The Origin of Planetary Impactors in the Inner Solar System

Robert G. Strom\textsuperscript{1}, Renu Malhotra\textsuperscript{1}, Takashi Ito\textsuperscript{2}, Fumi Yoshida\textsuperscript{2}, and David Kring\textsuperscript{1}

ABSTRACT

The impact crater record of the terrestrial planets and the Moon is owed to two populations of impactors that are distinct in their size distributions. We show that the ‘old’ population, responsible for an intense period of bombardment that ended 3.8 Gigayears ago, is virtually identical in size distribution to the present main belt asteroids; the second population, responsible for craters younger than 3.8 Gy, matches closely the size distribution of the near earth asteroids. These results confirm that an inner Solar System impact cataclysm occurred 3.9 Gy ago, and also provide compelling new evidence that identifies the main asteroid belt as the source of the impactors. A plausible dynamical explanation is that many asteroids were ejected from the main belt on a dynamical timescale by sweeping gravitational resonances during an epoch of orbital migration of the giant planets in early Solar System history. The impactors in the inner Solar System over the past 3.8 Gy are also derived from the asteroid belt, but by a combination of chaotic gravitational resonances and non-gravitational processes; the latter process being size-dependent, yields a size distribution quite distinct from the purely gravitational process that caused the cataclysm 3.9 Gy ago. Several important implications for the evolution of asteroids, for planetary geology, and for the origin and history of the Solar System follow from these results.

1. Introduction

The Moon and all the terrestrial planets were resurfaced during a period of intense impact cratering that occurred between the time of their accretion, \( \sim 4.5 \) Gigayears ago (Ga), and the time of formation of the Orientale multi-ring basin on the Moon, the last great basin-forming impact event, \( \sim 3.85 \) Ga. The lunar cratering record and the radiometrically dated Apollo samples have shown that the intense bombardment of the Moon ended at \( \sim 3.85 \) Ga; the impact flux since

\textsuperscript{1}Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ 85721, USA.

\textsuperscript{2}National Astronomical Observatory, Osawa, Mitaka, Tokyo 181-8588, Japan.

Submitted for publication to Science, April 12, 2005.
Correspondence should be addressed to Renu Malhotra, email: renu@lpl.arizona.edu
that time to the present has been at least an order of magnitude smaller. The 3.85 Ga epoch might represent the final end of an era of steadily declining large impacts (the tail end of the accretion of the planets). However, it has been argued that only a sudden injection of impacting objects into the terrestrial planet zone accounts for the abrupt end of the intense bombardment; thus, this event has been named the Late Heavy Bombardment (LHB), or sometimes the Lunar Cataclysm, to distinguish it from the prior final accretion of the planets at 4.5 Ga. Specifically, the lunar cataclysm hypothesis (1,2) postulates that the intense bombardment of the Moon lasted only a very short period of time, 20-200 My (2–5). Recent results on the impact ages of lunar meteorites (which represent a much broader region of the lunar surface than the Apollo samples) support this hypothesis (6–8). Furthermore, the impact-reset ages of meteoritic samples of asteroids (9–10) and the shock-metamorphosing at 3.92 Ga of the only known sample of the heavily cratered highlands of Mars, meteorite Alan Hills 84001 (11) indicate that the LHB affected the entire inner Solar System, not just the Moon.

Identifying the sources of planetary impactors has proven elusive. Dynamical models invoking both geocentric and heliocentric debris and both asteroidal and cometary reservoirs have been proposed (12), but recent chemical analyses of Apollo impact melts point to a dominantly asteroidal reservoir for the lunar cataclysm (10). In this paper, we provide compelling new evidence that the source of the LHB impactors was the main asteroid belt, that the dynamical mechanism that caused the LHB was unique in the history of the Solar System and distinct from the processes producing the flux of objects that currently hit planetary surfaces. A potentially unifying theory that explains many features of the outer Solar System as well as the LHB event, is to be found in the recent recognition that the giant planets underwent an epoch of orbital migration early in their history (13–18). We point out several important implications for the evolution of small body reservoirs in the solar system, as well as for planetary geology and for the origin and history of the Solar System.

2. The Terrestrial Planet Cratering Record

It is often implied that the crater size distribution has been the same throughout Solar System history, and that it is found on all rocky planet surfaces regardless of age (e.g., ref. 17). However, examining the oldest surfaces on the Moon, Mercury, and Mars, one finds a cratering record that is distinct in its size distribution from the crater record on the younger areas on those planets and on Venus.

Figure 1 shows the crater size distributions (20) of various surface units on the Moon, Mars and Mercury, based on published data in (21–22) supplemented by new crater counts. The uppermost curves represent the most heavily cratered surface units, and the lower curves are for progressively less densely cratered surfaces. In the progression from the uppermost curves to the lowermost curves, we can recognize a qualitative change in the shapes of the curves. We discuss these in more detail below.
Heavily Cratered Surfaces

Of the terrestrial planets, only the Moon, Mercury and Mars have heavily cratered surfaces. The heavily cratered highlands all have complex size distributions, as seen in Fig. 1. For the Moon, the curve slopes downward to the left (with a differential power law of index \( p = -2.2 \)) at diameters less than about 50 km, a nearly flat \( (p = -3) \) portion between 50 and 100 km, and sloping downward to the right \( (p \approx -4) \) at diameters between about 100 km to 300 km. The curve may slope upward at diameters greater than about 400 km, but the statistics are too poor at these large diameters to be sure. The curves for Mercury and Mars are steeper than the lunar curve at diameters less than about 40 km. This is due to the obliteration of a fraction of the craters by intercrater plains formation. Intercrater plains are common on Mercury and Mars, but are very scarce on the Moon. Therefore, the lunar highlands curve best represents the shape of what we shall call the Population 1 crater size distribution.

In addition to the highlands, the crater curves for Martian old plains east of the Tharsis region, old plains within the Helles basin and plains within and surrounding the Caloris basin on Mercury also have the same shape as the lunar highland curve over the same diameter range, but with a lower crater density \((23)\). The similar shape to the lunar highlands curve strongly indicates that Mars’ and Mercury’s highlands have been affected by intercrater plains formation, but not the Moon. The lower crater densities imply that these older plains probably formed near the tail end of the LHB, about 3.8 billion years ago.

Young Lightly Cratered Plains

The lowest crater curves in Fig. 1 are flat and distinctly different than the heavily cratered surfaces described above. These include the lightly cratered, hence younger, plains on Mars and the Moon, as well as Class 1 craters on the Moon. (Class 1 craters are relatively fresh craters with non-degraded morphologies, sharp rims, and well-defined ejecta blankets; all post-Mare craters are Class 1 craters.)

The Northern plains on Mars have a crater size distribution that is a horizontal line on the R plot, from about 100 km down to at least 2 km diameter. This size distribution is characteristic of what we shall call the Population 2 craters.\(^3\) There is also evidence for Population 2 on the Moon. The Class I craters, which are distributed all over the Moon, have a flat distribution on the R plot, similar to the young plains on Mars. The post-mare crater curve has a saw-tooth shape characteristic of poor statistics; however, within the large error bars the distribution is consistent with Population 2 but inconsistent with the lunar highlands curve.

Figure 2 shows the crater size distribution on Venus compared to the Northern Plains of Mars. Notice that the crater density on Venus is about an order of magnitude less than on Mars. Only

---

\(^3\)At diameters less than about 1 to 1.5 km the curve steepens sharply to a \( p \approx -4 \) distribution. This part of the crater size distribution is dominated by secondary impacts and does not represent the crater production size distribution \((24)\).
young craters are present because of multiple global resurfacing events (25). Furthermore, the Venus crater curve at smaller diameters is severely affected by atmospheric screening due to Venus’ thick 90 bar atmosphere (26); this explains the sharp downturn at diameters less than about 25 km. At larger diameters it is a horizontal line on the R plot, similar to Population 2 on the Moon and Mars shown in Figure 1.

Part of the Venus crater population consists of clusters of craters (multiples) that result from fragmentation of the impacting object in the dense atmosphere. These comprise 16% of all Venus craters. The size distribution of these multiples is also shown in Fig. 2, where the diameter is derived from the sum of the crater areas in the cluster. Multiples are probably due to stronger projectiles that could resist atmospheric disintegration better than most other impacting objects, but still weak enough that they broke up in the atmosphere. The turnover of the curve for multiple craters does not occur until diameters less than 9 km; at larger diameters the curve is flat. This, together with the much lower crater density, strongly suggests that the impacting population on Venus was the same as Population 2 on the Moon and Mars. It is also evidence that the turnover of the crater curve is indeed due to atmospheric screening.

Figure 3 summarizes the two characteristic shapes of the crater curves in the inner Solar System, those of the heavily cratered highlands and old cratered plains, and those of the younger lightly cratered plains. (The Venus curve is a composite of the production population for all craters greater than 9 km, including multiples in the range of 9–25 km diameter.) We conclude that the terrestrial planets have been impacted by two populations of objects that are distinguishable by their size distributions. Population 1 is responsible for the LHB, and Population 2 is responsible for the post-LHB period and up to the present time.

3. Projectile Size Distributions

A number of studies on the physics of impact cratering on solid bodies have derived projectile-crater scaling laws. We used the Pi scaling law (27) to derive the projectile size distribution for Population 1 and Population 2 impactors. We used the lunar highland crater curves as representative of Population 1, and the Martian young plains as representative of Population 2, as these provide the best crater statistics. (We did not use the data for diameters greater than 500 km, because statistics there is very poor.) We assumed projectile parameters appropriate for asteroidal impacts: density of 3 g cm$^{-3}$ (similar to basaltic rock), an average impact angle of 45° and average impact velocities of 17 km s$^{-1}$ and 7 km s$^{-1}$ on the Moon and on Mars, respectively. The resulting projectile size distributions are shown in Fig. 4.

Also shown in Fig. 4 are recent determinations of the size distributions of the Main Belt asteroids (MBAs) and Near Earth Asteroids (NEAs). (The vertical positions of these curves are arbitrary, and were selected for clarity of comparison with the projectile size distributions.) The MBA size distributions come from three sources: the Subaru main belt survey (28), the Spacewatch
survey (29), and the Sloan Digital Sky Survey (30). The NEA size distribution is from the bias-corrected LINEAR survey (31). (Detailed descriptions of the techniques used in deriving the size distributions from the observational counts are given in the references cited.)

The comparison of the Population 1 and Population 2 projectile size distributions with the main asteroid belt and near earth asteroid populations leads to several significant conclusions.

First, the size distribution of the current MBAs is virtually identical to the Population 1 projectile size distribution. This was also pointed out in reference (32). It strongly indicates that the objects responsible for the LHB originated from Main Belt asteroids. (Unless comets or Kuiper belt objects have the same size distribution, these objects could not have been significant contributors to the LHB.) This conclusion is consistent with the orbital constraints imposed by a comparison of Population 1 crater curves of the Moon, Mars and Mercury (21).

Second, the close match between the current MBA size distribution and that of the LHB projectiles implies that the main asteroid belt has remained unchanged in its size distribution over the past 3.8 Gy. There are two possible interpretations: either collisional processes produced a steady-state size distribution in the main asteroid belt at least as early as 3.8 Ga, or the collision frequency in the main asteroid belt was drastically reduced around 3.8 Ga so that little or no further evolution of its size distribution has occurred since that time.

Third, the LHB mechanism responsible for ejecting asteroids from the main asteroid belt and into terrestrial planet-crossing orbits had to be unique to the early Solar System (the first 700 My of Solar System history), because there is no evidence for any subsequent event of similar magnitude in the inner planets’ cratering history since then.

Fourth, that mechanism had to be one that ejected asteroids from the main belt in a size-independent manner, preserving the MBA size distribution in the inner planet impactor population. This precludes size-dependent non-gravitational transport processes, such as the Yarkovsky effect recently prominent in studies of the origin of NEAs (discussed below), and instead implicates a dynamical process, such as sweeping gravitational resonances, that was largely insensitive to asteroid mass. We discuss this mechanism in more detail in the next section.

Fifth, with regard to Population 2, we see from Fig. 4 that the size distribution of projectiles responsible for the post-LHB crater records (3.8 Ga to the present) is the same as that of the NEAs and quite different from that of the LHB projectiles. The fit between Population 2 and NEAs is remarkably good. This result is contrary to the findings in (33).

A plausible reason that the MBAs and the NEAs have such a different size distribution, even though the latter population is most probably derived from the former, is the Yarkovsky effect which causes secular changes in orbital energy of an asteroid due to the asymmetric way a spinning asteroid absorbs and reradiates solar energy (34). Over a few tens of millions of years the effect is large enough to transport a significant number of sub–20 km size asteroids into strong Jovian resonances (35); the latter then deliver them into terrestrial planet-crossing orbits. The magnitude
of the effect depends on the size of the asteroid: for diameters greater than about 10 m, the smaller the asteroid the larger the effect. This explains why the NEAs (Population 2 projectiles) have relatively more small objects compared to MBAs.

4. Dynamical Mechanism for the Late Heavy Bombardment

That the Population I projectiles have the same size distribution as the MBAs implicates an asteroid transport mechanism which has the following two properties: (a) it preserves the main belt size distribution, and (b) it must be a one-time event that occurred early in Solar System history. The former constraint, rules out physical transport mechanisms that have any significant dependence on the asteroid size, and instead indicates a dynamical (gravitational) transport mechanism that excites asteroid orbits directly into terrestrial planet crossing orbits, independent of asteroid size. The second constraint suggests a non-repeating circumstance in the Solar System. A plausible mechanism that satisfies these constraints is associated with the orbital migration of the outer planets (Jupiter—Neptune), which is thought to have occurred on a timescale of $10^7$–$10^8$ years in the early history of the Solar System (15, 16).

A rearrangement of the orbits of the outer planets would have had significant effects on the inner Solar System. The outer asteroid belt would have been severely depleted due to orbital instabilities that ensue as strong Jovian mean motion resonances sweep across this zone (17); indeed, this is possibly the only viable explanation for the absence of asteroids in a range of semimajor axes where dynamical stability arguments would otherwise argue for a large population of asteroids in long-term stable orbits. Levison et al. (18) suggested that the orbital migration of the outer planets would have caused the Moon and the other terrestrial planets to be intensely bombarded for a period of 10—100 Myr. Such bombardment would have to be considered cataclysmic on a geologic timescale. If giant planet migration is the correct explanation for the LHB, then it must be shown to be consistent both with the timing of the LHB, and with the provenance of the LHB impactors.

The timing of the LHB at 3.8–3.9 Ga provides a strong constraint on the timing of the planet migration process and on the timescale for the formation of Uranus and Neptune. This is because the migration of the giant planets begins only when Neptune and Uranus accrete a significant fraction of their mass, large enough to cause the large-scale mobilization of outer solar system residual planetesimals. The fraction may be on the order of 30% of their present mass, although detailed studies have yet to be done. Thus, one major inference from linking the LHB to the migration of the planets is that Uranus and Neptune accreted approximately 600 Myr after the formation of the other major planets. This timescale constraint has significant implications for theories of Uranus–Neptune formation that have long been plagued with difficulties (36, 37).

Regarding the provenance of the LHB impactors, we note that the giant planet migration process is due to the large-scale scattering of residual outer Solar System planetesimals (which are ice-rich bodies), but it also leads to a large-scale perturbation of inner solar system planetesimals.
(rocky asteroids) due to the sweeping by Jovian mean motion and secular resonances. Thus, in this model, the bombardment of the inner solar system would be due to a combination of icy outer solar system planetesimals and rocky asteroids; uncertainties in the model parameters allow room for different relative abundances of these impactor populations. Our results strongly constrain this model, as we show that the MBAs were the dominant impactors in the inner solar system during the LHB era.

One of the strongest resonances for the transport of MBAs to the terrestrial planets at the present epoch is the secular $\nu_6$ resonance which is associated with the mean apsidal precession rate of Saturn. The $\nu_6$ resonance, located at the inner edge of the main asteroid belt, at 2.1 AU, excites asteroid eccentricities to nearly 1 on a timescale of $10^6$ Myr (38), and is thought to be a major supply route for sun-grazing asteroids as well as NEAs originating in the main asteroid belt (39). The location and dynamics of this resonance is determined primarily by the secular interaction of Jupiter and Saturn. Therefore, it is interesting to consider how the location of this resonance would have shifted during the orbital migration of the giant planets. Classical secular perturbation theory for two planets (40) can be used to determine the past locations of this resonance provided one knows the initial (pre-migration) orbital eccentricities and semimajor axes of Jupiter and Saturn. Numerical simulations suggest that planetary orbital eccentricities do not vary significantly during migration (16), so we can reasonably adopt the present values of the planetary eccentricities. An estimate of 0.45 AU inward migration of Jupiter’s semimajor axis has been obtained from the orbital distribution of the Hilda asteroids (41). The initial (pre-migration) semimajor axis of Saturn is, however, not very well constrained. For plausible values in the range of $\pm$2 AU for the net migration of Saturn, we calculate that the sweeping of the $\nu_6$ could range from just a very small zone near its present location (at the inner edge of the main belt) to almost across the entire main belt. Thus the LHB impactors could, in principle, have originated from either a narrow region near the inner edge or indeed the entire asteroid belt. Because a strong radial compositional gradient exists in the main asteroid belt (42), it is possible that chemical analyses of LHB craters could identify more precisely within the main asteroid belt the provenance of the impactors, and thereby constrain the migration history of Saturn. This may also be possible using the MBA size distribution alone, if there were a radial variation of size distribution of the MBAs.

5. Dating Planetary Surfaces by the Cratering Record

There are several important implications for dating planetary surfaces on an absolute time scale using the cratering record. The LHB was a catastrophic event that occurred about 3.9 billion years ago. Therefore, it is not possible to use the cratering record to date surfaces older than 3.9 billion years, as that record has been obliterated by the cataclysmic event. The heavily cratered highlands of the Moon, Mars and Mercury are 3.9 billion years old. Surfaces that show the characteristic Population 1 size distribution but at a lower density than the heavily cratered highlands are probably closer to 3.8 billion years (the end of the LHB). Such surfaces include the
old plains on Mars and possibly the post-Caloris basin surface on Mercury (Fig. 1).

The younger post-LHB surfaces that show the Population 2 size distribution have been im-
pacted primarily by NEAs (Fig. 4). Therefore, the ages of these surfaces can be derived from the
crater production rate of NEAs. Because we do not know the comet crater production rate with
any certainty, the ages derived from the NEA impacts will be an upper limit.\textsuperscript{4} The exception is
Venus, because its crater population is unlikely to include comet impacts, as cometary impactors
are likely to have been almost entirely screened out by the thick Venusian atmosphere (43).

Finally, it is important to note that dating surfaces on outer planet satellites using the inner
planet cratering record is not valid. Attempts have been made to date outer planet surfaces on an
absolute time scale by assuming that the crater population found in the inner Solar System is the
same throughout the entire Solar System and has the same origin. But this assumption is false,
in light of our results. Indeed, additional evidence to support this conclusion is to be found in the
cratering record of the Jovian satellites: Callisto has a crater size distribution distinctly different
than both Population 1 and Population 2 craters on the inner planets (44). The differences cannot
be accounted for by differences in the impact velocities or target and projectile properties (21).
Whatever caused the Callisto crater record did not significantly affect the inner Solar System. The
inner planet impactors originated primarily from the asteroid belt, and these would likely be a
relatively minor component of the impactor populations for the outer planet satellites.

6. Conclusions

We have demonstrated that the cratering record of the inner planets and the Moon has been
caused by two distinct impactor populations that are distinguishable by their size distributions. We
have further shown that the old, Late Heavy Bombardment population has a size distribution very
similar to that of the current main belt asteroids, whereas the younger, post-LHB population is very
in similar size distribution to the present Near Earth Asteroids. Taken together with cosmochemical
evidence, these results imply that the origin of the LHB impactors was the dynamical ejection, over
a relatively short time interval \(\sim 100\) Myr, of asteroids from the main asteroid belt approximately
600 Myr after the formation of the inner planets; the craters on the inner planets since that time
are due to the NEAs whose different size distribution is owed to a combination of non-gravitational
effects (the Yarkovsky effect) and Jovian resonances that preferentially deliver smaller asteroids
(diameter \(\lesssim 20\) km) from the main asteroid belt to planet-crossing orbits. The results also imply
that the main asteroid has had its the presently-observed size distribution for at least 3.8 Gy.

The cataclysmic Late Heavy Bombardment event was unique in the history of the Solar System,
and quite plausibly the result of asteroids ejected from the main belt by the inward sweeping of

\textsuperscript{4}The cometary contribution is likely low, judging by the good match between the NEA and the Population 2
projectile size distributions.
gravitational resonances caused by the orbital migration of the giant planets. The timing of the LHB at about 3.9 Gy ago implies that Uranus and Neptune did not reach critical masses to gravitationally scatter and clear their associated planetesimal debris (and thereby trigger the giant planet migration process) until \( \sim 600 \) My after the formation of the other planets. This result resolves a long-standing question on the formation timescale of Uranus and Neptune.

Our results are of major import for planetary geology and the dating of solid planetary surfaces in the inner Solar System. The absolute ages of surfaces older than about 3.9 Gy cannot be determined from the cratering record because that record has been destroyed by the cataclysmic bombardment. Surfaces younger than about 3.8 billion years (the end of the LHB) can be dated by the NEA impact flux at the Moon, Mars, Venus and Mercury. We also note that the lunar and terrestrial planet cratering record provides a poor, and probably misleading, guide for the dating of solid body surfaces in the outer Solar System, such as the satellites of the giant planets.

We acknowledge research support from NASA, NAOJ and JSPS.

REFERENCES

20. Throughout this paper we display the crater and projectile size distributions using the “Relative” plot method (see NASA Technical Memorandum 79730, 1978) which was devised to better show the size distribution of craters and crater number densities for determining relative ages. The R plot provides a more sensitive and discriminating comparison tool than Cumulative Distribution plots which tend to smear out important details of the crater size distribution curves, and can lead to erroneous interpretations. For an R plot, the size distribution is normalized to a power law differential size distribution function, \(dN(D) \sim D^p dD\), where \(D\) is diameter and \(p = -3\), because most crater size distributions are observed to be within \(\pm 1\) of a \(p = -3\) power law distribution. The discretized equation for R is: \(R = \frac{D^3 N}{A(b_2 - b_1)}\), where \(D\) is the geometric mean diameter of the size bin (\(\sqrt{b_1 b_2}\)), \(N\) is the number of craters in the size bin, \(A\) is the area over which the counts were made, \(b_1\) and \(b_2\) are the lower and upper limit of the size bin, respectively. The size bins are usually defined in \(\sqrt{2}\) increments because there are many more small craters than large craters. In an R plot, log R is plotted on the y-axis and log(D) is plotted on the x-axis. A \(p = -3\) distribution plots as a horizontal straight line; a \(p = -2\) distribution slopes down to the left at an angle of 45°, and a \(p = -4\) distribution slopes down to the right at 45°. The vertical position of the line is a measure of crater density; the higher the vertical position, the higher the crater density.
23. This also demonstrates that the shape of the lunar highlands curve has not been affected by crater saturation because they show the same size distribution as the highlands but the crater densities are well below saturation density, confirming a theoretical result that a surface impacted by a population with the same size distribution as the one observed for the


Figure Captions

**Figure 1.** The crater size distributions on the Moon, Mars and Mercury, shown as R plots (20). The blue curves are for the Moon (highlands, post-mare, and Class I craters), red are for Mars (highlands, old plains and younger plains craters), and green are for Mercury (highlands and post-Caloris craters).

**Figure 2.** Size distributions of all Venus craters and, separately, multiple craters, compared to the Mars Northern Plains. The downturn in the Venus curves is due to atmospheric screening of projectiles.

**Figure 3.** These crater curves summarize the inner Solar System cratering record, with two distinctly different size distributions. The red curves are Population 1 craters that represent the period of Late Heavy Bombardment. The lower density blue curves (Population 2) represent the post-LHB era on the Moon, Mars and Venus.

**Figure 4.** The size distributions of the projectiles (derived from the crater size distributions) compared with those of the Main Belt Asteroids (MBAs) and Near Earth Asteroids (NEAs). The red circles are for the lunar highlands (Population 1), and the red squares are for the Martian young plains (Population 2). The other colors and point styles are for the asteroids derived by various authors. The MBA size distribution is virtually identical with Population 1 projectiles responsible for the LHB crater record. The NEA size distribution is the same as Population 2 projectiles responsible for the post-LHB crater record.
Figure 1

Figure 3
Strom, Malhotra, Ito, Yoshida & Kring, 2005
Figure 4

Strom, Malhotra, Ito, Yoshida & Kring, 2005