

**THE KUIPER BELT
AND THE PRIMORDIAL EVOLUTION
OF THE SOLAR SYSTEM**

A. MORBIDELLI

*Observatoire de la Côte d'Azur, Nice, France
morby@obs-nice.fr*

and

M.E. BROWN

California Institute of Technology, Pasadena, California

October 7, 2002

Invited review for *COMETS II*, M. Festou et al. eds., University Arizona Press, Tucson, AZ.

send correspondence to:

Alessandro Morbidelli
Observatoire de la Côte d'Azur
B.P. 4229, 06304 Nice Cedex 4
France
email: morby@obs-nice.fr
tel: 33-492003126
fax: 33-492003033

Abstract.

We discuss the structure of the Kuiper belt as it can be inferred from the first decade of observations. In particular, we focus on its most intriguing properties –like mass deficit, inclination distribution, the apparent existence of an outer edge and of a correlation between inclinations and colors– which clearly show that the belt has lost its pristine structure of a dynamically cold proto-planetary disk. Understanding how the Kuiper belt acquired its present structure would provide insight into the formation of the outer planetary system and on its early evolution. We critically review the scenarios that have been proposed so far on the primordial sculpting of the belt. None of them can explain in a unitary model all the observed properties; the real history of the Kuiper belt probably requires a combination of some of the proposed mechanisms.

1. Introduction

When Edgeworth and Kuiper first conjectured the existence of a belt of small bodies beyond Neptune –the presently called Kuiper belt– they certainly were imagining a disk of planetesimals preserving the pristine conditions of the proto-planetary disk. But, since the first discoveries of trans-Neptunian objects, astronomers have realized that it is not so: the disk has been altered by a number of processes which altered its original structure. Therefore, the Kuiper belt is a sort of “smoking gun”, which may provide us with a large number of clues to understand what happened in the outer solar system during the primordial ages. Potentially, the Kuiper belt might teach us more on the formation of the giant planets than the planets themselves. And, as in a domino game, a better knowledge of giant planets formation would inevitably boost our understanding of the subsequent formation of the terrestrial planets. Consequently, Kuiper belt research is now considered a top priority of modern planetary science.

A decade after the discovery of 1992 QB₁ (Jewitt and Luu, 1993), we now know 642 trans-Neptunian objects (semi-major axis $a > 30$ AU)¹. Of these, 307 have been observed during at least 2 oppositions, and 193 during at least three oppositions. Observations at 2 and 3 oppositions are necessary for the Minor Planet Center to compute the objects’ orbital elements with, respectively, moderate and good accuracy. Therefore, the trans-Neptunian population is gradually taking shape, and we can start to seriously examine the “smoking gun” and learn what it has to teach us. We should not forget, however, that our view of the trans-Neptunian population is still partial, and strongly biased by a number of factors, some of which are difficultly modeled.

A primary goal of this chapter is to present the orbital structure of the Kuiper belt as it stands from the current observations. We start in section 2 by presenting the various sub-classes that constitute the trans-Neptunian population. Then in section 3 we describe some striking properties of the population, such as its mass deficit, inclination excitation, radial extent and a puzzling correlation between orbital elements and physical properties. In section 4 we finally review the models that have been proposed so far on the primordial sculpting of the Kuiper belt. Some of these models date from the very beginning of Kuiper belt science, when only a handful of objects were known, and have been at least partially dismissed by the new data. Paradoxically, however, as the data increase in number and quality, it becomes increasingly difficult to explain all the properties of the Kuiper belt in the framework a unitary scenario. The conclusions are in Sect. 5.

¹ All numbers are updated as of June 15, 2002.

2. The Trans-Neptunian Populations

The trans-Neptunian population is traditionally subdivided in two sub-populations: the *scattered disk* and the *Kuiper belt*. The definition of these sub-populations is not unique, the Minor Planet center and various authors often using slightly different criteria. Here we propose and discuss a partition based on the dynamics of the objects and their relevance for the reconstruction of the primordial evolution of the outer Solar System.

In principle, one would like to call *Kuiper belt* the population of objects that, even if characterized by a chaotic dynamics, do not suffer close encounters with Neptune and thus do not undergo macroscopic migration in semi-major axis. Conversely, the bodies that are transported in semi-major axis by close and distant encounters with Neptune should constitute the *scattered disk*. The problem for the precise partition of the trans-Neptunian population into Kuiper belt and scattered disk is related to timescale. On which timescale should we see semi-major axis migration to classify an object in the scattered disk? The question is relevant, because bodies trapped in resonances can significantly change their perihelion distance, and pass from a scattering phase to a non-scattering phase and vice-versa, even several times over the age of the solar system.

For this reason, we prefer to link the definition of the scattered disk to its formation mechanism. We call *scattered disk* the region of the orbital space that can be visited by bodies that have encountered Neptune within a Hill's radius at least once during the age of the Solar System, assuming no substantial modification of the planetary orbits. We then call *Kuiper belt* the complement of the scattered disk in the $a > 30$ AU region.

To categorize the observed trans-Neptunian bodies into scattered disk and Kuiper belt, we refer to previous works on the dynamics of trans-Neptunian bodies in the framework of the current architecture of the planetary system. For the $a < 50$ AU region, we use the results by Duncan et al. (1995) and Kuchner et al. (2002), who numerically mapped the regions of the (a, e, i) space with $32 < a < 50$ AU that can lead to a Neptune encountering orbit within 4 Gy. Because dynamics are reversible, these are also the regions that can be visited by a body after having encountered the planet. Therefore, according to our definition, they constitute the scattered disk. For the $a > 50$ AU region, we use the results by Levison and Duncan (1997) and Duncan and Levison (1997), who followed for another 4 Gy time-span the evolutions of the particles that encountered Neptune in Duncan et al. (1995). Despite the fact that the initial conditions did not cover all possible configurations, we can reasonably assume that these integrations cumulatively show the regions of the orbital space that can be possibly visited by bodies transported to $a > 50$ AU by Neptune encounters. Again, according to our definition, these regions constitute the scattered disk.

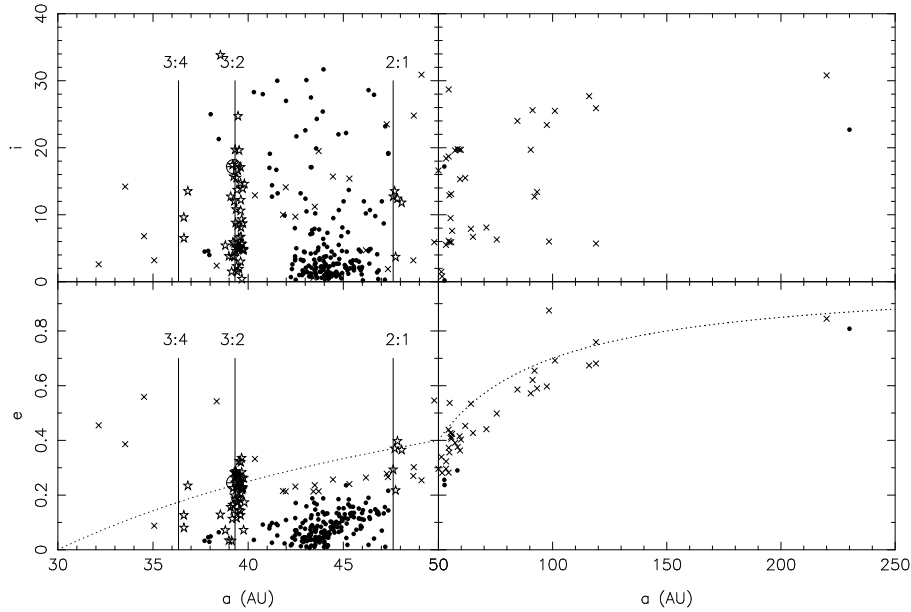


Fig. 1. The orbital distribution of multi-opposition trans-Neptunian bodies, as of June 15, 2002. Scattered disk bodies are represented as a cross, classical Kuiper belt bodies as dots and resonant bodies as stars. We qualify that, in absence of long term numerical integrations of the evolution of all the objects, a few bodies could have been miss-classified. Thus, the figure should be considered as an indicative representation of the various subgroups that compose the trans-Neptunian population. The dotted curve denotes $q = 30$ AU. The vertical solid lines mark the locations of the 3:4, 2:3 and 1:2 mean motion resonances with Neptune. The orbit of Pluto is represented by a crossed circle.

In Fig. 1 we show the (a, e, i) distribution of the trans-Neptunian bodies which have been observed during at least two oppositions. The bodies that belong to the scattered disk according to our criterion are represented as crosses, while Kuiper belt bodies are represented by dots and stars (whose difference will be explained below).

We believe that our definition of scattered disk and Kuiper belt is meaningful for what concerns the major goal of Kuiper belt science, that is to reconstruct the primordial evolution of the outer solar system. In fact, all bodies in the solar system must have been formed on orbits with very small eccentricities and inclinations, typical of an accretion disk. The current orbits of scattered disk bodies might have been reached starting from quasi-circular orbits in Neptune's zone by pure dynamical evolution, in the framework of the current architecture of the solar system. Therefore, they do not provide us any relevant clue to uncover the primordial architecture, although their relative orbital distribution might tell us something interesting

(Emel'yanenko, 2002). The opposite is true for the orbits of the Kuiper belt objects with non-negligible eccentricity and/or inclination. Their existence reveals that some excitation mechanism, which is no longer at work, occurred in the past (see section 4).

In this respect, particularly important is the existence of Kuiper belt bodies with $a > 50$ AU, on highly eccentric orbits (4 objects in Fig. 1). Among them 2000CR₁₀₅ ($a = 230$ AU, perihelion distance $q = 44.17$ AU and inclination $i = 22.7^\circ$) is a challenge by itself for what concerns the explanation of its origin. We like to call them *extended scattered disk* objects for two reasons: (i) they do not belong to the scattered disk according to our definition but are very close to its boundary and (ii) a body of ~ 300 km like 2000CR₁₀₅ presumably formed much closer to the Sun, where the accretion timescale was sufficiently short (Stern, 1996), implying that it has been subsequently transported in semi-major axis until its current location was reached. This hypothesis suggests that in the past the true scattered disk extended well beyond its present boundary in perihelion distance. Given that the observational biases become rapidly more severe with increasing perihelion distance and semi-major axis, the currently known extended scattered disk objects may be like the tip of an iceberg, e.g. the emerging representatives of a conspicuous population, possibly outnumbering the scattered disk population (Gladman et al., 2002).

In addition to the extended scattered disk, we distinguish two other sub-populations of the Kuiper belt. We call *resonant population* the Kuiper belt bodies that are located in some major mean motion resonance with Neptune (essentially the 3:4, 2:3 and 1:2 resonances; star symbols in Fig 1). It is well known that mean motion resonances offer a protection mechanism against close encounters with the resonant planet (see for instance chapter 9 of Morbidelli, 2002). For this reason, the resonant population –which, being part of the Kuiper belt, by definition must not encounter Neptune within the age of the solar system– can have perihelion distances much smaller than the other Kuiper belt objects, and even Neptune-crossing orbits ($q < 30$ AU) as in the case of Pluto. The bodies in the 2:3 resonance are often called *Plutinos*, for the analogy of their orbit with that of Pluto. We call *classical belt* the collection of Kuiper belt objects with $a < 50$ AU that are not in any notable resonant configuration. Because they are not protected from close encounters with Neptune by any resonance, the stability criterion confines them to the region with small to moderate eccentricity, typically on orbits with $q > 35$ AU. The adjective “classical” is justified because, among all sub-populations, this is the one whose orbital properties are the most similar to those expected for the Kuiper belt prior to the first discoveries. We remark however that the classical population is not that “classical”. Although moderate, the eccentricities are larger than those that should characterize a proto-planetary disk. Moreover, several bodies have very large inclinations

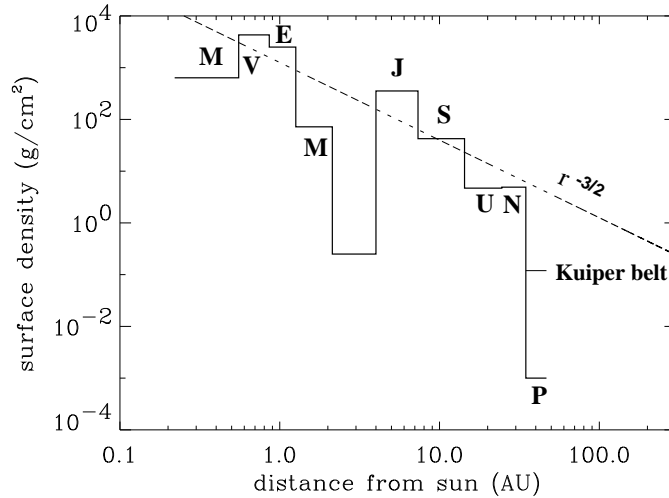


Fig. 2. The mass distribution of the solar nebula inferred from the masses of the planets augmented by the mass needed to bring the observed material to solar composition (data from Lewis 1995). The surface density in the Kuiper belt has been computed assuming a current mass of $\sim 0.1 M_{\oplus}$ (Jewitt et al. 1996, Chiang and Brown 1999, Trujillo et al. 2001, Gladman et al. 2001) in the 42–48 AU annulus, and scaling the result by a factor 70 in order to account for the inferred primordial local ratio between volatiles and solids. The estimate of the total mass in the Kuiper belt overwhelms that of Pluto, but still does not bring the mass to the extrapolation of the $\sim r^{-3/2}$ line.

(see section 3.2). Finally, the total mass is only a small fraction of the expected pristine mass in that region (section 3.1). All these elements indicate that also the classical belt has been labored by some primordial excitation and depletion mechanism(s).

3. The structure of the Kuiper belt

3.1. THE MISSING MASS OF THE KUIPER BELT

The original argument followed by Kuiper (1951) to conjecture the existence of a band of small planetesimals beyond Neptune was related to the mass distribution in the outer solar system. The minimum mass solar nebula inferred from the total planetary mass (plus lost volatiles) smoothly declines from the orbit of Jupiter until the orbit of Neptune (see Fig. 2); why should it abruptly drop beyond the last planet? However, while Kuiper’s conjecture on the existence of a trans-Neptunian belt is correct, the total mass in the 30–50 AU range inferred from observations is two orders of magnitude smaller than the one he expected.

Kuiper's argument is not the only indication that the mass of the primordial Kuiper belt had to be significantly larger. Further evidence for a mass deficit was uncovered by Stern (1995) who found that the objects currently in the Kuiper belt were incapable of having formed in the present population: collisions are sufficiently infrequent that 100 km objects cannot be built by pairwise accretion of the current population over the age of the solar system. Moreover, owing to the large eccentricities and inclinations of Kuiper belt objects –and consequently to their high encounter velocities– collisions that do occur tend to be erosive rather than accretional, making bodies smaller rather than larger. Stern suggested that the resolution of this dilemma is that the primordial Kuiper belt was both more massive and dynamically colder, so that more collisions occurred, and they were gentler and therefore generally accretional.

Following this idea, detailed modeling of accretion in a massive primordial Kuiper belt was performed by Stern (1996), Stern & Colwell (1997a,b) and Kenyon and Luu (1998, 1999a, 1999b). While each model includes different aspects of the relevant physics of accretion, fragmentation, and velocity evolution, the basic results are in approximate agreement. First, with $\sim 10 M_{\oplus}$ (Earth masses) or more of solid material in an annulus from about 35 to 50 AU on very low eccentricity orbits ($e \leq 0.001$), all models naturally produce of order a few objects the size of Pluto and approximately the right number of ~ 100 km objects. The models suggest that the majority of mass in the disk was in bodies approximately 10 km and smaller. The accretion stopped when the formation of Neptune or other dynamical phenomena (see section 4) began to induce eccentricities and inclinations in the population high enough to move the collisional evolution from the accretional to the erosive regime (Stern 1996).

While the general formation picture of an initial massive Kuiper belt appears secure, a fundamental question remains to be addressed: how did the initial mass disappear? Collisions can grind bodies down to dust particles, which are subsequently transported away from the belt by radiation pressure and/or Poynting Robertson drag, causing a net mass loss. The major works on the collisional erosion of a massive primordial belt have been done by Stern and Colwell (1997b) and Davis and Farinella (1997, 1998), achieving similar conclusions (see Farinella et al., 2000 for a review). As long as the planetesimal disk was characterized by small eccentricities and inclinations, the collisional activity could only moderately reduce the mass of the belt. However, when the eccentricities and inclinations became comparable to those currently observed, bodies smaller than 50–100 km in diameters could be effectively destroyed. The total amount of mass loss depends on the primordial size distribution. To reduce the total mass from $30 M_{\oplus}$ to a fraction of an Earth mass, the primordial size distribution had to be steep enough that essentially all of the mass was carried by these small bodies,

while the number of bodies larger than ~ 100 km had to be basically equal to the present number. Though this outcome seems consistent with the suggestions of the accretional models, there is circumstantial (but nonetheless compelling) evidence suggesting the primordial existence of a much larger number of large bodies (Stern 1991). The creation of Pluto-Charon likely required the impact of two approximately like-sized bodies which would be the two largest currently known bodies in the Kuiper belt. The probability that the two largest bodies in the belt would collide and create Pluto-Charon is vanishingly small, arguing that many such-sized bodies must have been present and subsequently vanished. Similarly, the existence of Triton and the large obliquity of Neptune are best explained by the one-time existence of many large bodies being scattered through the Neptune system. The elimination of these large bodies could not be due to the collisional activity, but requires a dynamical explanation.

Understanding the ultimate fate of the 99% of the initial Kuiper belt mass that is no longer in the Kuiper belt is the first step in reconstructing the history of the outer solar system.

3.2. THE EXCITATION OF THE KUIPER BELT

An important clue to the history of the early outer solar system is the dynamical excitation of the Kuiper belt. While eccentricities and inclinations of resonant and scattered objects are expected to have been affected by interactions with Neptune, those of the classical objects should have suffered no such excitation. Nonetheless, the confirmed classical belt objects have an inclination range up to at least 32 degrees and an eccentricity range up to 0.2, significantly higher than expected from a primordial disk, even accounting for mutual gravitational stirring.

The observed distributions of eccentricities and inclinations in the Kuiper belt are highly biased. High eccentricity objects have closer approaches to the sun and thus become brighter and more easily detected. High inclination objects spend little time at low latitudes² at which most surveys take place, while low inclination objects spend zero time at the high latitudes where some searches have occurred.

Determination of the eccentricity distribution of the Kuiper belt requires disentanglement of eccentricity and semi-major axis, which is only possible for objects with well determined orbits, for which a large enough well-characterized sample is not yet available. Determination of the inclination distribution, however, is much simpler because the inclination of an object is well determined even after a small number of observations, and the lat-

² Latitude and inclination are defined with respect to the invariable plane, which is a better representation for the plane of the Kuiper belt than is the ecliptic (Brown and Trujillo 2003).

itude of discovery of each object is a known quantity. Using these facts, Brown (2001) developed general methods for de-biasing object discoveries to discern the underlying inclination distribution. The simplest method removes the latitude-of-discovery biases by considering only objects discovered within one degree of the invariable plane equator and weights each object by $\sin(i)$, where i is the inclination of each object, to account for the proportional fraction of time that objects of different inclination spend at the equator (strictly, one should use only objects found precisely at the equator; expanding to one degree around the equator greatly increases the sample size while biasing the sample slightly against objects with inclinations between 0 and 1 degree). An important decision to be made in constructing this inclination distribution is the choice of which objects to include in the sample. One option is to use only confirmed classical objects, by which we mean those that have been observed at least 2 oppositions and for which the orbit is reasonably assured of fitting the definition of the classical Kuiper belt, as defined above. The possibility exists that these objects are biased in some way against unusual objects which escape recovery at a second opposition because of unexpected orbits, but we expect that this bias is likely to be in the direction of under-reporting high inclination objects. On the other hand, past experience has shown that if we use all confirmed and unconfirmed classical bodies, we pollute the sample with misclassified resonant and scattered objects, which generally have higher inclinations and therefore artificially inflate the inclination distribution of the classical belt. We therefore chose to use only confirmed classical belt bodies with the caveat that some high inclination objects might be missing. Figure 3 shows the inclination distribution of the classical Kuiper belt derived from this method. This method has the advantage that it is simple and model independent, but the disadvantage that it makes no use of the information contained in high latitude surveys where most of the high inclination objects are discovered. For example, the highest inclination classical belt body found within 1 degree of the equator has an inclination of 10.1 degrees, while an object with an inclination of 31.9 degrees has been found at a latitude of 11.2 degrees. The two high inclination points in Fig. 3 attempt to partially correct this deficiency by using discoveries of objects between 3 and 6 degrees latitude to define the high inclination end of the inclination distribution. The latter is then matched to the low inclination end of the distribution in the range $6 < i < 10$ degrees.

Brown (2001) developed a more general method to use all objects simultaneously by comparing inclinations of all objects to those found from Monte Carlo observations of simple model inclination distributions at the latitudes of discovery. The simplest reasonable model distribution has a form where $f(i)di$, the number of objects between inclinations i and $i + di$, is proportional to $\sin i \exp(-i^2/2\sigma^2)di$ where σ is a measure of the excitation of

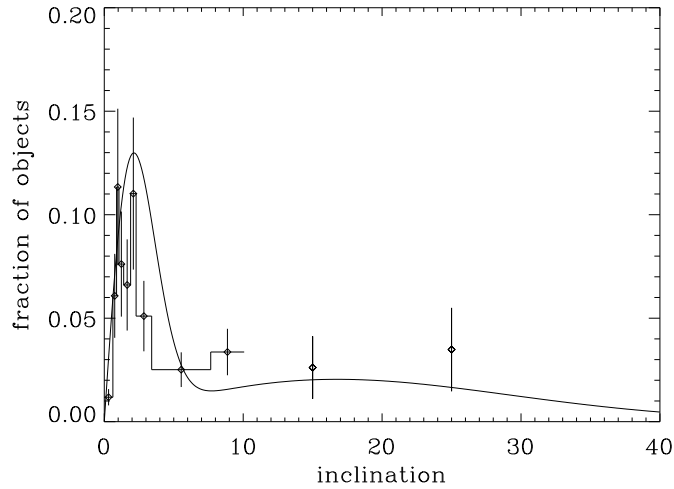


Fig. 3. The inclination distribution of the classical Kuiper belt. The points with error bars show the model-independent estimate constructed from a limited subset of confirmed classical belt bodies, while the smooth line shows the best fit two population model.

the population. The resonant and scattered objects are both well fit by this functional form with $\sigma = 10 \pm 2$ and 20 ± 4 degrees, respectively. The best single Gaussian fit for the confirmed classical belt objects can be ruled out at a high level of confidence; the observed inclination distribution of the classical Kuiper belt is more complex than can be described by the simplest model. Guided by Figure 3, we make the assumption that the inclination distribution between about 0 and 3 degrees appears adequately described by a single Gaussian times sine inclination, and search for a functional form to describe the higher inclination objects. The next simplest functional form is one with a second Gaussian added to the distribution: $f(i)di = \sin(i)[a_1 \exp(-i^2/2\sigma_1^2) + a_2 \exp(-i^2/2\sigma_2^2)]di$. The best fit to the two Gaussian model, found by modeling the latitudes and inclinations of all confirmed classical belt objects, has parameters $a_1 = 96.4$, $a_2 = 3.6$, $\sigma_1 = 1.8$ and $\sigma_2 = 12$ and is shown in Figure 3. For this model $\sim 60\%$ of the objects reside in the high inclination population.

A clear feature of this modeled distribution is the presence of distinct high and low inclination populations. While Brown (2001) concluded that not enough data existed at the time to determine if the two populations were truly distinct or if the model fit forced an artificial appearance of two populations, the larger amount of data now available, and shown in the model-independent analysis of Figure 3, confirms that the distinction between the

populations is real. The sharp drop around 4 degrees is independent of any model, while the extended distribution to 30 degrees is demanded by the presence of objects with these inclinations.

3.3. PHYSICAL EVIDENCE FOR TWO POPULATIONS IN THE CLASSICAL BELT

The existence of two distinct classical Kuiper belt populations, which we will call the hot ($i > 4$) and cold ($i < 4$) classical populations, could be caused in one of two general manners. Either a subset of an initially dynamically cold population was excited, leading to the creation of the hot classical population, or the populations are truly distinct and formed at separate times or in separate places. One manner in which we can attempt to determine which of these scenarios is more likely is to examine the physical properties of the two classical populations. If the objects in the hot and cold populations are physically different it is less likely that they were initially part of the same population.

The first suggestion of a physical difference between the hot and the cold classical objects came from Levison and Stern (2001) who noted that the intrinsically brightest classical belt objects (those with lowest absolute magnitudes) are preferentially found with high inclination. A potentially serious observational bias exists in this analysis, however. Many of the brightest objects are found in shallow wide-field surveys in which most observations necessarily occur at high latitude, where no low inclination objects can be found, while the fainter objects tend to be found in deeper surveys at lower latitude. Levison and Stern were aware of this caveat, but they concluded that it was not important, from the data available at the time. With increased number of data, however, it is now evident that the absolute magnitude and the latitude-of-discovery are indeed correlated. One simple method of roughly correcting this bias is again to use only objects found at latitudes less than one degree (Fig. 4). For this sample a Kolmogorov-Smirnov (K-S) test shows no statistical evidence that the brightest and faintest objects are drawn from different distributions. More relevant to our examination of the hot and cold classical populations, a K-S test also shows that there is no statistical evidence that the hot and cold classical objects are drawn from different absolute magnitude distributions. In summary, using the most recent dataset and accounting for the observational biases, no solid statistical evidence is found that the hot and cold classical populations differ in their size distributions. The possibility remains that the brightest objects reside exclusively in the hot classical Kuiper belt, as still appears to be the case from a cursory examination of the known objects, but from the currently available data we cannot make a statistically supportable claim of this effect. Wide field surveys currently being undertaken should soon find the largest objects in both populations and conclusively determine if the largest bodies

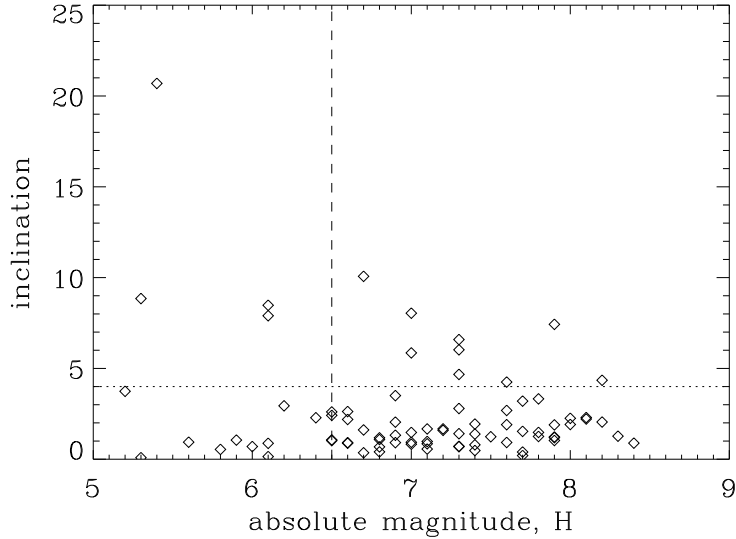


Fig. 4. Absolute magnitude versus inclination in a sample of the classical Kuiper belt corrected for observational biases. No statistical difference is seen in the inclination distribution of large and small objects, nor do any differences appear in the magnitude distribution of the hot and cold classical objects.

are indeed absent in the cold classical belt.

The second possible physical difference between hot and cold classical Kuiper belt objects is their colors, which relates in an unknown manner to surface composition. Several possible correlations between orbital parameters and color were suggested by Tegler and Romanishin (2000) and further investigated by Doressoundiram et al. (2001). The issue was clarified by Trujillo and Brown (2002) who quantitatively showed that for the classical belt, inclination, and no other independent orbital parameter, is correlated with color. In essence, the low inclination classical objects tend to be redder than higher inclination objects. Hainut (2002) has compiled a list of all published Kuiper belt colors which more than doubles the sample of Trujillo and Brown. A plot of color versus inclination for the classical belt objects in this expanded sample (Fig. 5) confirms the correlation between color and inclination. This expanded sample also conclusively demonstrates that no other dynamical correlations occur.

More interestingly, we see that the colors naturally divide into distinct low inclination and high inclination populations at precisely the location of the divide between the hot and cold classical objects. These populations differ at a 99.9% confidence level. Interestingly, the cold classical population

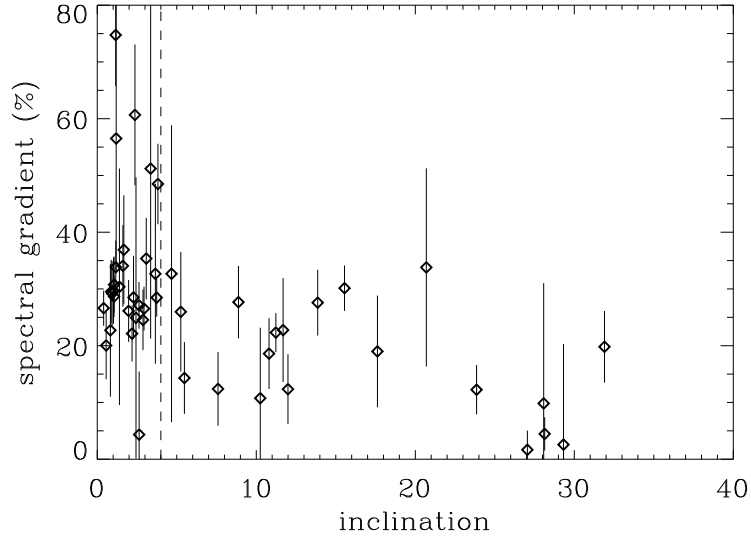


Fig. 5. Color gradient versus inclination in the classical Kuiper belt. Color gradient is the slope of the spectrum, in % per 100nm, with 0% being neutral and large numbers being red. The hot and cold classical objects have significantly different distributions of color.

also differs in color from the Plutinos and the scattered objects at the 99.8% and 99.9% confidence level, respectively, while the hot classical population appears identical in color to these other populations. The possibility remains, however, that the colors of the objects, rather than being markers of different populations, are actually *caused* by the different inclinations. Stern (2002), for example, has suggested that the higher average impact velocities of the high inclination objects will cause large scale resurfacing by fresh water ice which could be blue to neutral in color. If this hypothesis were correct, however, we would also expect to see correlations between colors and semi-major axis or eccentricity, which also determine impact velocities. These correlations do not exist. We would also expect to see correlations between color and inclination within the hot and cold populations. Again, these correlations do not exist. Finally, we would expect to see correlations between color and inclination or semi-major axis or eccentricity for all populations, not just the classical belt objects. Once again, no such correlations exist. While collisional resurfacing of bodies may indeed affect colors, there is clearly no causal relationship between average impact velocity and color. In summary, the significant color distinction between the hot and cold classical objects implies that these two populations are compositionally distinct

(at least in surface composition) in addition to being dynamically distinct. A confirmation of the surface composition differences between the hot and cold populations could be made with infrared reflectance spectroscopy, but to date no spectrum of a cold classical Kuiper belt object has been published.

3.4. THE RADIAL EXTENT OF THE KUIPER BELT

Another important property of interest for understanding the primordial evolution of the Kuiper belt is its radial extent. While initial expectations were that the mass of the Kuiper belt should smoothly decrease with heliocentric distance –or perhaps even increase in number density by a factor of ~ 100 back to the level of the extrapolation of the minimum mass solar nebula beyond the region of Neptune’s influence– the lack of detection of objects beyond about 50 AU soon began to suggest a drop off in number density (Dones 1997, Jewitt et al. 1998, Chiang and Brown 1999, Trujillo et al. 2001, Allen et al. 2001), though it was often argued that this lack of detections was the consequence of a simple observational bias caused by the extreme faintness of objects at greater distances from the sun (Gladman et al. 1998).

Determination of the reality of a density drop beyond 50 AU was hampered by the small numbers of objects and thus weak statistics in individual surveys. Trujillo and Brown (2001) developed a method to use all detected objects to estimate a radial distribution of the Kuiper belt. The method relies on the fact that the heliocentric distance (*not* semi-major axis) of objects, like the inclination, is well determined in a small number of observations and that within ~ 100 AU surveys have no biases against discovering distant objects other than the intrinsic radial distribution and the easily quantifiable brightness decrease with distance. Thus, at a particular distance, a magnitude m_0 will correspond to a particular object size s , but, assuming a power-law differential size distribution, each detection of an object of size s can be converted to an equivalent number n of objects of size s_0 by $n = (s/s_0)^{q-1}$ where q is the differential power-law size index. Thus the observed radial distribution of objects with magnitude m_0 , $O(r, m_0)dr$ can be converted to the true radial distribution of objects of size s_0 by

$$R(r, s_0)dr = O(r, m_0)dr \left[\frac{r(r-1)10^{(m-24.55)/5}}{15.60s_0} \right]^{q-1},$$

where albedos of 4% are assumed, but only enter as a scaling factor. Measured values of q for the Kuiper belt have ranged from 3.5 to 4.8 (see Trujillo and Brown 2001 for a review). We will assume the steepest currently proposed value of $q = 4.45$ (Gladman et al. 2001), which puts the strongest constraints on the existence of distant objects.

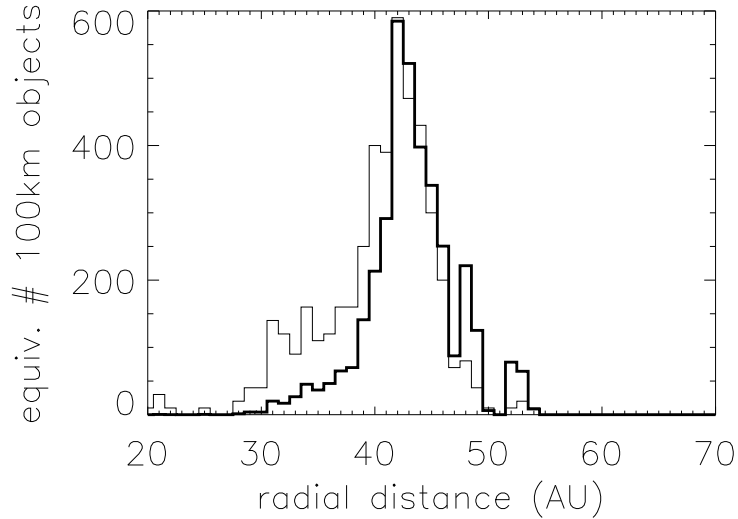


Fig. 6. The radial distribution of the Kuiper belt. The light line shows the observed number of trans-Neptunian objects per AU interval (times ten), while the thick bold line shows the true radial distribution inferred from this observed distribution taking into account biases due to brightness, distance, and size of the object. All discovered trans-Neptunian objects are considered in this analysis, regardless of their dynamical class.

Figure 6 shows the total equivalent number of 100 km objects as a function of distance implied by the detection of the known trans-Neptunian objects. One small improvement has been made to the Trujillo and Brown (2001) method. The power law size distribution is only assumed to be valid from 50 to 1000 km in diameter, corresponding to an expected break in the power law at some small diameter (Kenyon and Luu, 1999a) and a maximum object's size. The effect of this change is to only use objects between magnitudes 22.7 and 24.8, which makes the analysis only valid from 30 (where a 50 km object would be magnitude 24.8) to 80 AU (where a 1000 km object would be magnitude 22.7). Changes in the maximum object size assumed, s_{\max} are equivalent to changing the outer limit of the validity of the analysis by $80\text{AU}(s_{\max}/1000\text{km})^{1/2}$. Alternatively, one could further restrict the magnitude limits considered to limit the maximum size while maintaining validity to a particular distance. Different choices of minimum and maximum diameters have little effect on the final result unless extreme values for the maximum are chosen.

The analysis clearly shows that the known Kuiper belt is a localized increase in number density. Several implicit assumptions go into the above

method, but only extreme changes in these assumptions substantially change the results. For example, a change in the object size distribution beyond 43 AU could mimic a drop in object number density, but only if, by 50 AU, the distribution is so extreme that most of the mass is either in a few (undiscovered) large objects or a large number of (too faint) small objects. A physical reason for such a change is not apparent, particularly when one of the largest known Kuiper belt objects, 2002 AW₁₉₇, has a semi-major axis of 47.5 and a size of almost 900 km (Margot et al. 2002). Likewise, a lowering of albedo beyond 50 AU could make the appearance of a drop in number density, but, again, such a lowering is not physically motivated. A change in the inclination distribution beyond 50 AU could have the effect of hiding objects if they are concentrated in low inclination orbits close to the invariable plane, but repeating the analysis considering only objects found within 1 degree of the invariable plane still shows the sharp drop. While changing of these assumptions could indeed invalidate the analysis method above, the much simpler conclusion is that the number density of the Kuiper belt peaks strongly at 42 AU and quickly drops off beyond.

While the Trujillo and Brown (2001) method is good at giving an indication of the radial structure of the Kuiper belt where objects have been found, it is less good at determining upper limits to the detection of objects where none have been found. A simple extension, however, allows us to easily test hypothetical radial distributions against the known observations by looking at observed radial distributions of all objects found at a particular magnitude m_0 independent of any knowledge of how these objects were found. Assume a true radial distribution of objects $R(r)dr$ and again assume the above power law differential size distribution and maximum size. For magnitudes between m and $m + dm$, we can construct the expected observed radial distribution of all objects found at that magnitude, $o(r, m)drdm$ by

$$o(r, m)drdm = R(r)dr \left[\frac{r(r-1)10^{(m-24.55)/5} - q + 1}{15.6s_0} \right] dm,$$

where r ranges from that where the object of brightness m has a size of 50 km to that where the object of brightness m has a size of s_{\max} . The overall expected observed radial distribution is then simply the sum of $o(r, m)$ over the values of m corresponding to all detected objects. We can then apply a K-S test to determine the probability that the observed radial distribution could have come from the modeled radial distribution. We first apply this test to determine the magnitude of the drop off beyond 42 AU. Standard assumption about the initial solar nebula suggest a surface density drop off of $r^{-3/2}$. Figure 7 shows the observed radial distribution of objects compared to that expected if the surface density of objects dropped off as $r^{-3/2}$ beyond 42 AU. This distribution can be ruled at at the many sigma level. Assuming

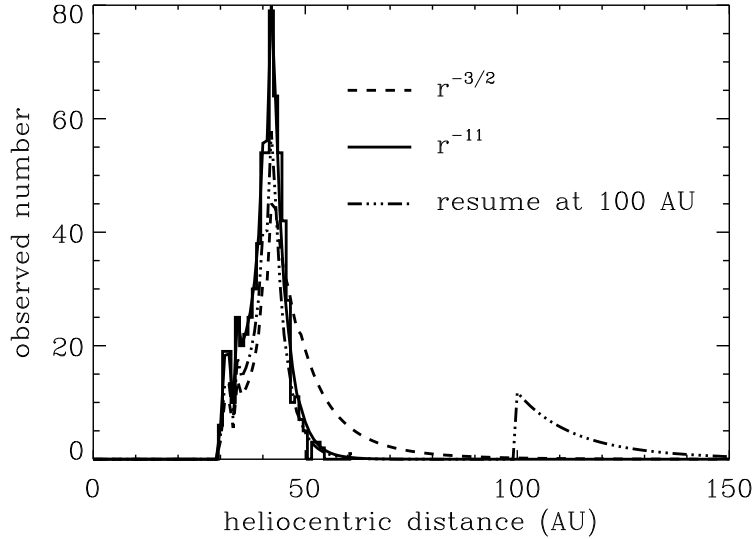


Fig. 7. The observed radial distribution of Kuiper belt objects (solid histogram) compared to observed radial distributions expected for models where the surface density of Kuiper belt objects decreases by $r^{-3/2}$ beyond 42 AU (dashed curve), where the surface density decreases by r^{-11} beyond 42 AU (solid curve), and where the surface density at 100 AU increases by a factor of 100 to the value expected from an extrapolation of the minimum mass solar nebula (dashed-dotted curve).

that the surface density drops as some power law, we model a range of different distribution $r^{-\alpha}$ and find a best fit of $\alpha = -11 \pm 4$ where the error bars are 3σ . This radial decay function should hold up to ~ 60 AU, beyond which we expect to encounter an almost flat distribution due to the scattered disk objects.

It has been conjectured that beyond some range of Neptune's influence the number density of Kuiper belt objects could increase back up to the level expected for the minimum mass solar nebula (see sect. 3.1). We therefore model a case where the Kuiper belt from 42 to 60 AU falls off as r^{-11} but beyond that the belt reappears at a certain distance δ with a number density found by extrapolating the $r^{-3/2}$ power law from the peak density at 42 AU and multiplying by 100 to compensate for the mass depletion of the classical belt (Fig. 7). Such a model of the radial distribution of the Kuiper belt can be ruled out at the 3σ level for all δ less than 115 AU (around this distance biases due to the slow motions of these objects also become important, so few conclusion can be drawn from the current data about objects beyond this distance). If the model is slightly modified to make the maximum object mass proportional to the surface density at a particular radius, a 100 times

resumption of the Kuiper belt can be ruled out inside 94 AU. Similar models can be made where a gap in the Kuiper belt exists at the presently observed location but the belt resumes at some distance with no extra enhancement in number density. These models can be ruled out inside 60 AU at a 99% confidence level.

While all of these results are necessarily assumption dependent, several straightforward interpretations are apparent. First, the number density of Kuiper belt objects drops sharply from its peak at around 42 AU. Second, a distant Kuiper belt with a mass approaching that of the minimum mass solar nebula is ruled out inside at least ~ 100 AU. And finally, a resumption of the Kuiper belt at a density of about 1% expected from a minimum mass solar nebula is ruled out inside ~ 60 AU.

4. The primordial sculpting of the Kuiper belt

The previous section makes clear that the Kuiper belt has lost its accretional disk-like primordial structure, sometime during the solar system history. The goal of modelers is to find the scenario, or the combination of compatible scenarios, that can explain how the Kuiper belt acquired the structural properties discussed above. Achieving this goal would probably shed light on the primordial architecture of the planetary system and its evolution.

Several scenarios have been proposed so far. Some of the Kuiper belt properties discussed in section 3 were not yet known when some of these scenarios have been first presented. Therefore in the following – going beyond the original analysis of the authors – we attempt a critical re-evaluation of the scenarios, confronting them to all the aspects enumerated in the previous section. We divide the proposed scenarios in three groups: (i) those invoking sweeping resonances, which offer a view of gentle evolution of the primordial solar system; (ii) those invoking the action of massive scatterers (lost planets or passing stars), which offer an opposite view of violent and chaotic primordial evolution; (iii) those aimed at building the Kuiper belt as the superposition of two populations with distinctive dynamical histories, somehow combining (i) and (ii).

4.1. RESONANCE SWEEPING SCENARIOS

Fernández and Ip (1984) showed that, while scattering primordial planetesimals, Neptune should have migrated outwards. Malhotra (1993, 1995) realized that, following Neptune’s migration, the mean motion resonances with Neptune also migrated outwards, sweeping the primordial Kuiper belt until they reached their present position. From adiabatic theory (Henrard, 1982), most of the Kuiper belt objects swept by a mean motion resonance would have been captured into resonance; they would have subsequently

followed the resonance in its migration, while increasing their eccentricity. This model accounts for the existence of the large number of Kuiper belt objects in the 2:3 mean motion resonance with Neptune (and also in other resonances) and explains their large eccentricities (see Fig. 8). Reproducing the observed range of eccentricities of the resonant bodies requires that Neptune migrated by 7 AU. Malhotra's simulations also showed that the bodies captured in the 2:3 resonance can acquire large inclinations, comparable to that of Pluto and other objects. The mechanisms that excite the inclination during the capture process have been investigated in detail by Gomes (2000). The author concluded that, although large inclinations can be achieved, the resulting proportion between the number of high inclination vs. low inclination bodies and their distribution in the eccentricity vs. inclination plane do not well reproduce the observations.

The mechanism of adiabatic capture into resonance requires that Neptune's migration happened very smoothly. If Neptune had encountered a significant number of large bodies (Lunar mass or more), its jerky migration would have jeopardized capture into resonances. Hahn and Malhotra (1999), who simulated Neptune's migration using a disk of Lunar to Martian-mass planetesimals, did not obtain any permanent capture. The precise constraints set by the capture process on the size distribution of the largest disk's planetesimals have never been quantitatively computed, but they are likely to be severe.

In the mean motion resonance sweeping model the eccentricities and inclinations of the non-resonant bodies are also excited by the passage of many weak resonances, but the excitation that does occur is too small to account for those observed (compare Fig. 8 with Fig. 1). Some other mechanism (like those discussed below) must also have acted to produce the observed overall orbital excitation of the Kuiper belt. It is an open question whether this other mechanism acted before or after the resonance sweeping and capture process. Had it occurred afterwards, it would have probably ejected from the resonances most of the previously captured objects (not necessarily a problem if the number of captured bodies was large enough). Had it happened before, then the mean motion resonances would have had to capture particles from an excited disk. It has generally been believed that capture into resonance is impossible for particles with moderate or large eccentricities, but in fact it is just less probable (Levison, private communication; see also Gomes, 2000). Another long debated question concerning the sweeping model is the relative proportion between the number of bodies in the 2:3 and 1:2 resonances. The original simulations by Malhotra indicated that the population in the 1:2 resonance should be comparable to – if not greater than – that in the 2:3 resonance. This prediction seemed to be in conflict with the absence of observed bodies in the former resonance at that time. Ida et al. (2000a) showed that the proportion between the two populations

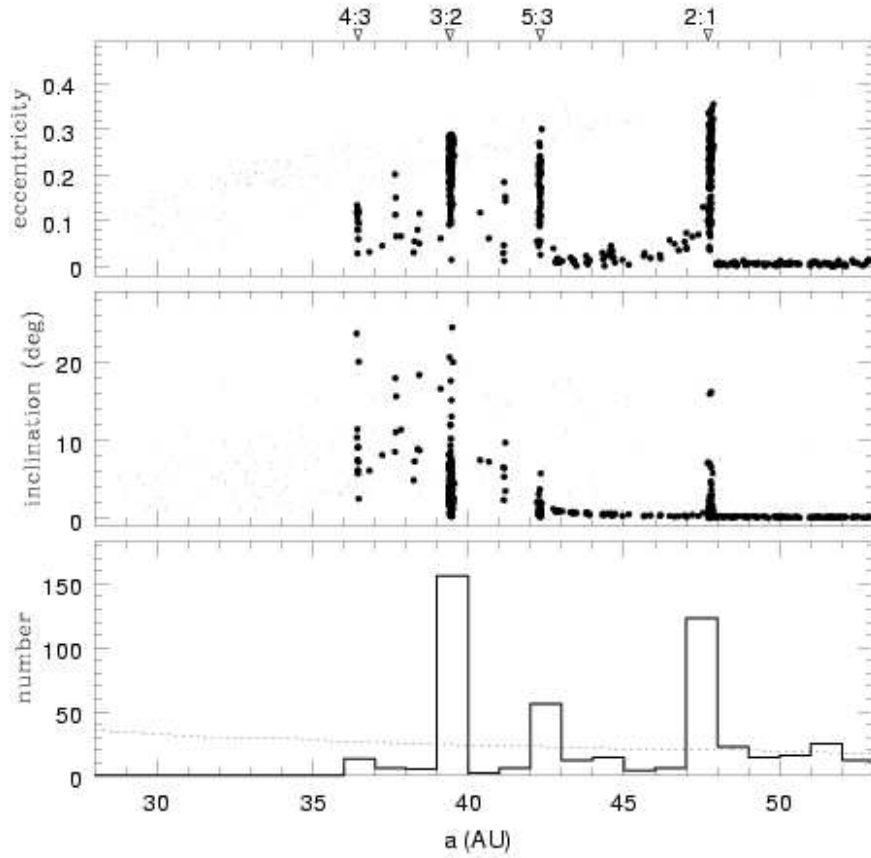


Fig. 8. Final distribution of the Kuiper belt bodies according to the sweeping resonances scenario (courtesy of R. Malhotra). The simulation is done by numerical integrating, over a 200 Myr timespan, the evolution of 800 test particles on initial quasi-circular and coplanar orbits. The planets are forced to migrate (Jupiter: -0.2 AU; Saturn: 0.8 AU; Uranus: 3 AU; Neptune: 7 AU) and reach their current orbits on an exponential timescale of 4 Myr. Large solid dots represent ‘surviving’ particles (i.e., those that have not suffered any planetary close encounters during the integration time); small dots represent the ‘removed’ particles at the time of their close encounter with a planet. In the lowest panel, the solid line is the histogram of semi-major axis of the ‘surviving’ particles; the dotted line is the initial distribution

is very sensitive on Neptune’s migration rate and that the small number of 1:2 resonant bodies, suggested by the lack of observations, would just be indicative of a fast migration (10^5 – 10^6 y timescale). Since then, 5 objects have been discovered in or close to the 1:2 resonance (given orbital uncertainties it is not yet possible to guarantee that all of them are really *inside*

the resonance). There is no general consensus on the de-biased ratio between the populations in the 2:3 and 1:2 resonances, because the de-biasing is necessarily model-dependent and the current data on the population of the 1:2 resonance are sparse. Trujillo et al. (2001) estimated a 2:3 to 1:2 ratio close to 1/2, while Chiang and Jordan (2002) obtained a ratio closer to 3. Chiang and Jordan (2002) also noted that the positions of the five potential 1:2 resonant objects are unusually located with respect to a reference frame rotating with Neptune, which may also have implications for migration rates and capture mechanisms.

The migration of secular resonances could also have contributed to the excitation of the eccentricities and inclinations of Kuiper belt bodies. Secular resonances occur when the precession rates of the orbits of the bodies are in simple ratio with the precession rates of the orbits of the planets. There are several reasons to think that secular resonances could have been in different locations in the past and migrated to their current location at about 40–42 AU. A gradual mass loss of the belt due to collisional activity, the growth of Neptune’s mass and also Neptune’s orbital migration would have moved the secular resonance with Neptune’s perihelion outward. Levison et al. (1997) showed that Kuiper belt interior to 42 AU would have suffered a strong eccentricity excitation. However, the quantitative simulations show that the orbital distribution of the surviving bodies in the 2:3 resonance would not be similar to the observed one: the eccentricities of most simulated bodies would range between 0.05 and 0.1, while those of the observed Plutinos are between 0.1 and 0.3. Also, in this model there is basically no eccentricity and inclination excitation for the Kuiper belt bodies with $a > 42$ AU, in contrast with what is observed.

The dissipation of the primordial nebula would also have caused the migration of the secular resonances. Nagasawa and Ida (2000) showed that the secular resonances involving the precession rates of the perihelion longitudes would have migrated from beyond 50 AU to their current position during the nebula dispersion. This could have caused eccentricity excitation of the Kuiper belt in the 40–50 AU region. In addition, if the midplane of the nebula was not orthogonal to the total angular momentum vector of the planetary system, a secular resonance involving the precession rates of the node longitudes would also have swept the Kuiper belt, causing inclination excitation. The magnitude of the eccentricity and inclination excitation depends on the timescale of the nebula dissipation. A dissipation timescale of $\sim 10^7$ y is required in order to excite the eccentricities up to 0.2–0.3 and the inclinations up to 20–30 degrees. The major failure of the model is that, because only one nodal secular resonance sweeps the belt, all the Kuiper belt bodies acquire orbits with comparably large inclinations. In other words, the model does not reproduce the observed spread of inclinations, nor their bi-modal distribution. No correlation between inclination and color can be

explained either. The same is not true for the eccentricities, because the belt is swept by several perihelion resonances, which causes a spread in the final values. Because of the eccentricity excitation, several Kuiper belt bodies became Neptune-crossing and were subsequently eliminated (following a dynamical path analog to that discussed in Duncan et al. chapter). This process would have removed some fraction of the total mass from the primordial belt, but unfortunately Nagasawa et al. never quantified if this fraction could have been substantial enough to explain the current mass deficit of the Kuiper belt, without the need to invoke collisional grinding. The secular resonance sweeping model cannot explain the existence of significant populations in mean motion resonances, so that the mean motion resonance sweeping model would still need to be invoked.

None of the above discussed models can explain the existence of the edge of the belt at ~ 50 AU.

4.2. SCATTERING SCENARIOS

A radically different view has been proposed by Morbidelli and Valsecchi (1997), who first proposed that massive Neptune scattered planetesimals (mass of order of 1 Earth mass), temporarily on Kuiper belt-crossing orbits, could have excited by close encounters the eccentricities and inclinations of the majority of Kuiper belt objects. This idea has been investigated in details by Petit et al. (1999), who made direct numerical simulations of the effects of scattered massive planetesimals on an initially dynamically cold Kuiper belt. Fig. 9 shows snap shots of the status of the Kuiper belt after respectively 20, 50 and 100 Myr of evolution of an Earth-mass planetesimal in the scattered disk. The test particles (initially 500) were assumed at start on circular and coplanar orbits between 35 and 55 AU. The simulation shows that in the inner belt ($a < 40$ AU) less than 1% of the bodies are found on what would become “stable orbits” once the massive planetesimal is dynamically removed. The depletion factor in the 40–47 AU region is 74% after 20 Myr, 91% after 50 Myr and 96% after 100 Myr. This would completely explain the mass deficit of the current belt. However, beyond 50 AU, $\sim 50\%$ of the original test particles are found on “stable orbits” ($q > 35$ AU) after 100 Myr, which is inconsistent with the observed “edge”. In general terms, this model implies a quite steep positive gradient of the number density of bodies versus semi-major axis, which is not observed in the real population. In particular, the relative population in the 2:3 resonance would be well smaller than that (4% of the classical belt population) claimed from observations by Trujillo et al. (2001). In Petit et al. simulations the median eccentricity and inclination of the survivors after 100 Myr is 0.19 and 8.6° in the 40–47 AU region and 0.27 and 7.4° beyond 47 AU. If the eccentricity distribution correctly reproduces the observations, the inclination distribu-

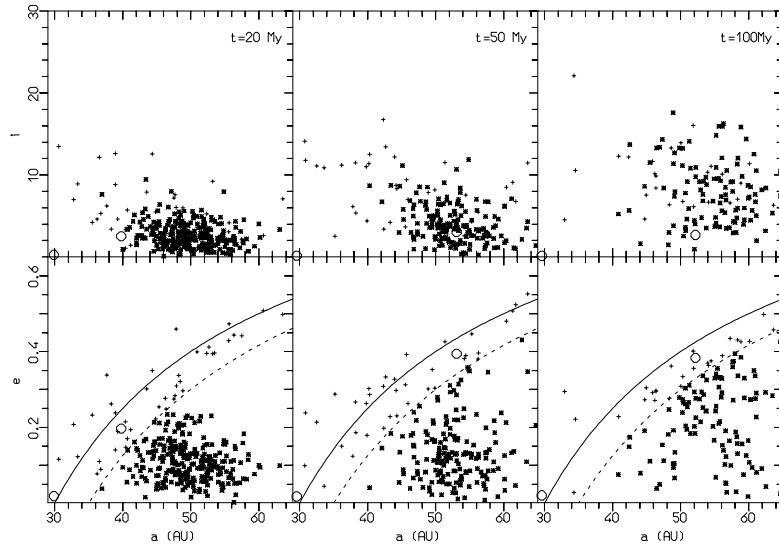


Fig. 9. Snap shots of the Kuiper belt under the scattering action of an Earth-mass planetesimal, itself evolving in the scattered disk. The bold and the dash curves denote $q = 30$ AU and $q = 35$ AU respectively. The latter approximately defines the present limit for stability in the Kuiper belt beyond 42 AU, and therefore marks the transition from the classical belt to the scattered disk. The test particles (initially 500, uniformly distributed on circular and coplanar orbits between 35 and 55 AU) are plotted as an asterisk if $q > 35$ AU, and as a cross otherwise. Neptune and the scattered planetesimal are shown by open circles. From Petit et al. (1999).

tion is not bimodal, and completely misses objects with inclination larger than 20 degrees, in contradiction with the observations. No correlation between inclination and colors could be explained within the framework of this model.

A variant of Petit et al. scenario has been invoked by Brunini and Melita (2002) to explain the apparent edge of the Kuiper belt at 50 AU. They showed with numerical simulations that a Martian mass body residing for 1 Gy on an orbit with $a \sim 60$ AU and $e \sim 0.15-0.2$ could have scattered into Neptune-crossing orbits most of the Kuiper belt bodies originally in the 50–70 AU range, leaving this region strongly depleted and dynamically excited. Such a massive body should have been a former Neptune scattered planetesimal, which decoupled from Neptune due to some diffusive dynamical phenomenon (like the interaction with the 1:3 resonance with Neptune) that temporarily reduced its eccentricity. Although dynamically plausible, this scenario is improbable. The orbital distribution inside ~ 50 AU is not severely affected by the massive planetesimal, once on its decoupled orbit at $a \sim 60$ AU (see also Melita et al., 2002). However, a strong dynamical

excitation could be obtained during the transfer phase, when the massive planetesimals was transported by Neptune encounters towards $a \sim 60$ AU, similarly to what happens in Petit et al. simulations. Some of the simulations by Brunini and Melita that include this transfer phase lead to a (a, e) distribution that is perfectly consistent with what is currently observed in the classical belt in terms of mass depletion, eccentricity excitation, and outer edge (see for instance their Fig. 10). The corresponding inclination distribution is not explicitly discussed, but it is less excited than in the Petit et al. scenario (Melita, private communication). Similarly, a correlation between inclination and color cannot be reproduced by this mechanism, and a distinctive Plutino population is not formed. Finally, our numerical simulations show that an Earth mass planet in Kuiper belt cannot transport bodies up to 200 AU or more by gravitational scattering. Therefore, neither the Petit et al. scenario nor that of Brunini and Melita can explain the origin of the orbit of objects such as 2000 CR₁₀₅.

Motivated by the observation that the eccentricity of the classical belt bodies on average increases with semi-major axis (a fact certainly enhanced by the observational biases, which strongly favor the discovery of bodies with small perihelion distances), Ida et al. (2000b) suggested that the structure of the classical belt records the footprint of the close encounter with a passing star. In that paper and in the follow-up work by Kobayashi and Ida (2001) the resulting eccentricities and inclinations were computed as a function of a/D , where a is the original body's semi-major axis and D is the heliocentric distance of the stellar encounter, for various choices of the stellar parameters (inclination, mass and eccentricity). The eccentricity distribution in the classical belt suggested to the authors a stellar encounter at about ~ 150 AU. The same parameters, however, do not lead to an inclination excitation comparable to the observed one. The latter would require a stellar passage at ~ 100 AU or less. From Kobayashi and Ida simulations we argue that a bi-modal inclination distribution could be possibly obtained, but a quantitative fit to the de-biased distribution discussed in sect. 3.2 has never been attempted. A stellar encounter at ~ 100 AU would make most of the classical belt bodies so eccentric to intersect the orbit of Neptune. Therefore, it would explain not only the dynamical excitation of the belt (although a quantitative comparison with the observed distributions has never been done) but also its mass depletion. Melita et al. (2002) showed that a stellar passage at about 200 AU would be sufficient to explain the edge of the classical belt at 50 AU. Our numerical simulations show that a star passing at ~ 100 AU could scatter objects initially in the 50–60 AU region into the region of the (a, e, i) space occupied by 2000 CR₁₀₅, which is another success of the stellar encounter scenario. The major drawbacks of the stellar scenario are:

- 1) The correlations between inclination and colors cannot be explained.

2) The passage of a star at ~ 100 AU is a low probability event. We tend to reject the idea, championed by Ida et al. (2000b), that the Sun formed in a loose binary system with a companion on a high eccentric orbit, which was rapidly lost on hyperbolic orbit due to the perturbations exerted by the primitive dense stellar environment. In fact, in this case the stellar encounter would have presumably occurred very early in the history of the Solar System, on a timescale of the order of the binary orbital period ($\sim 10^6$ y). But the growth of 100 km size Kuiper belt bodies requires $\sim 10^8$ y even in a dense and dynamically cold environment (Stern, 1996), so that the excitation of the Kuiper belt would have occurred before the accretion of its bodies! A later stellar encounter must necessarily have been a random event. Analyzing the Ipparcos data, Garcia-Sanchez et al. (2001) concluded that, with the current stellar environment, 11.5 stellar encounters within 1 parsec should be expected per My. This implies that the probability that a stellar encounter occurred within 100 AU from the Sun during the age of the Solar System is 1.2%. However, there are arguments based on the Sun's metallicity that the Sun formed closer to the galactic center, which would argue for a denser stellar environment during the early history. Also, the Sun is currently between galactic spiral arms, so the present stellar environment is less dense than the average one and the long-term stellar encounter rate should be higher than the rate derived from Ipparcos data (Weissman, private communication). Nevertheless, it is unlikely that that the probability of a stellar encounter at 100 AU could exceed $\sim 10\%$.

The compatibility of the stellar encounter scenario with the Plutino population is also a debatable issue. Ida et al. (2000b) imagined a 2 stages scenario, in which first the star excited and depleted the outer disk, leaving the disk inside 40 AU dynamically cold, and then Neptune's migration built the Plutino population by resonance sweeping. In this case, the stellar encounter had to occur very early in the solar system history, which raises the problems discussed above. If the stellar encounter occurred after the formation of the Plutino population by resonance sweeping, then many of the high-eccentricity members (those close to aphelion at the time of the encounter) would have probably been ejected from the resonance. Finally, if the star passed sufficiently close (~ 100 AU), as needed to explain the inclination excitation of the classical belt, also the region at 30–40 AU would have been strongly perturbed, destabilizing the Plutino population.

4.3. SCENARIOS FOR A TWO-COMPONENT KUIPER BELT

None of the scenarios discussed above successfully reproduce the existence of a cold and a hot population in the classical belt (see sect. 3.2–3.3) and the correlation between inclination and colors. The reason is obvious. All these scenarios start from a unique population (the primordial, dynamically cold,

Kuiper belt). It is very difficult to produce, from a unique population, two populations with distinct orbital properties. And even in the case where it might be possible (like in the stellar encounter scenario), the orbital histories of grey bodies cannot differ statistically from those of the red bodies, because the dynamics does not depend on the physical properties. The correlations between colors and inclination can be explained only postulating that the hot and the cold populations of the current Kuiper belt originally formed in distinctive places of the solar system. The scenario suggested by Levison and Stern (2001) is that initially the proto-planetary disk in the Uranus–Neptune region and beyond was uniformly dynamically cold, with physical properties that varied with heliocentric distance. Then a dynamical violent event cleared the inner region of the disk, dynamically scattering the inner disk objects outward. In the scattering process, large inclinations were acquired. Most of these objects have been dynamically eliminated, or persist as members of the scattered disk. However a few of these objects somehow were deposited in the main Kuiper belt, becoming the hot population of the classical belt currently observed.

Two dynamical scenarios have been proposed so far to explain how planetesimals in the Uranus–Neptune zone could be permanently trapped in the Kuiper belt.

Thommes et al. (1999) proposed a radical view of the primordial architecture of our outer Solar System, in which Uranus and Neptune formed in the Jupiter–Saturn zone. In their simulations, Uranus and Neptune were rapidly scattered outwards by Jupiter, where the interaction with the massive disk of planetesimals damped their eccentricities and inclinations by dynamical friction; as a consequence, the planets escaped from the scattering action of Jupiter before ejection on hyperbolic orbit could occur. In about 50% of the cases, the final states resembled the current structure of the outer Solar System, with 4 planets roughly at the correct locations. In this scenario Neptune experienced a high-eccentricity phase lasting for a few million years, during which its aphelion distance was larger than the current one. The planetesimals scattered by Neptune during the dynamical friction process therefore formed a scattered disk which extended well beyond its current perihelion distance boundary. When Neptune’s eccentricity decreased down to its present value, the large- q part of the scattered disk became “fossilized”, being unable to closely interact with Neptune again. This scenario therefore explains how a population of bodies, originally formed in the inner part of the disk, could be trapped in the classical belt. However, the inclination excitation of this population, although relevant, is smaller than that of the observed hot population. This is probably due to the fact that Neptune’s eccentricity is rapidly damped, so that the particles undergo Neptune’s scattering action for only a few million years, too short a timescale to acquire large inclinations. For the same reason, the “fossilized” scattered disk does

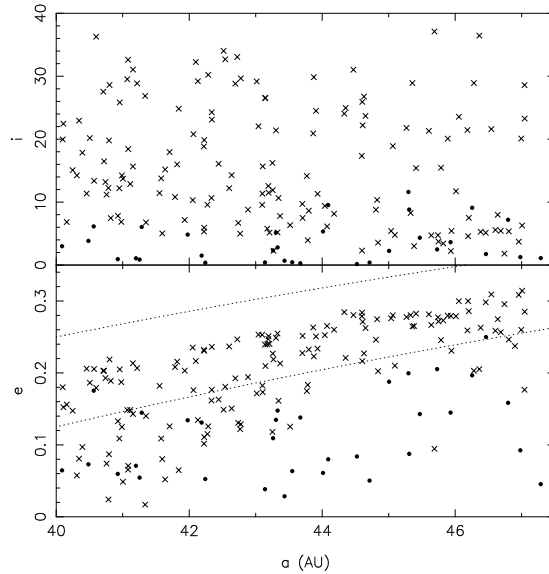


Fig. 10. The orbital distribution in the classical belt according to Gomes' simulations. The dots denote the local population, which is only moderately dynamically excited. The crosses denote the bodies that were originally inside 30 AU. Therefore, the resulting Kuiper belt population is the superposition of a dynamically cold population and of a dynamically hot population, which gives a bi-modal inclination distribution comparable to that observed. The dotted curves in the eccentricity vs. semi-major axis plot correspond to $q = 30$ AU and $q = 35$ AU. Courtesy of R. Gomes.

not extend very far in semi-major axis, so that objects like 2000 CR₁₀₅ are not produced in this scenario. Also, the high eccentricity of Neptune would destabilize the bodies in the 2:3 resonance, so that the Plutinos could have been captured only after Neptune's eccentricity damping, during a final quiescent phase of radial migration similar to that in Malhotra's scenario. Nevertheless, a Plutino population was never formed in the Thommes et al. simulations, possibly because Neptune's migration was too jerky owing to the encounters with the massive bodies used in the numerical representation of the disk.

Gomes (2002) revisited Malhotra's model. Like Hahn and Malhotra (1999) he attempted to simulate Neptune's migration, starting from about 15 AU, by the interaction with a massive planetesimal disk extending from beyond Neptune's initial position. But, taking advantage of the improved computer technology, he used 10,000 particles to simulate the disk population, with individual masses roughly equal to twice the Pluto's mass, while Hahn and Malhotra used only 1,000 particles, with Lunar to Martian masses. In his simulations, during its migration Neptune scattered the planetesimals and

formed a massive scattered disk. Some of the scattered bodies decoupled from the planet, by decreasing their eccentricity through the interaction with some secular or mean-motion resonance. If Neptune had not been migrating, as in Duncan and Levison (1997) integrations, the decoupled phases would have been transient, because the eccentricity would have eventually increased back to Neptune-crossing values, the dynamics being reversible. But Neptune's migration broke the reversibility, and some of the decoupled bodies managed to escape from the resonances, and remained permanently trapped in the Kuiper belt. As shown in Fig. 10, the current Kuiper belt would therefore be the result of the superposition of these bodies with the local population, originally formed beyond 30 AU and only moderately excited (by the resonance sweeping mechanism, as in Hahn and Malhotra, 1999). Unlike in Thommes et al. simulations, the migration mechanism is sufficiently slow (several 10^7 y) that the scattered particles have the time to acquire very large inclinations, consistent with the observed hot population. The resulting inclination distribution of the bodies in the classical belt is bimodal, and quantitatively reproduces the de-biased inclination distribution computed by Brown (2001) from the observations. For the same reason (longer timescale) the extended scattered disk in Gomes simulations reaches much larger semi-major axes than in Thommes et al. integrations. Although bodies on orbits similar to that of 2000 CR₁₀₅ are not obtained in the nominal simulations, other tests done in Gomes (2002) are suggestive that such orbits could be achieved in the framework of the same scenario. A significant Plutino population is also created in Gomes' simulations. This population is also the result of the superposition of the population coming from the Neptune's region with that formed further away and captured by the 2:3 resonance during the sweeping process. Assuming that the bodies' color varied in the primordial disk with heliocentric distance, this process would explain why the Plutinos, scattered objects, and hot classical belt objects, which mostly come from regions inside ~ 30 AU, all appear to have identical color distributions, while only the cold classical population, the only objects actually formed in the trans-Neptunian region, has a different distribution.

Of all the models discussed in this paper, Gomes' scenario is the one that seems to best account for the observed properties of the classical belt. A few open questions persist, though. The first concerns the mass deficit of the Kuiper belt. In Gomes' simulations about 0.2% of the bodies initially in the Neptune-swept disk remained in the Kuiper belt at the end of Neptune's migration. Assuming that the primordial disk was ~ 100 Earth masses, this is well compatible with the estimated current mass of the Kuiper belt. But the local population was only moderately excited and not dynamical depleted, so that it should have preserved most of its primordial mass. The latter should have been several Earth masses, in order to allow the growth of

~ 100 km bodies within a reasonable timescale (Stern, 1996). How did this local population lose its mass? This problem is open also for the Thommes et al. scenario. The only plausible answer seems to be the collisional erosion scenario, but which has the limitations discussed in sect. 3.1. Quantitative simulations need to be done. A second problem, also common to Thommes et al. scenario, is the existence of the Kuiper belt edge at 50 AU. In fact, in neither scenario is significant depletion of the pristine population beyond this threshold obtained. A third problem with Gomes' scenario concerns Neptune's migration. Why did it stop at 30 AU? There is no simple explanation within the model, so that Gomes had to artificially impose the end of Neptune's migration by abruptly dropping the mass surface density of the disk at ~ 30 AU. A possibility is that, by the time that Neptune reached that position, the disk beyond 30 AU had already been severely depleted by collisions. A second possibility is that something (a massive planetary embryo?) opened a gap in the disk at about 30 AU, so that Neptune run out of material and could not sustain further its migration.

5. Conclusions

Ten years of dedicated surveys have revealed unexpected and intriguing properties of the trans-Neptunian population, such as the existence of a large number of bodies trapped in mean motion resonances, the overall mass deficit, the large orbital eccentricities and inclinations, the apparent existence of an outer edge at ~ 50 AU and of a correlation between inclinations and colors. Understanding how the Kuiper belt acquired all these properties would probably constraint several aspects of the formation of the outer planetary system and of its primordial evolution.

Up to now, a portfolio of scenarios have been proposed by theoreticians. None of them can account for all the observations alone, and the solution of the Kuiper belt primordial sculpting problem probably passes through a sapient combination of the proposed models. The Malhotra–Gomes scenario does a quite good job at reproducing the observed orbital distribution inside 50 AU, although we still do not properly understand how the cold classical population lost most of its primordial mass. However, the model runs in conflict with the apparent edge of the belt at 50 AU. Such an edge is currently explained only in two scenarios: that of the stellar encounter of Ida et al., and that of the rogue planet trapped at ~ 60 AU, championed by Brunini and Melita. The first one predicts a sharp boundary of the belt, while the second one predicts the existence of a gap in the 50–70 AU range, beyond which the density of the population should gradually increase again. But neither of these scenarios explains very well the architecture of the belt inside 50 AU. Can a combination of these scenarios account for all observables? Was the primordial evolution of the trans-Neptunian region really that tormented?

Or are we misinterpreting some observational constraints, so that the various pieces of information cannot be coherently fit by any unitary model? It is certainly true that of the observational constraints discussed here, the existence of an edge to the Kuiper belt is the most model dependent and therefore subject to revision.

Kuiper belt science is a rapidly evolving one. New observations change our view of the belt every year. Since the discovery of the first trans-Neptunian object 10 years ago several review papers have been written, and most of them are already obsolete. No doubt that this will also be the fate of this chapter, but it can be hoped that the ideas presented here can continue to guide us in the direction of further understanding of what present observations of the Kuiper belt can tell us about the formation and evolution of the outer solar system.

6. References

- Allen, R.L., Bernstein, G.M., Malhotra, R. 2001. The edge of the solar system. *Astroph. J.*, **549**, L241-L244.
- Brown M. 2001. The Inclination Distribution of the Kuiper Belt. *Astron. J.*, **121**, 2804-2814.
- Brown, M.E., Trujillo, C.A. 2003. The plane of the Kuiper belt. *in preparation*.
- Brunini A., Melita M. 2002. The existence of a planet beyond 50 AU and the orbital distribution of the classical Edgeworth Kuiper belt objects. *Icarus*, in press.
- Chiang, E.I., Brown, M.E. 1999. Keck pencil-beam survey for faint Kuiper belt objects. *Astron. J.*, **118**, 1411-1422.
- Chiang E. I., Jordan A. B. 2002. On the plutinos and Twotinos of the Kuiper belt. *Astron. J.*, in press.
- Davis D. R., Farinella P. 1997. Collisional Evolution of Edgeworth-Kuiper Belt Objects. *Icarus*, **125**, 50-60.
- Davis D. R., Farinella P. 1998, Collisional Erosion of a Massive Edgeworth-Kuiper Belt: Constraints on the Initial Population. In *Lunar Planet. Science Conf.* **29**, 1437-1438.
- Dones, L. 1997. in ASP Conf. Ser. 122, From Stardust to Planetesimals, ed. Y.J. Pendleton & A.G.G.M. Tielens (San Francisco: ASP), 347.
- Doressoundiram A., Barucci M.A., Romon J., Veillet C. 2001. Multicolor Photometry of Trans-neptunian Objects. *icarus*, **154**, 277-286.
- Duncan, M. J., Levison, H. F., Budd, S. M. 1995, The long-term stability of orbits in the Kuiper belt, *Astron. J.*, **110**, 3073-3083.
- Duncan, M. J., Levison, H. F. 1997, Scattered comet disk and the origin of Jupiter family comets, *Science*, **276**, 1670-1672.
- Emel'yanenko V. 2002. A flux of distant trans-Neptunian objects. In *Asteroid, Comet, Meteors*, ESA Spec. Publ. series, in press.

- Farinella P., Davis D.R., Stern S.A. 2000. Formation and Collisional Evolution of the Edgeworth-Kuiper Belt. In *Protostars and Planets IV* Mannings V., Boss A.P., Russell S. S. eds., University of Arizona Press, Tucson, 125–133.
- Fernández J. A., Ip W. H. 1996. Orbital expansion and resonant trapping during the late accretion stages of the outer planets. *Pl. Sp. Sci.*, **44**, 431–439.
- Garcia-Sanchez J., Weissman P.R., Preston R.A., Jones D.L., Lestrade J.F., Latham D.W., Stefanik R.P., Paredes J.M. 2001. Stellar encounters with the solar system. *Astron. Astropys.*, **379**, 634–659.
- Gladman, B., Kavelaars, J.J., Nicholson, P.D., Loredó, T.J., Burns, J.A. 1998. Pencil-beam surveys for faint trans-Neptunian objects. *Astron. J.*, **116**, 2042–2054.
- Gladman, B., Kavelaars, J.J., Petit, J.M., Morbidelli, A., Holman, M.J., Loredó, Y., 2001. The structure of the Kuiper belt: Size distribution and radial extent. *Astron. J.*, **122**, 1051–1066.
- Gladman B., Holman M., Grav T., Kaavelars J.J., Nicholson P., Aksnes K., Petit J.M. 2002. Evidence for an extended scattered disk. *Icarus*, **157**, 269–279.
- Gomes R. S. 2000. Planetary Migration and Plutino Orbital Inclinations *Astron. J.*, **120**, 2695–2707.
- Gomes R.S. 2002. The origin of the Kuiper belt high inclination population. *Icarus*, submitted.
- Hahn J. M., Malhotra R. 1999. Orbital Evolution of Planets Embedded in a Planetesimal Disk. *Astron. J.*, **117**, 3041–3053.
- Hainaut, O. 2002 <http://www.sc.eso.org/~ohainaut/MBOSS/>
- Henrard J. 1982. Capture into resonance - An extension of the use of adiabatic invariants. *Cel. Mech.*, **27**, 3–22.
- Ida S., Bryden G., Lin D. N., Tanaka H. 2000a. Orbital Migration of Neptune and Orbital Distribution of Trans-Neptunian Objects. *Astroph. J.*, **534**, 428–445.
- Ida S., Larwood J., Burkert A. 2000b. Evidence for Early Stellar Encounters in the Orbital Distribution of Edgeworth-Kuiper Belt Objects. *Astroph. J.*, **528**, 351–356.
- Jewitt, D. C., Luu, J. X. 1993, Discovery of the candidate Kuiper belt object 1992 QB1, *Nature*, **362**, 730–732.
- Jewitt, D., Luu, J., Chen, J. 1996. The Mauna-Kea-Cerro-Totololo (MKCT) Kuiper belt and Centaur survey. *Astron. J.*, **112**, 1225–1232.
- Jewitt, D., Luu, J., Trujillo, C. 1998. Large Kuiper belt objects: The Mauna Kea 8K CCD survey. *Astron. J.*, **115**, 2125–2135.
- Kenyon, S.J., Luu, J.X. 1998. Accretion in the early Kuiper belt: I. Coagulation and velocity evolution. *Astron. J.*, **115**, 2136–2160.

- Kenyon, S.J., Luu, J.X. 1999a. Accretion in the early Kuiper belt: II. Fragmentation. *Astron. J.*, **118**, 1101-1119.
- Kenyon, S.J., Luu, J.X. 1999b. Accretion in the early outer solar system. *Astrophys. J.*, **526**, 465-470
- Kobayashi H., Ida S. 2001. The Effects of a Stellar Encounter on a Planetary Disk. *Icarus*, **153**, 416-429.
- Kuchner M.J., Brown M.E., Holman M. 2002. Long-Term Dynamics and the Orbital Inclinations of the Classical Kuiper Belt Objects. *Astron. J.*, **124**, 1221-1230.
- Kuiper, G.P. 1951. On the origin of the Solar System. In *Astrophysics*, ed. Hynek, J.A., McGraw-Hill, New York, 357 pp.
- Levison, H.F., & Duncan, M. J. 1997, From the Kuiper Belt to Jupiter-Family Comets: The Spatial Distribution of Ecliptic Comets, *Icarus*, **127**, 13-32.
- Levison H.F., Stern S. A., Duncan, M. J. 1997. The role of a massive primordial Kuiper belt on its current structure. *Icarus*, submitted.
- Levison H.F., Stern S.A. 2001. On the Size Dependence of the Inclination Distribution of the Main Kuiper Belt. *Astron. J.*, **121**, 1730-1735.
- Lewis, J.S. 1995. *Physics and Chemistry of the Solar System*, Academic Press, San Diego.
- Malhotra R. 1993. The origin of Pluto's peculiar orbit. *Nature*, **365**, 819-821.
- Malhotra R. 1995. The Origin of Pluto's Orbit: Implications for the Solar System Beyond Neptune. *Astron. J.*, **110**, 420-432.
- Margot, J.-L., Trujillo, C.A., Brown, M.E., Bertoldi, F. 2002. BAAS, in press.
- Melita M., Larwood J., Collander-Brown S, Fitzsimmons A., Williams I.P., Brunini A. 2002. The edge of the Edgeworth-Kuiper belt: stellar encounter, trans-Plutonian planet or outer limit of the primordial solar nebula? In *Asteroid, Comet, Meteors*, ESA Spec. Publ. series, in press.
- Morbidelli A., Valsecchi G. B. 1997. Neptune scattered planetesimals could have sculpted the primordial Edgeworth-Kuiper belt, *Icarus*, **128**, 464-468.
- Morbidelli A. 2002 *Modern celestial mechanics: aspects of solar system dynamics*, Taylor and Francis, London and New York.
- Nagasawa M., Ida S. 2000. Sweeping Secular Resonances in the Kuiper Belt Caused by Depletion of the Solar Nebula. *Astron. J.*, **120**, 3311-3322.
- Petit J. M., Morbidelli A., Valsecchi, G. B. 1999. Large scattered planetesimals and the excitation of the small body belts. *Icarus*, **141**, 367-387.
- Stern S. A. 1991. On the number of planets in the outer solar system - Evidence of a substantial population of 1000-km bodies, *Icarus*, **90**, 271-281.
- Stern SA. 1995. Collisional time scales in the Kuiper disk and their implications. *Astron. J.*, **110**, 856-868.
- Stern, S. A. 1996, On the Collisional Environment, Accretion Time Scales, and Architecture of the Massive, Primordial Kuiper Belt., *Astron. J.*,

- 112**, 1203–1210.
- Stern, S. A., Colwell, J. E. 1997a, Accretion in the Edgeworth-Kuiper Belt: Forming 100-1000 KM Radius Bodies at 30 AU and Beyond. *Astron. J.*, **114**, 841-849.
- Stern, S. A., Colwell, J. E. 1997b, Collisional Erosion in the Primordial Edgeworth-Kuiper Belt and the Generation of the 30-50 AU Kuiper Gap, *Astroph. J.*, **490**, 879–885.
- Stern, S.A. 2002. Evidence for a collisional mechanism affecting Kuiper belt object colors. *Astron. J.*, in press.
- Tegler, S. C.; Romanishin, W. 2000. Extremely red Kuiper-belt objects in near-circular orbits beyond 40 AU. *Nature*, **407**, 979-981.
- Thommes E. W., Duncan M. J., Levison H. F. 1999. The formation of Uranus and Neptune in the Jupiter-Saturn region of the Solar System. *Nature*, **402**, 635–638.
- Trujillo C.A., Brown M.E. 2001. The Radial Distribution of the Kuiper Belt. *Astroph. J.*, **554**, 95-98.
- Trujillo, C. A., Jewitt D. C., Luu J. X. 2001. Properties of the Trans-Neptunian Belt: Statistics from the Canada-France-Hawaii Telescope Survey. *Astron. J.*, **122**, 457-473.
- Trujillo C.A., Brown M.E. 2002. A Correlation between Inclination and Color in the Classical Kuiper Belt. *Astroph. J.*, **566**, 125-128.