THE EVOLUTION OF SMALL BODY POPULATIONS: FROM PLANET MIGRATION TO THERMAL DRIFT FORCES

Community White Paper to the Planetary Science and Astrobiology Decadal Survey 2023-2032

Sanctioned by the NASA Small Bodies Assessment Group

July 15, 2020

Author:
William F. Bottke

Endorsers:

<table>
<thead>
<tr>
<th>Endorser</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qing-Zhu Yin</td>
<td>University of California at Davis, Department of Earth and Planetary Sciences</td>
</tr>
<tr>
<td>Kris Zacny</td>
<td>Honeybee Robotics</td>
</tr>
<tr>
<td>Shashwat Shukla</td>
<td>University of Twente, The Netherlands</td>
</tr>
<tr>
<td>Jonti Horner</td>
<td>University of Southern Queensland</td>
</tr>
<tr>
<td>Jennifer Scully</td>
<td>Jet Propulsion Laboratory, California Institute of Technology</td>
</tr>
<tr>
<td>Luke Dones</td>
<td>Southwest Research Institute</td>
</tr>
<tr>
<td>Björn J. R. Davidsson</td>
<td>Jet Propulsion Laboratory, California Institute of Technology</td>
</tr>
<tr>
<td>Ross A. Beyer</td>
<td>SETI Institute and NASA Ames Research Center</td>
</tr>
<tr>
<td>Name</td>
<td>Institution</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Stephen Mojzsis</td>
<td>University of Colorado at Boulder</td>
</tr>
<tr>
<td>Simone Marchi</td>
<td>SwRI</td>
</tr>
<tr>
<td>Vincent F Chevrier</td>
<td>University of Arkansas</td>
</tr>
<tr>
<td>Geol. Planet. Sci. Caltech</td>
<td></td>
</tr>
<tr>
<td>Devanshu Jha</td>
<td>MVJ College of Engineering, India</td>
</tr>
<tr>
<td>Amy J. V. Riches</td>
<td>UK</td>
</tr>
<tr>
<td>Renu Malhotra</td>
<td>Lunar and Planetary Laboratory, The University of Arizona</td>
</tr>
<tr>
<td>Oriel A Humes</td>
<td>Northern Arizona University</td>
</tr>
<tr>
<td>Carol A. Raymond</td>
<td>Jet Propulsion Laboratory, California Institute of Technology</td>
</tr>
<tr>
<td>Marina Gemma</td>
<td>American Museum of Natural History</td>
</tr>
<tr>
<td>Ronald A. Fevig</td>
<td>University of North Dakota</td>
</tr>
<tr>
<td>Joe Masiero</td>
<td>JPL</td>
</tr>
<tr>
<td>Jared Espley</td>
<td>GSFC</td>
</tr>
<tr>
<td>Christian Koeberl</td>
<td>University of Vienna</td>
</tr>
<tr>
<td>Maria Gritsevich</td>
<td>Finnish Geospatial Research Institute</td>
</tr>
</tbody>
</table>
EXECUTIVE SUMMARY

Over the last several decades, there has been a transformation in our understanding of early Solar System evolution. Historically, it was thought that the planets and small bodies formed near their current locations, but it has now become clear that the structure of the Solar System changed as the planets grew and migrated [e.g., 1,2]. Evidence for giant planet migration can be found in the orbits of the giant planets (and exoplanets) and the orbital structures of the Kuiper belt, scattered disk, irregular satellites, Trojans, and asteroid belt [1,2]. At least one period of migration was driven by interactions between the giant planets and a massive outer cometary disk [3]. As the giant planets moved to their current orbits, reservoirs of small bodies became destabilized, leading to enhanced bombardment rates on all Solar System worlds [4].

Additional evidence for migration comes from meteorite samples and spacecraft data, with new interpretations arguing that much of the asteroid belt is not indigenous to that region [5, 6,7]. Instead, the asteroid belt may have been filled with planetesimals and planetary fragments from the terrestrial planet region, giant planet zone, and primordial Kuiper belt as a byproduct of planet formation processes and planet migration [e.g., 8-13]. From a sample perspective, this implies the entire early Solar System may be accessible by visiting near-Earth asteroids, fragments of main belt bodies that have been driven into Earth’s orbital neighborhood.

The timing of giant planet migration, however, is unclear. Late migration models, in which a delay between planet formation and migration produced a discrete “late heavy bombardment” (LHB) episode ~500 Myr after the Earth-Moon system formed, now appear to violate key constraints [e.g., 14]. New work instead argues for early migration (< 100 Myr after CAIs), with powerful consequences for terrestrial planet formation, the origin of the Moon, the dynamical evolution of asteroids, and the bombardment history of all these worlds [e.g., 15, 16].

The post-migration history of the asteroid belt is dominated by the disruption of large asteroids, which produce clusters of rocky fragments, and Yarkovsky thermal forces, which help drive the bodies onto planet-crossing orbits [17]. Some breakups have even led to surges in the Earth’s impact rate, with the astrobiology implications virtually unexplored! [18,19].

In short, the story of our Solar System cannot be told without understanding planet migration, while the story of life on Earth may be surprisingly dependent on the history of the main belt. Small body missions are well positioned to glean insights into each idea, given their ability to visit worlds that are largely unchanged since the earliest times of Solar System history. In this white paper, we briefly discuss the advances made over the last decade, some questions driven by our progress, and how missions could potentially provide answers to key questions of the 2010 Decadal Survey (DS10):

**DS10 Questions: Building New Worlds:**

- What were the initial stages, conditions, and processes of solar system formation and the nature of the interstellar matter that was incorporated?
- How did the giant planets and their satellite systems accrete, and is there evidence that they migrated to new orbital positions?
- What governed the accretion, supply of water, chemistry, and internal differentiation of the inner planets and the evolution of their atmospheres, and what roles did bombardment by large projectiles play?

**Key bodies to investigate:**

Asteroids, comets, Trojans, irregular satellites, and Kuiper belt objects. Meteorite samples from all worlds. The bombarded surfaces of ancient rocky and icy worlds.
A BRIEF HISTORY OF PLANET MIGRATION

It has been argued that inner Solar System bodies, giant planet cores, and the primordial Kuiper belt formed by the accumulation of initially small objects. While much progress has been made with this model, simulating a planetary system that reproduces the one we live in is still difficult. Planetary orbits are often too excited, Mars analogs are too big, and asteroid belts are too patchy. This has led to proposed solutions that involve planetary migration. Here we briefly discuss some of the major ideas in the field, their implications, and how they might be tested.

Planet Migration in Grand Tack Model

The first opportunity for planet migration in our Solar System occurs when the protoplanetary gas nebula is still in existence, approximately ~3–10 Myr after the birth of the Sun; [20, 21]). As our giant planets grow, they begin to trade angular momentum within the gas disk. This can cause them to radially migrate, with some traveling far enough to be captured in orbital resonances with a neighboring body [1]. The consequence is often a compact system of giant planets in mutual mean motion resonances whose long-term stability is uncertain.

A model designed to address this phase of giant planet migration that also takes on the issue of Mars and asteroid belt formation is the “Grand Tack” [9]. It builds on the ideas of Hansen (2009) [23], who showed that all of the terrestrial planets could form with masses and orbits like those observed today if the Venus-Earth region between 0.7 and 1 au initially contained small bodies, while the regions where Mercury and Mars orbit today were empty. While Hansen’s paper was provocative, his initial conditions were ad hoc. Walsh et al. [9] noted that in hydrodynamic simulations that followed the orbital migration of Jupiter and Saturn in a gas disk, Jupiter could have migrated inward all the way to ~1.5 au from the Sun. At that point Jupiter and Saturn could have become trapped in their mutual 2:3 resonance and would have begun to move outward. This outward migration continued until the final removal of gas in the disk, which the model assumes happened when Jupiter reached a distance of ~5.5 AU.

**Figure 1:** A sketch of the layout and evolution of the Grand Tack scenario (Walsh et al. 2011). Jupiter and Saturn are in a gas disk with planetesimals interior and exterior. The inner disk of planetesimals is labeled ‘S-type’, a moniker to represent less-primitive, volatile-poor asteroids. The outer disk, labeled ‘C-type’, represents a population of more primitive asteroids. The inward gas-migration of Jupiter scatters S-type bodies, and truncates the inner planetesimal disk. Such a disk can form a low-mass Mars (Hansen 2009). When Jupiter and Saturn enter the 2:3 resonance, they begin to migrate outward. Jupiter’s outward migration scatters both S-type and C-type bodies into the asteroid belt, thereby producing the gradient of taxonomic types observed. This scenario leaves the asteroid belt in a highly excited state – more excited than is currently observed.

The Grand Tack model examined the gravitational scattering of small bodies in the inner and outer solar system by Jupiter and Saturn during their inward, then outward migration. It showed that bodies that were once denizens of the Saturn-Neptune zone, which would presumably have characteristics similar to carbonaceous chondrite-like asteroids (broadly C-types), could be captured into the main belt region. It also found that S-type asteroids, bodies with characteristics representative of planetesimals from the Mars zone, could also be captured in the main belt. Intriguingly, this model not only reproduces the mass and mixture of spectral types in the asteroid belt, but also truncates the planetesimal disk from which the terrestrial planets form, as in Hansen’s work. This allowed [9] to form a low-mass Mars. In this scenario,
Mars was one of many embryos forming near 1 AU at the time of Jupiter and Saturn’s migration, but was later scattered out to 1.5 AU by its siblings.

The idea that many C-type asteroids originated in the giant planet zone is provocative, but potential supporting evidence for this idea comes from the analysis of meteorites. Using isotopes of titanium, chromium, oxygen, molybdenum, and tungsten, it has been shown that all stony and iron meteorites fall into two fundamentally different groups: carbonaceous and non-carbonaceous bodies [5,6]. Here it is argued that carbonaceous planetesimals accreted in the outer Solar System while non-carbonaceous planetesimals accreted in the inner Solar System. The difference between the two groups implies little mixing of solids between the reservoirs after some early time. One way to accomplish this would be the growth of Jupiter, whose large mass would cut off communication between the inner/outer solar systems.

Additional evidence comes from NASA’s DAWN mission. Ceres, the largest C-type asteroid in the main belt, has ammoniated phyllosilicates on its surface [7]. Ammonia is found in the cold outer Solar System, and its presence suggests that it was incorporated into Ceres prior to Ceres’s transport into the main asteroid belt.

It should be noted that there are alternative scenarios that can also explain some of the above constraints. For example, C-type bodies from the outer solar system might migrate into the main belt via gas drag and interactions with the growing giant planets [e.g., 12]. In addition, several scenarios now exist that explain the relatively small size of the Mars and the inclusion of material form the terrestrial planet region into the main belt [e.g., 15]. Tests of these scenarios can be accomplished by new numerical modeling work, meteorite studies, new ground-based and space-based observations (e.g., LSST; PS4; NEOSM), and data returned by small body missions.

Finally, the introduction of carbonaceous bodies into the inner solar system would have produced numerous impacts on growing worlds [see 1,2,12]. Carbonaceous chondrites are organic- and volatile-rich compared to non-carbonaceous samples, so it is possible that the majority of these substances were delivered to Earth, Mars, and other worlds by this population.

A List of Some of the Major Questions Related to this Scenario

- Was the asteroid belt initially empty, or did planetesimal formation take place within the main belt region? If so, which asteroids are indigenous planetesimals?
- Did Jupiter and Saturn migrate across the primordial main asteroid belt? If not, can main belt planetesimals be sufficiently dynamically excited by a giant plane instability?
- Did most C-type asteroids form in the giant planet zone? If so, what was the implantation mechanism, and how did it affect the orbit and size distribution of these bodies?
- Were planetesimals from the terrestrial planet region (and the progenitors of Earth) implanted into the main belt? If so, which ones, and what are their orbit and size distributions?
- Was the early main belt massive, with most bodies lost when Jupiter/Saturn migrated across the main belt, or was it never larger than a few times its present-day population?
- How were implanted planetesimals from across the solar system affected by collisional evolution prior to the dissipation of the gas disk?
- Do asteroid surfaces record the earliest bombardment history from this epoch, or has subsequent collisional and thermal history erased it?
- How can we use asteroids and meteorites to constrain these early dynamical processes?

Planet Migration After a Giant Planet Instability

After the gas disk disperses, our solar system was in a very different configuration than the one seen today. Gas accretion left the giant planets on nearly circular coplanar orbits between 5 and 20 au, with most/all locked in mutual mean motion resonances with one another. Moreover,
the primordial Kuiper belt existed just beyond the original orbit of Neptune, with an estimated comet population of ~20 Earth masses that stretched beyond 20 au [1, 2].

This configuration, however, eventually went unstable. The most natural explanation for the dynamical structure of the present-day Kuiper Belt is that the outermost giant planets radially migrated through the primordial comet disk in a violent exchange of orbital energy and angular momentum [e.g., 3]. This led to a reorganization of the giant planets in what is now called a “giant planet instability” [1].

![Diagram showing stages of giant planet migration and dynamical instability](image)

**Fig. 1. Giant planet migration and dynamical instability.** The giant planets scatter each other, while Neptune migrates into a massive outer disk of cometsimals (i.e. primordial comets). This migration dynamically affects the terrestrial planets/asteroid belt (e.g., Nesvorný 2018).

Uranus and Neptune were scattered into the comet disk at this time, flinging its members throughout the Solar System as they migrated through it. It caused the primordial Kuiper belt to lose a factor of ~1000 in population, with roughly this number of Pluto-sized bodies -- and enumerable smaller bodies -- scattered into the giant planet zone and/or out into a scattered disk of comets associated with Neptune. The interaction between Neptune and Pluto-sized bodies made the migration grainy rather than smooth, and this in turn affected the capture of Kuiper belt objects into Neptune resonances [23].

The number of giant planets gravitationally interacting with one another at the time of the instability is unknown. Dynamical studies suggest systems with 5 or 6 giant planets (e.g., 3 or 4 Uranus/Neptune-size bodies) have greater success reproducing dynamical constraints across the solar system than those that start with 4 giant planets [e.g., 1]. In the most successful cases, a Neptune-sized body interacting with Jupiter causes it to migrate slightly inward via numerous tiny jumps [e.g., 2]. These events, as well as the interaction of a Neptune-sized body with the asteroid belt itself, may help describe the dynamical structure of the asteroid belt.

In addition, giant planet encounters with one another allowed a small fraction of ejected comets from the primordial disk to be captured in stable orbits across the Solar System via three body encounters. This mechanism most easily explains why primitive D- and P-type asteroids are found in the central and outer main belt, the Jupiter and Neptune Trojan populations, the irregular satellite populations, and so on [e.g., 1, 2, 9, 10].

Finally, there are indications that bodies larger than Earth yet smaller than Neptune were also active in the giant planet zone (and beyond). The most obvious evidence comes from the large obliquity of Uranus, which is often attributed to the impact of an Earth-mass body [23]. More controversially, perplexing orbital signatures among very distant Kuiper belt objects could point to the existence of an undiscovered sub-Neptune-sized body in the outer reaches of the Solar System [24]. While the existence of this “Planet X”-type body is hotly-debated, finding such a world would help to rewrite our understanding of the early solar system.

As the giant planets moved to their present-day orbits, the dynamical resonances associated
with them also moved to their current locations. This may have depleted and/or eliminated portions of the inner main belt and created the Hungaria asteroid population [25, 26].

The timing of the giant planet instability is controlled by the unknown size of the gap between Neptune’s initial orbit and the cometsesimal disk’s inner edge (e.g., [1]). The smaller the gap, the earlier the instability. If the gap is < 1.5 au, there is effectively no delay in the onset of Neptune’s migration, and it would take place shortly after dissipation of the solar nebula.

Originally, it was proposed that giant planet migration/instability occurred ~4 Ga and triggered the lunar Late Heavy Bombardment [e.g., 4]. New studies of the collisional evolution of the outer disk, however, contradict this idea; they indicate that the disk must have dispersed within the first 100 Myr of Solar System history [e.g., 14], too early to account for the youngest lunar basins. This new constraint implies that giant planet migration occurred simultaneously with the final stages of terrestrial planet formation. Early migration (and an early instability) also avoids a serious problem, namely the over-excitation of terrestrial planet orbits typically associated with late migration [e.g., 1, 15]. The issue of how terrestrial planet formation is affected by an early giant planet instability is a critical area for future work (e.g., 15).

It is also possible that planetesimals (or their fragments) from the terrestrial planet region were captured into the main belt region during this epoch as a byproduct of interactions with both protoplanets and resonances associated with migrating giant planets [8, 12].

All of this small body movement led to a heavy bombardment of all existing bodies [4]. The signatures of these events may still be found on ancient surfaces across the solar system (e.g., Mercury, the Moon, Mars, the icy satellites of the giant planets, Kuiper belt objects like Pluto). The interpretation of these worlds is complicated by the fact that some surfaces did not become fully stable until well after some phases of bombardment had begun [e.g., 16, 27, 28, 29].

The impact of P- and D-type material on the Earth by the giant planet instability are thought to be too limited to deliver more than 10% or so its volatiles, but these impacts may play a critical role in providing organics and noble gases to many worlds [e.g., 2, 30]. They could also be a source of water for the remaining terrestrial planets and possibly the Moon.

A List of Some of the Major Questions Related to this Scenario

- How are the signatures of early giant planet migration recorded in small body populations and larger worlds that span the Solar System?
- What was the collisional evolution of the asteroid and Kuiper belts during this epoch, and how can these scars be used to tell us the story of planet migration?
- How many giant planets and Earth-sized worlds existed prior to the giant planet instability?
- Can missions and/or meteorite analysis show that dormant comet-like asteroids were implanted in the main belt, Hildas, Trojans, irregular satellites by the giant planet instability?
- Were components of the primordial Kuiper belt left unchanged by giant planet migration (e.g., cold classical main belt containing primordial objects like Arrakoth)?
- Did the primordial Kuiper belt interact with a sub-Neptune-sized “Planet X”?
- Did the primordial asteroid belt interact with a “extra Neptune”?
- What is the role of hyper-volatiles in the Kuiper belt, and how can it be used to interpret the nature of active and dormant comet-like bodies (many of the latter which were captured)?

SMALL BODY EVOLUTION IN THE AFTERMATH OF PLANET MIGRATION

Small body populations continue to evolve after planet migration has concluded. The driver of this activity changes from giant planet migration to collisions and, in some cases, smaller-scale dynamical mechanisms such as migration of $D < 30$ km bodies produced by the Yarkovsky
effect, a thermal radiation force that causes asteroids to drift in semimajor axis via the absorption and reemission of sunlight [e.g., 31]. Even here, though, there are open questions.

In the immediate aftermath of planet migration, many small bodies find themselves residing in dynamically unstable regions. Examples are comets in the scattered disk, objects near but not in stable resonances in the Kuiper belt, and bodies within main belt resonances where the resonances were moved to those locations by giant planet migration [e.g., 26]. A period of transition then takes place as bodies move onto planet-crossing orbits, where they hit the planets, the Sun, or are thrown out of the solar system by encounters with Jupiter. The timescale of this transitional period will vary depending on the dynamical state of each small body reservoir.

Collisional evolution is also a steady mechanism that will create fragments before, during, and after planet migration [17]. The rate of collisional evolution in a population depends on several factors: population size, collision probabilities and impact velocities between the bodies, the time when collisions take place, and the possibility that external impactors could affect the population in question. Collisional evolution can be rapid when populations are massive and extremely slow when they are depleted. Accordingly, in the primordial comet disk, the initial excitation of the disk prior to and during the giant planet instability may have produced more collisional grinding than the quantity that occurred over the subsequent 4.5 Gyr. This would make the Kuiper belt size distribution something of a fossil from ancient times [e.g., 17].

In the main belt, the quantity of early collision evolution depends on the initial size if the primordial main belt, which is unknown. Numerical models suggest that most  D < 100 km bodies in the main belt are fragments produced by the disruption of  D > 100 km bodies, the most probable size of initial planetesimals [17]. If true, the main belt was deficient in sub-100 km bodies for a short or an extended time, depending on the nature of early Solar System evolution.

Key constraints that may relate to this issue are the ages of so-called asteroid families [32]. Families are produced by main belt collisions, with the fragments staying close enough to the impact site that the orbital concentration of bodies can be identified using clustering algorithms. As family members with diameter  D < 30 km obtain mobility from the Yarkovsky effect, they take on dynamical signatures that can be used to date the time of the break-up [31]. The age of many asteroid families has been identified using such methods, with ground truth provided by craters observed on asteroids within families (e.g., craters on Gaspra, residing in the Flora family, and returned samples from Itokawa, a possible Flora family member [e.g., 33]).

Curiously, few 3-4 Gyr old families can be found in the main belt, even though models predict they should be there [17]. Does this suggest ancient families never formed, or did they undergo sufficient collisional and dynamical dispersal over time to be effectively erased? The answer may tell us about the early history of the main belt.

Another key attribute of asteroid families is that family members (fragments) can migrate into resonances that drive them into the planet-crossing region. This implies that the breakup of a large asteroid at the right time can lead to surges in both near-Earth objects (NEOs) and impacts on the terrestrial planets [e.g., 19; see Zellner/Ghent white paper]. There are intriguing indications that the disruption of some asteroids has strongly influenced the history of astrobiology (e.g., the disruption of the L chondrite parent body 470 Myr ago corresponds with an upheaval in several astrobiological markers [18]). Accordingly, it is possible that the history of life can only be told by better understanding the evolution of families in the asteroid belt.

Families have also been identified in populations dominated by dormant comet-like objects, such as the Hildas and Trojans, but less so in the Kuiper belt [32]. Understanding the nature of such breakups may be needed to interpret the early history of the solar system and whether the inferred deficit of sub-km bodies in the Kuiper belt was driven by primordial processes or collisional evolution.
A List of Some of the Major Questions Related to this Phase of History

- When did the majority of collisional evolution take place for different small body reservoirs?
- Can large impact events in the asteroid belt be linked to meteoritical evidence?
- Where are the oldest families in the main belt and Kuiper belt?
- Which families can be linked to surges in the impact flux on Earth?
- How much influence have asteroid and comet impacts had in the history of astrobiology?

METHODS TO ANSWER QUESTION AND MEET GOALS

Many of the advances discussed here are the fruit of NASA’s Research and Analysis programs. We strongly advocate that support be continued for fundamental research ranging from laboratory studies to remote observations to theoretical/computational studies.

Small body missions can also help us answer major questions. For example, observations of Arrakoth, a primordial object in the cold classical Kuiper belt, have given us insights into planetesimal formation (shape/nature of Arrakoth), the sub-km population that comprises the Kuiper belt (craters on Arrakoth), and how that sub-km population evolved [34, 35].

Potential missions that could help us better understand the questions above include:

- **Sample return missions to large main belt asteroids.** What primordial signatures do they have, how have they been affected by 4.5 Gyr of collisional evolution, how have they thermally-evolved, and what can they tell us about collisional processes in general.
- **Sample return missions to strategic near-Earth asteroids.** We have argued that some NEAs may be fragments of objects from the terrestrial planet region, giant planet zone, and primordial Kuiper belt. Can the samples from these bodies tell us about the origin and evolution of their parent body and where they originally formed? Can family members tell us about the timing of specific events?
- **Visits to poorly-explored or unexplored populations (e.g., Hilda asteroids; Jupiter’s irregular satellites).** It has been suggested both populations originated in the primordial disk once located beyond Neptune. Can this be demonstrated by studying their bodies *in situ*?

REFERENCES