.45 Mercury masses, with relative velocities of  $28\text{--}30\text{ km s}^{-1}$ , around six times the escape velocity of the proto-Mercury. They also find one marginally successful case at  $20\text{ km s}^{-1}$  for a directly head-on impact. Although Mercury formation by such a process is sometimes thought of as a ‘mantle-stripping’ event, the final Mercury is less than half the mass of the target, placing this collision in the catastrophic disruption regime (e.g. Leinhardt & Stewart 2012). The final Mercury produced in this case is composed of the gravitationally re-accumulated remnants of the disrupted core of proto-Mercury plus a lesser amount of mantle material.

As this is a very violent impact and the final product contains only a fraction of the total mass of the colliding bodies, the assumption that the orbit of the target will be approximately the same as that of the present-day final body is likely to break down. As such, rather than assuming that the proto-Mercury follows the same orbit as the present-day Mercury, here we will assume that it follows a circular orbit at the present-day semimajor axis of Mercury. Although it is not clear that this would adequately approximate the true orbit of proto-Mercury, it does seem likely that proto-Mercury would have had a less excited orbit than the remnant of the giant impact that is the present-day Mercury. At a semimajor axis of 0.387 au, the relative velocity of  $30\text{ km s}^{-1}$  translates into  $v_{\text{rel}}/v_{\text{k}} = 0.63$ . We show the resulting distribution of allowed impactor orbits in Fig. 6.





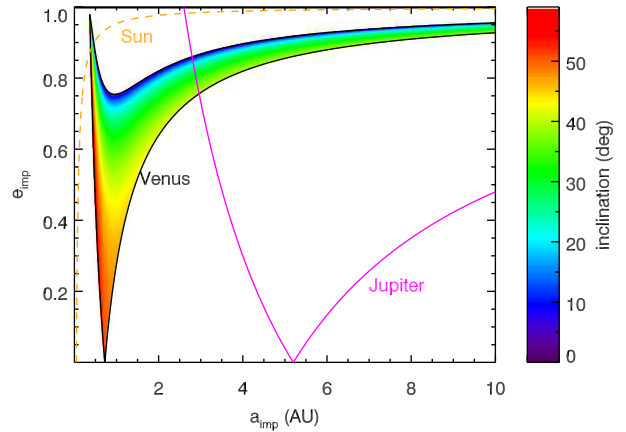
**Figure 6.** Allowed semimajor axis – eccentricity – inclination distribution for impactors striking a proto-Mercury at  $v_{\text{rel}} = 30 \text{ km s}^{-1}$  per the Benz et al. (2007) model. We assume that proto-Mercury has a circular orbit at 0.387 au. Orbits above the magenta line are Jupiter crossing, while those above the orange line are Sun grazing (passing within  $10 R_{\oplus}$  of the Sun).

#### 4.2 Proto-Mercury as the impactor

There are two main problems with the catastrophic disruption scenario where Mercury is the target. First, after the collision the newly formed Mercury would be orbiting through a massive torus of silicate debris, its catastrophically disrupted mantle, most of which continues orbiting the Sun. One must ensure that Mercury does not accrete too much of this material or else the core fraction will become too low. Secondly, the event is more intensely energetic than the formation of Earth’s Moon; thus, Mercury would be expected to be at least as devolatilized as the Moon, when in fact it is comparatively rich in volatiles (Paige et al. 2013) and indeed Earth- and Mars-like in semivolatiles (Peplowski et al. 2011).

Asphaug & Reufer (2014) propose an alternative giant impact scenario for the formation of Mercury. Rather than the proto-Mercury being struck by a smaller impacting body that results in its disruption and the loss of its mantle, proto-Mercury is itself the impactor and collides in a hit-and-run impact with a larger body, either proto-Venus or proto-Earth. This hit-and-run scenario solves the problem of mantle re-accumulation by Mercury, since proto-Venus or proto-Earth dominate the accretion of most of the mantle material liberated from proto-Mercury by the giant impact. Of course, the hit-and-run scenario introduces a new problem in that the newly formed Mercury will still be on an orbit that crosses one or both of the orbits of Venus and Earth and is reliant on other dynamical interactions, perturbing it away from further collisions. This can happen (Chambers 2013), and although most such Mercuries would be accreted by Venus or Earth, those are gone (now part of Venus or Earth), so Asphaug & Reufer (2014) argue that the probabilities of survival of the core of a single hit-and-run proto-Mercury is consistent.

Although the hit-and-run scenario is conceptually very different from the Benz et al. (2007) scenario, the relative velocity required in the two cases is actually quite similar. The best fit found by Asphaug & Reufer (2014) for producing Mercury in a single hit-and-run event uses a relative velocity of around  $30 \text{ km s}^{-1}$ , and they find that all successful single hit-and-run scenarios have relative velocities of  $>25 \text{ km s}^{-1}$ . Since the orbital velocity at Venus or Earth is lower than at the semimajor axis of Mercury, this actually results in a higher  $v_{\text{rel}}/v_k$  of  $>0.71$  if the target is Venus or  $>0.84$  if the target is Earth, assuming that the proto-Venus and proto-Earth



**Figure 7.** Allowed pre-impact orbits for proto-Mercury impacting Venus at  $3 v_{\text{esc}}(\text{Venus})$  as suggested by Asphaug & Reufer (2014) for a single hit-and-run collision. Orbits above the magenta line are Jupiter crossing, while those above the orange line are Sun grazing.

in this scenario have orbits close to their present-day ones, as is likely since they are near their present-day masses ( $0.85 M_{\oplus}$ ).

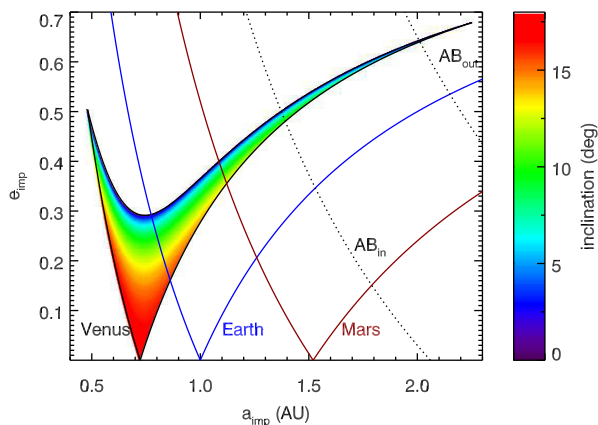
We show the orbital distribution for the pre-impact proto-Mercury for an impact with Venus at  $3 v_{\text{esc}}$  ( $v_{\text{rel}} = 29.3 \text{ km s}^{-1}$ ,  $v_{\text{rel}}/v_k = 0.84$ ) in Fig. 7. As we can see the distribution is broader than that in Fig. 6 due to the higher  $v_{\text{rel}}/v_k$  at Venus. If the target is Earth, then the distribution becomes even more extreme as  $v_k$  drops further and the escape velocity of Earth is slightly higher, such that a  $3 v_{\text{esc}}$  impact at Earth requires a relative velocity above the orbital velocity of Earth,  $v_{\text{rel}}/v_k = 1.06$ . This leads to retrograde orbits for proto-Mercury in the upper left of the distribution.

Though Asphaug & Reufer (2014) demonstrated a successful scenario for the production of Mercury in a single hit-and-run impact with either proto-Earth or proto-Venus, their preferred scenario is for the proto-Mercury to undergo two or three hit-and-run collisions with proto-Earth and/or proto-Venus, losing a fraction of its mantle in each collision and progressively increasing the core–mantle mass ratio. They argue for multiple hit-and-runs statistically, on the grounds that a hit-and-run survivor is likely to have survived more than one such collision, and geochemically, on the grounds that two or more individual impacts would each be less violent, with lower relative velocities than the single hit-and-run scenario, thus preserving more volatiles. Asphaug & Reufer (2014) do not identify a specific impact sequence, but generalizing from previous hit-and-run results (Asphaug 2010), they argue that two to three impacts at a relative velocity roughly equal to the escape velocity of the target (proto-Earth or proto-Venus) would be sufficient to increase the core–mantle mass ratio of an initially chondritic proto-Mercury to what is observed today.

In Fig. 8, we show the allowed distribution of pre-impact orbits for proto-Mercury for an impact with Venus at  $v_{\text{rel}} = v_{\text{esc}}(\text{Venus})$ . In contrast with both the single hit-and-run model and the Benz et al. (2007) model, the allowed distribution of pre-impact orbits is now much smaller and fully contained within the terrestrial planet system with no Jupiter-crossing region. If the target is changed to Earth, then a small region of Jupiter-crossing orbits for proto-Mercury is allowed.

#### 4.3 Discussion of Mercury-formation scenarios

The relative velocities required for both classes of single impact scenarios, those with proto-Mercury as the target and those with



**Figure 8.** Allowed pre-impact orbits for proto-Mercury impacting Venus at  $v_{\text{rel}} = v_{\text{esc}}(\text{Venus})$ , as suggested by Asphaug & Reufer (2014) for producing Mercury through a sequence of 2-3 hit-and-run impacts. Orbits above the blue line are Earth crossing, while those above the dark red line are Mars crossing. The black dotted lines mark the approximate boundaries at which the impactor would be crossing the inner and outer edges of the asteroid belt.

proto-Mercury as the impactor, are similar at around  $30 \text{ km s}^{-1}$  in both cases. This is also similar to the relative velocity in the Moon-formation scenario of Čuk & Stewart (2012) (Section 3.3). As such, both scenarios result in similarly broad distributions of allowed orbits for the impactor and fall into the same high-velocity category. The distribution for the Benz et al. (2007) proto-Mercury-as-target scenario is slightly narrower than that for the Asphaug & Reufer (2014) proto-Mercury-as-impactor scenario, which is in turn slightly narrower than that for the Čuk & Stewart (2012) Moon-formation scenario, since the higher orbital velocities closer to the Sun translate into lower  $v_{\text{rel}}/v_k$  for the same  $v_{\text{rel}}$ .

Occupying the same high relative velocity category as the Čuk & Stewart (2012) Moon-formation scenario, many of the same conclusions regarding the relative likelihood of that scenario also apply to the single impact Mercury formation scenarios. Achieving the relative velocities necessary without the influence of Jupiter is difficult, and outer Solar system origins for the impactor are possible. As previously, if the impactor originates in the outer Solar system this places tight constraints on the pre-impact inclination to allow for both scattering with Jupiter and collision with the relevant terrestrial planet.

If the impactor did indeed originate in the outer Solar system then this has different implications for the model of proto-Mercury as the target versus proto-Mercury as the impactor. For the proto-Mercury-as-target model, an origin of the impactor in the outer Solar system may be beneficial in the same way as for the Čuk & Stewart (2012) and Reufer et al. (2012) Moon-formation models. An outer Solar system impactor would likely be water rich, which would decrease the mass of silicate material in the impactor and reduce the problem of avoiding re-accretion of the silicate material. Although a water-rich impactor could potentially introduce a new problem of avoiding the re-accretion of water, it will be vapourized and dispersed by the energetic giant impact leaving behind a situation analogous to the formation of Earth's Moon. In addition, it would be difficult for the low-mass Mercury to retain substantial amounts of water in its thermal and radiation environment in the long term, albeit that small amounts can be retained in shadowed regions at the poles (Lawrence et al. 2013; Paige et al. 2013).

For the single hit-and-run origin of Mercury, however, an origin for the impactor (in this case proto-Mercury itself) in the outer Solar

system, which would thus likely be water rich, presents a more problematic issue. In a water-rich proto-Mercury, the iron fraction of the total body would be smaller, presuming a chondritic silicate-iron ratio with an addition of water. This increases the required total mass of the body if it is to contain the same mass of iron. In addition, for a fully differentiated body with a water ice outer mantle, a rocky inner mantle and an iron core, the silicates and iron would now be more strongly concentrated at the centre of the body, with consequently higher (negative) total binding energy, and it would thus require more energy to remove the silicates overlying the iron. The increased concentration of the iron towards the centre of the body, combined with the larger mass required to obtain the same total mass of iron, both suggest that the impact would need to be more violent than originally proposed by Asphaug & Reufer (2014).

Although such an energetic hit-and-run would be plausible for a proto-Mercury scattered by Jupiter, it raises questions about the consequences for Earth or Venus if they were struck with such intensity. For example, mantle stripping of the target (Earth or Venus) could become significant, as could the loss of volatile elements from a very strongly shock heated and highly deformed target planet. Conversely, as it is the target that accumulates most of the mantle ejected from proto-Mercury in the hit-and-run scenario, the hit and run by a water-rich impactor would additionally contribute a late complement of outer Solar system water on to the target, in addition to the stripped mantle silicates. These complexities do not rule the scenario out – Venus and Earth evolved quite differently for some reason – but raises the importance of understanding the implications for the target of the hit-and-run scenario.

In the multiple hit-and-run Mercury-as-impactor scenario proposed by Asphaug & Reufer (2014), while proto-Mercury could have been water rich if it originated in the asteroid belt, this is not a necessary condition. Furthermore, a relative velocity of around the escape velocity of Earth or Venus is readily expected for an impact involving a body that has been scattered by the terrestrial planets, and indeed Jupiter-crossing orbits are not allowed if Venus is the target.

Although a scenario involving a chain of impacts might seem at first glance unlikely, the individual impacts are fairly representative of the impacts one would expect during terrestrial planet formation. Kokubo & Genda (2010) and Chambers (2013) find that around half of all impacts are hit and run, and so a chain of several hit and runs is not especially unlikely, for a given unaccreted embryo. According to Asphaug & Reufer (2014), the accreted embryos are biased to include all of the slow, direct hits, causing the unaccreted embryos to be biased to include an increasing fraction of hit-and-run survivors. Indeed in one of the handful of cases in Chambers (2013) they find a Mercury-sized survivor of four successive hit-and-run collisions.

The single impact scenarios however require rather extreme events. Following the same arguments as for the Čuk & Stewart (2012) Moon-formation model in Section 3.6, the single impact scenarios are thus unlikely and on the basis of our work we suggest that the multiple hit-and-run scenario of Asphaug & Reufer (2014) is statistically more favourable, although requiring more detailed study in terms of Solar system dynamics and collision geophysics.

## 5 THE BOREALIS BASIN AND ECCENTRIC TARGETS

A dominant characteristic of Mars is the hemispheric crustal dichotomy, with the Northern hemisphere lying on average 5 km lower

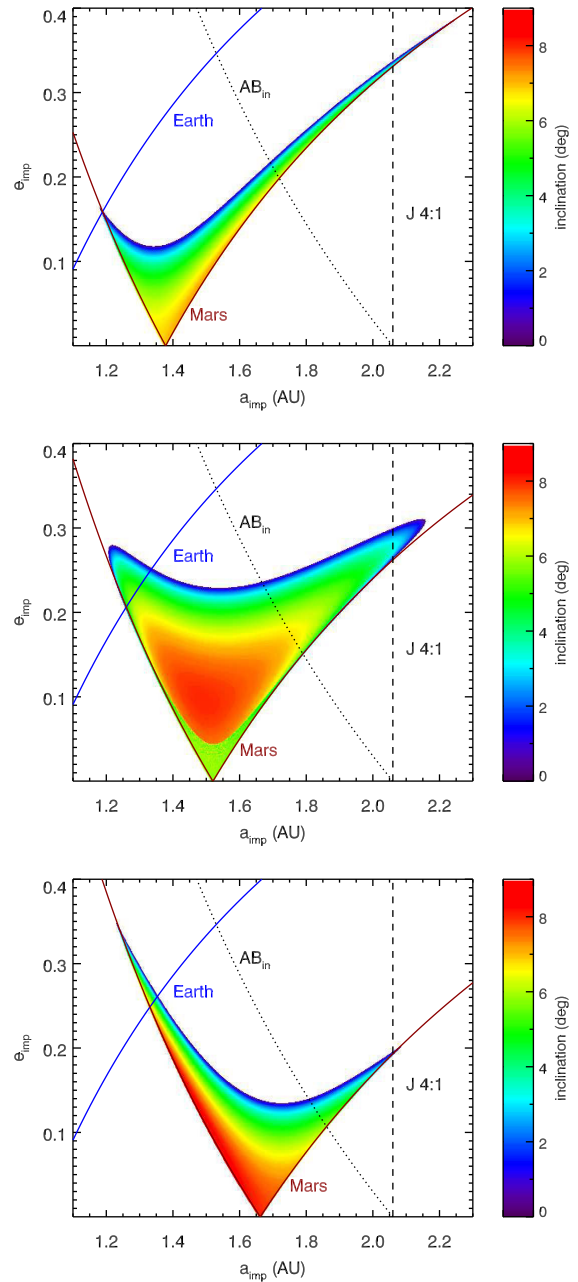
than the Southern hemisphere (e.g. Smith et al. 1999). It has long been recognized that a massive impact might be able to reproduce the hemispheric dichotomy (e.g. Wilhelms & Squyres 1984). More recently, improvements in our knowledge of the gravity field of Mars (Andrews-Hanna, Zuber & Banerdt 2008) and numerical modelling of giant impacts into Mars (Marinova et al. 2008; Nimmo et al. 2008; Marinova et al. 2011) has shown that a giant impact indeed seems to well reproduce the hemispheric dichotomy and fits well with other features of Mars, such as its magnetic field. Additionally, Citron, Genda & Ida (2015) demonstrate that a Borealis basin-forming collision places enough mass into a circum-Mars disc to produce a proto-lunar disc from which the Martian moons Phobos and Deimos could derive. Under this scenario, the hemispheric dichotomy is then the scar of a giant impact involving a planetesimal around twice the mass of Ceres.

The Borealis basin impact also represents a useful demonstration case, since the eccentricity of the Martian orbit is significant and so must be taken into account when examining the potential pre-impact orbit of the Borealis impactor. As we can see from the equations in Section 2, the key difference is that when the target has an eccentric orbit the range of impactor orbital parameters that will produce the correct relative velocity now depends on the true anomaly of the target at the time of impact, since the velocity of the target varies as it moves around its orbit. Since we cannot easily represent the variation in the allowed impactor orbit distribution as a function of the true anomaly of Mars at impact over the whole range of true anomaly, we instead chose three representative locations, when Mars is at apocentre, pericentre and when the Sun–Mars separation is equal to the Martian semimajor axis ( $r = a$ ). Apocentre and pericentre represent the extremes, when Mars is moving slowest and fastest, respectively, while  $r = a$  provides a useful middle ground, since at this location, the velocity of Mars is equal to that of a body on a circular orbit at  $a_{\text{Mars}}$ . These three cases are shown in the top, middle and bottom panels of Fig. 9 – for which we use the best-fitting scenario of Marinova et al. (2011) in which a roughly 2000 km body impacts Mars in a glancing collision at around  $1.2 v_{\text{esc}}$  or  $v_{\text{imp}} = 6 \text{ km s}^{-1}$ . For the Martian escape velocity of  $5.027 \text{ km s}^{-1}$ , this corresponds to  $v_{\text{rel}}/v_k = 0.14$  or  $v_{\text{rel}} = 3.33 \text{ km s}^{-1}$ .

As we would expect from a low velocity impact the pre-impact orbits for the Borealis impactor lie close to Mars. Regardless of where on the Martian orbit the impact takes place, a significant region of impactor orbits that cross into the asteroid belt is allowed due to the proximity of the asteroid belt to Mars. In contrast, Earth-crossing orbits are allowed only when the impact takes place away from the pericentre of the Martian orbit. This may seem counter intuitive because Mars is closest to Earth at pericentre, but is due to the faster orbital speed of Mars at this point. Because the allowed range for pre-impact orbits that also cross the orbit of Earth is small, this suggests that it is unlikely that the impactor would have been passed outwards by the other terrestrial planets. As such we recommend that any future detailed  $N$ -body examination of the origin of the impactor proposed for the Borealis basin should focus on the range from 1.2 to 2.2 au near Mars.

## 6 OTHER SOLAR SYSTEM GIANT IMPACTS

The terrestrial planets are not the only places where there are features that are best explained by a giant impact. In the outer reaches of the Solar system, a giant impact has also been proposed as the origin for the Pluto–Charon system (e.g. Canup 2005, 2011; Stern et al. 2006). The successful simulations of Canup (2011) had low impact velocities of  $< 1.2 v_{\text{esc}}$ ; however, translating this into



**Figure 9.** Allowed semimajor axis – eccentricity – inclination distributions for the pre-impact orbit of the Borealis impactor (assuming the best-fitting parameters of Marinova et al. 2011) for an impact occurring when Mars is at pericentre (top), at an orbital distance of 1.52 au (equal to the Martian semimajor axis, middle), and at apocentre (bottom). Orbits above the blue line are Earth crossing, while those above the dotted line have apocentres beyond 2.06 au, the inner edge of the asteroid belt. The black dashed line indicates the 4:1 mean motion resonance with Jupiter.

predictions for the pre-impact orbit of the impactor is difficult since the orbit of proto-Pluto at the time of impact is highly uncertain. Today, Pluto–Charon is in a 3:2 mean-motion resonance with Neptune, a result of Neptune’s outward migration into the Kuiper belt. It is likely that the formation of Charon occurred before or during the migration of Neptune as the primordial Kuiper belt would have been much more massive with many more potential impactors (e.g. Tsiganis et al. 2005). As such, the orbit of proto-Pluto at the time of impact was likely not the same as Pluto–Charon today. It is however

worth noting that, whatever the orbit of proto-Pluto at the time of impact, the low impact velocity effectively requires that the impactor must have originated from the same population as proto-Pluto. If the impact occurred after capture into the present, eccentric, orbit of Pluto–Charon, then impactors with eccentricities below around 0.1 can be excluded, indicating that the impactor would be another resonant or scattered disc object. Conversely, if the impact occurred before capture into resonance then the impactor would have been a low eccentricity object from nearby. If future constraints, for example on the timing of the Pluto–Charon-forming impact, allow us to estimate what the orbit of proto-Pluto would have been at the time of impact, then we would be able to distinguish between these possibilities.

Smaller scale events are evident in the creation of the Hirayama asteroid families (e.g. Durda et al. 2007). In particular, the Vesta family is believed to be traceable to the formation of the massive Rheasilva basin  $\lesssim 1$  Gyr ago (Binzel et al. 1997). This basin dominates the Southern hemisphere of Vesta and models suggest it was caused by an impactor around 66 km in diameter (Jutzi et al. 2013), large compared with the 525 km mean diameter of Vesta itself. Similarly in the outer Solar system, the dwarf planet Haumea also has a collisional family associated with it (Brown et al. 2007), which was presumably also created in a large impact, perhaps with a precursor of the current two satellites of Haumea (Schlichting & Sari 2009; Ćuk, Ragozzine & Nesvorný 2013). It has also long been recognized that a large impact could explain the large obliquity of Uranus (e.g. Safronov 1972; Parisi & Brunini 1997).

## 7 CONCLUSIONS

The impact velocity is one of the key parameters in any giant impact scenario and we have demonstrated how this can be used to provide constraints on the orbit of the impactor immediately prior to the collision. These constraints are very fast to compute, taking only a few seconds on a laptop computer. This allows us to quickly compare a wide range of giant impact scenarios in a uniform way and make broad judgements on the likelihood of different scenarios as well as narrow down the range of initial orbital parameters within which more detailed examinations with expensive  $N$ -body simulations should be targeted. We apply this method to examine the currently proposed scenarios for giant impacts in the Solar system for the formation of the Moon, Mercury and the Borealis basin on Mars. Any subsequent proposals for giant impact scenarios in the Solar system can then be compared in the same manner.

The Ćuk & Stewart (2012) scenario for the formation of the Moon, both current scenarios for the formation of Mercury through a single impact (Benz et al. 2007; Asphaug & Reufer 2014) and the multiple impact Moon-formation scenario of Rufu et al. (2017) all have similar distributions of pre-impact orbits and can be grouped together as high-velocity impacts. For all of the scenarios in this category achieving the required impact velocity through scattering with the terrestrial planets is unlikely, but can be readily achieved by scattering with Jupiter. Scattering of an object by Jupiter on to an impact trajectory with a terrestrial planet is still a low probability event however and so we disfavour all of the scenarios in this high impact velocity category. If we rule out these high-velocity scenarios, this means that in the case of Mercury, the only remaining scenario we consider favourable from a probabilistic perspective is the multiple hit-and-run proposal of Asphaug & Reufer (2014).

Although we cannot distinguish the three low-velocity scenarios for the formation of the Moon in terms of relative probability, we can predict the range of orbital parameters from which we expect the

impactor to have originated. Comparing these predictions with the detailed  $N$ -body study of Quarles & Lissauer (2015) demonstrates that our predicted parameter ranges do indeed encompass the origins of successful impactors. As such we recommend that any future  $N$ -body examination of the origin of the Borealis basin impactor should focus on the range from 1.2 to 2.2 au near Mars.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge funding through NASA grant NNX16AI31G (Stop hitting yourself). The results reported herein benefited from collaborations and/or information exchange within NASA's Nexus for Exoplanet System Science (NExSS) research coordination network sponsored by NASA's Science Mission Directorate. The authors are grateful to Billy Quarles for providing the data from fig. 2d of QL15 for use in Fig. 4. APJ thanks Dan Tamayo for useful discussions. The authors thank the anonymous referee for insightful suggestions that have improved the manuscript.

## REFERENCES

- Andrews-Hanna J. C., Zuber M. T., Banerdt W. B., 2008, *Nature*, 453, 1212
- Anic A., 2006, PhD thesis, Univ. Bern, <http://space.unibe.ch/fileadmin/media/pdf/taps/diss-augustina.pdf>
- Asphaug E., 2010, *Chemie der Erde/Geochemistry*, 70, 199
- Asphaug E., 2014, *Annu. Rev. Earth Planet. Sci.*, 42, 551
- Asphaug E., 2017, in Elkins-Tanton L. T., Weiss B. P., eds, *Planetesimals: Early Differentiation and Consequences for Planets*. Cambridge Univ. Press, p. 16
- Asphaug E., Reufer A., 2014, *Nature Geosci.*, 7, 564
- Benz W., Anic A., Horner J., Whitby J. A., 2007, *Space Sci. Rev.*, 132, 189
- Binzel R. P., Gaffey M. J., Thomas P. C., Zellner B. H., Storrs A. D., Wells E. N., 1997, *Icarus*, 128, 95
- Brown M. E., Barkume K. M., Ragozzine D., Schaller E. L., 2007, *Nature*, 446, 294
- Cameron A. G. W., Ward W. R., 1976, *Lunar and Planetary Science Conference Abstract*, p. 120
- Canup R. M., 2004a, *ARA&A*, 42, 441
- Canup R. M., 2004b, *Icarus*, 168, 433
- Canup R. M., 2005, *Science*, 307, 546
- Canup R. M., 2008, *Icarus*, 196, 518
- Canup R. M., 2011, *AJ*, 141, 35
- Canup R. M., 2012, *Science*, 338, 1052
- Canup R. M., Asphaug E., 2001, *Nature*, 412, 708
- Chambers J. E., 2007, *Icarus*, 189, 386
- Chambers J. E., 2013, *Icarus*, 224, 43
- Citron R. I., Genda H., Ida S., 2015, *Icarus*, 252, 334
- Clayton R. N., 2003, *Space Sci. Rev.*, 106, 19
- Ćuk M., Stewart S. T., 2012, *Science*, 338, 1047
- Ćuk M., Ragozzine D., Nesvorný D., 2013, *AJ*, 146, 89
- Durda D. D., Bottke W. F., Nesvorný D., Enke B. L., Merline W. J., Asphaug E., Richardson D. C., 2007, *Icarus*, 186, 498
- Ebel D. S., Alexander C. M. O., 2011, *Planet. Space Sci.*, 59, 1888
- Halliday A. N., 2000, *Earth Planet. Sci. Lett.*, 176, 17
- Hartmann W. K., Davis D. R., 1975, *Icarus*, 24, 504
- Hauck S. A. et al., 2013, *J. Geophys. Res. (Planets)*, 118, 1204
- Jackson A. P., Wyatt M. C., 2012, *MNRAS*, 425, 657
- Jackson A. P., Wyatt M. C., Bonsor A., Veras D., 2014, *MNRAS*, 440, 3757 (J14)
- Jutzi M., Asphaug E., Gillet P., Barrat J.-A., Benz W., 2013, *Nature*, 494, 207
- Kaib N. A., Cowan N. B., 2015, *Icarus*, 252, 161
- Kenyon S. J., Bromley B. C., 2006, *AJ*, 131, 1837
- Kokubo E., Genda H., 2010, *ApJ*, 714, L21
- Lawrence D. J. et al., 2013, *Science*, 339, 292
- Leinhardt Z. M., Stewart S. T., 2012, *ApJ*, 745, 79



- Leinhardt Z. M., Marcus R. A., Stewart S. T., 2010, *ApJ*, 714, 1789
- Marinova M. M., Aharonson O., Asphaug E., 2008, *Nature*, 453, 1216
- Marinova M. M., Aharonson O., Asphaug E., 2011, *Icarus*, 211, 960
- Nimmo F., Hart S. D., Korycansky D. G., Agnor C. B., 2008, *Nature*, 453, 1220
- O'Brien D. P., Morbidelli A., Levison H. F., 2006, *Icarus*, 184, 39
- Ozima M., Podosek F. A., Higuchi T., Yin Q.-Z., Yamada A., 2007, *Icarus*, 186, 562
- Pahlevan K., Stevenson D. J., 2007, *Earth Planet. Sci. Lett.*, 262, 438
- Paige D. A. et al., 2013, *Science*, 339, 300
- Parisi M. G., Brunini A., 1997, *Planet. Space Sci.*, 45, 181
- Peplowski P. N. et al., 2011, *Science*, 333, 1850
- Quarles B. L., Lissauer J. J., 2015, *Icarus*, 248, 318 (QL15)
- Quintana E. V., Barclay T., Borucki W. J., Rowe J. F., Chambers J. E., 2016, *ApJ*, 821, 126
- Raymond S. N., O'Brien D. P., Morbidelli A., Kaib N. A., 2009, *Icarus*, 203, 644
- Reufer A., Meier M. M. M., Benz W., Wieler R., 2012, *Icarus*, 221, 296
- Rivera E. J., 2002, PhD thesis, State Univ. New York Stony Brook
- Rufu R., Aharonson O., Perets H. B., 2017, *Nature Geosci.*, 10, 89
- Russell C. et al., 2016, *Science*, 353, 1008
- Safronov V. S., 1972, *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets*. Academy of Sciences, Russian
- Salmon J., Canup R. M., 2012, *ApJ*, 760, 83
- Schlichting H. E., Sari R., 2009, *ApJ*, 700, 1242
- Smith D. E. et al., 1999, *Science*, 284, 1495
- Stern S. A. et al., 2006, *Nature*, 439, 946
- Svetsov V., 2011, *Icarus*, 214, 316
- Thomas P. C., Parker J. W., McFadden L. A., Russell C. T., Stern S. A., Sykes M. V., Young E. F., 2005, *Nature*, 437, 224
- Touma J., Wisdom J., 1994, *AJ*, 108, 1943
- Tsiganis K., Gomes R., Morbidelli A., Levison H. F., 2005, *Nature*, 435, 459
- Weidenschilling S. J., 1978, *Icarus*, 35, 99
- Wilhelms D. E., Squyres S. W., 1984, *Nature*, 309, 138
- Wisdom J., Tian Z., 2015, *Icarus*, 256, 138
- Wurm G., Tieloff M., Rauer H., 2012, *Meteorit. Planet. Sci. Suppl.*, 75, 5300
- Wyatt M. C., Bonsor A., Jackson A. P., Marino S., Shannon A., 2017, *MNRAS*, 464, 3385
- Young E. D., Kohl I. E., Warren P. H., Rubie D. C., Jacobson S. A., Morbidelli A., 2016, *Science*, 351, 493

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