
Kuiper Belt and Comets: An Observational Perspective

David Jewitt¹

1. Institute for Astronomy, University of Hawaii,
2680 Woodlawn Drive, Honolulu, HI 96822
jewitt@hawaii.edu

Note to the Reader

These notes outline a series of lectures given at the Saas Fee Winter School held in Murren, Switzerland, in March 2005. As I see it, the main aim of the Winter School is to communicate (especially) with young people in order to inflame their interests in science and to encourage them to see ways in which they can contribute and maybe do a better job than we have done so far. With this in mind, I have written up my lectures in a less than formal but hopefully informative and entertaining style, and I have taken a few detours to discuss subjects that I think are important but which are usually glossed-over in the scientific literature.

1 Preamble

Almost exactly 400 years ago, planetary astronomy kick-started the era of modern science, with a series of remarkable discoveries by Galileo concerning the surfaces of the Moon and Sun, the phases of Venus, and the existence and motions of Jupiter's large satellites. By the early 20th century, the focus of astronomical attention had turned to objects at larger distances, and to questions of galactic structure and cosmological interest. At the start of the 21st century, the tide has turned again. The study of the Solar system, particularly of its newly discovered outer parts, is one of the hottest topics in modern astrophysics with great potential for revealing fundamental clues about the origin of planets and even the emergence of life. New technology has been crucial to each of these steps. Galileo's refractor gave a totally new view of the sky. A hundred years ago, photographic plates and large telescopes allowed the first spectroscopic observations of distant galaxies revealing, through Hubble's law, the 3rd dimension of distance into the plane of the sky. In our own time, highly sensitive, wide-field electronic detectors have enabled the discovery and the exploration of the Kuiper Belt, while fast computers allow

us to make numerical simulations with a degree of sophistication that was previously unimaginable.

As a result of all this, our view of the Solar system is in the middle of a great change. Our appreciation of the different types of objects (planets, asteroids, comets, etc) orbiting the Sun is changing in response to new observations. Our understanding of their evolutionary connections with each other and with the formation epoch is changing as we develop more and more elaborate schemes to synthesize the new data. Additionally, our perception of the Solar system in the bigger context of the galactic disk is changing, particularly as we detect planets encircling other stars (in systems that are, for the most part, dynamically not very like our own). All of this makes it a great time to review what we know about the Solar system in the context of the Saas Fee winter school series, one of very few Saas Fee lectures to be dedicated to the universe at $z \sim 0$.

This article parallels five lectures given in Murren, Switzerland in March 2005, as part of the Saas Fee Lecture Series entitled “Trans-Neptunian Objects and Comets”. Some of these lectures were given “off the cuff” and I have tried to reconstruct them from memory and a few notes. The degree to which this succeeds is unknown and it doesn’t matter: the participants, like this lecturer, have no doubt forgotten most of what was said while readers who were not in Saas Fee for the Lecture Series never knew. The style of the write-up is deliberately informal.

1.1 The Conduct of Research into the Subject

In this section I want to take advantage of the open format of the Saas Fee lecture series to briefly discuss the conduct of modern science, particularly as it relates to the new study of the Solar system. Partly this is for fun and for my own entertainment, but I also have a serious purpose: there are real misconceptions about what is happening (as opposed to what should happen), sometimes even in the minds of the best scientists. Most of us probably possess vaguely Popperian ([123]) notions about the conduct of science. Essentially, Popper argued that we advance in science by the falsification of hypotheses. Observations suggest hypotheses which make predictions which can be confirmed or refuted by new observations, and so on. But not all of us work within this framework and there are few clues as to the real methods or motivations of scientists in the stylized and frequently dry presentations that are demanded for publication in the refereed journals. It is the absence of discussion about the realities of the practice of science that has allowed false ideas to spread unchecked. The Saas Fee participants, especially those likely to become major figures in the future exploration of the Solar system, are the main targets of my remarks.

Observations

Observationally, the goal is to determine objective reality through careful studies that are unbiased (or at least well calibrated), fully understood, independently reproducible and motivated by the desire to test a hypothesis. Several things must be said about this idealized goal.

- Real science is much more affected by chance discoveries than one would guess from the simple description of Popper's scheme, above. Sometimes, the biggest advance comes from simply looking, not from testing a hypothesis.
- The flip-side of this is that the human brain is rarely able to perceive or assimilate things that it does not expect to see and so fundamental discoveries made by chance are very rare (but disproportionately important). We are like ants in the city: comfortable with the dirt in front of us but unable to perceive the buildings above.



Fig. 1. Observers, doing what they do best. Photo courtesy Nik Wilhelm.

- Although it seems that it should be otherwise, taking good observations is incredibly hard. Too many things can go wrong, there are many sources of error both random and systematic and it is often difficult or impossible to accurately quantify these uncertainties. As a result, observations that seem secure (or “statistically significant” as we say with a misleading air of detachment) are often wrong, leading us up blind alleys that can take years to escape.

- An equally serious problem is that it is easy to take the “wrong” measurement, by which I mean a measurement that has no great impact on our perception of the big picture. In fact most observers, including this one, spend most of their time taking measurements that are unimportant. The simple reason is that we usually cannot see clearly enough to predetermine which measurements will be of the greatest value. Theories and models are supposed to help us here: usually they don’t.

As observers, we are swimming in mud (Figure 1): it’s hard work, we can’t see where we are going but sometimes we bump into interesting things as we crawl our way along.

Theories and Models

The purpose of theories and models is to use available data together with established physical laws to make observationally testable predictions. Predictions provide an objective and indispensable way to test the theories and models. Unfortunately, theory rarely works this way, because the systems under consideration are very complicated and a large number of processes interact in a way that is difficult to treat. Making observationally testable predictions is difficult because a given model, with changes to one or two of its many free parameters, can usually accommodate a wide range of outcomes, regardless of whether the model is correct. Making predictions that are falsifiable is the hard part of making models, which is why many modelers don’t do it.

Here are some problems with theory and theorists.

- The main problem for theorists and modelers is that the world is very complex and most problems are observationally under-constrained. Analytical approaches offer real insight and understanding, but are mostly confined to the study of highly simplified approximations. Numerical approaches provide a way to deal with the complexity, but at the expense of adding typically large numbers of under-determined model parameters and initial conditions.
- It has become common to present models which fit the available data but which offer no observationally testable predictions, leaving the reader to speculate about what predictions the model might make if only the authors had written them down. The reason for this is clear enough: making observationally testable predictions is difficult (and scary too: you could be wrong!). But without predictions the models have no scientific value. Some have argued that the mere fact that a model can fit many and varied observations in a self-consistent way is evidence in itself for the correctness of the model. Nonsense!
- The meaning of the word “predict” is also under attack. Sometimes, the authors say that their model “predicts” some quantity or property but

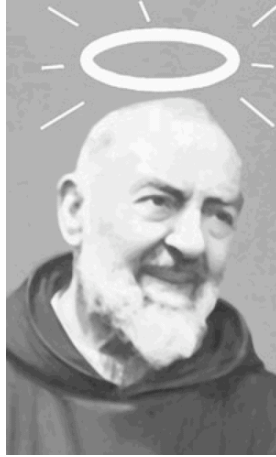


Fig. 2. The theorist, spotlessly clean, whose theory explains everything and has no free parameters. The halo and the facial expression signify his wisdom and purity. Courtesy Virginia A. Tikan.

closer inspection shows that the thing has already been measured. One cannot predict something which is already known! What the modelers mean is that they can fit the data, not predict new data. There's a big difference.

- Models are frequently over-sold (Figure 2). It is almost *de rigueur* for modelers to add comforting phrases like “our conclusions are insensitive to the parameters assumed in the model...” and “our model has only one free parameter...” whether or not these statements are true!

Of course, it is the interaction between the observers and the theorists that gives our subject its extraordinary vitality and power. Science without observations would collapse into dull paralysis within months. Science without models would soon degenerate into stamp collecting. But this doesn't mean that we have to accept either the observations or the models uncritically. In particular, we should not accept models which fail to make observationally testable predictions. They may offer beautiful descriptions of what we observe but, without predictions, we'll never know if they have deeper meaning.

The Kuiper belt is still very much in the discovery phase and we should not expect a scientifically compelling picture of its formation and evolution to emerge overnight. With this warning of a turbulent and uncertain background, we are ready to launch into an overview of the modern Solar system.

2 The Modern Solar System

2.1 Protoplanetary Disk

Scale Constraints

The most noticeable feature of the Solar system is that the planets follow nearly circular orbits about the Sun in roughly the same plane. This architecture strongly suggests that the planets formed by accretion in a circum-Solar disk. The properties of this disk, now long-gone, can be inferred only approximately from the modern-day system.

The extent of the Solar nebula is not tightly observationally constrained, but again we can set limits. At the inner edge it is reasonable to suppose that the disk extended practically to the surface of the protosun. Indeed, material flowed through the disk into the Sun as part of its formation. At the outer edge, we surmise that the disk extended to roughly the outer extremity of the well-established part of the Kuiper belt (roughly 50 AU). Observations of disks around other stars show that disks are commonly much larger, extending to 100's AU around stars of Solar mass. The timescales for the growth of solid bodies scale with heliocentric distance, R , as R^3 , give or take one power of R . One possibility is that the protoplanetary disk may initially have been 100's of AU in extent but that no large bodies grew in the outer parts. In this case deeper survey observations should reveal smaller bodies beyond the ~ 50 AU edge, something that seems not to be true. Another possibility is that the small size of the Kuiper belt (specifically of the classical belt) results from tidal truncation by a passing star, as argued by Ida et al. ([66]) and others since.

Structure Constraints

The current mass of the objects in the Solar system (excluding the Sun) is about $10^{-3} M_{\odot}$, most of which is in Jupiter. Obviously, this sets a strong lower limit to the initial mass of the disk. A more realistic limit is set by careful consideration of the compositions of the planets and the (probably good) assumption that the disk had a basically cosmic composition. For instance, consider the Earth. Its mass consists mostly of heavy elements (called “metals” by terminology-bending astrophysicists) whereas, in a mixture containing a cosmic proportion of H and He, the “metals” carry only ~ 0.01 of the mass. Therefore, the so-called augmented mass of the Earth (the mass of material of cosmic composition containing an Earth mass of metals) is about $100 M_{\oplus}$. This same treatment of the other planets leads to a best estimate of the minimum disk mass of order $0.01 M_{\odot}$. Models with this mass are known as MMSN models: Minimum Mass Solar Nebula models.

The distribution of mass and temperature within the protoplanetary disk are usually approximated by power laws

$$\Sigma(R) = \Sigma(R_0) \left[\frac{R_0}{R} \right]^p \quad (1)$$

$$T(R) = T(R_0) \left[\frac{R_0}{R} \right]^q \quad (2)$$

where $\Sigma(R)$ [kg m^{-2}] and $T(R)$ are the column density and temperature of the disk at radius R , R_0 is a reference radius, often taken as 1 AU (the orbit of Earth) or 10 AU (orbit of Saturn), and the indices p and q describe the radial fall-off of the density and temperature, respectively. Estimates of Σ_0 and p can be obtained by studying the distribution of mass within the Solar system. If we smear the augmented masses of the planets over annuli extending half way to the nearest planet (e.g. Saturn would be smeared from 7.5 AU to 15 AU) we obtain $p \sim 3/2$ (with an uncertainty of at least $\pm 1/2$) and $\Sigma(R_0) \sim 50 \text{ kg m}^{-2}$ at $R_0 = 10 \text{ AU}$. This is the total (gas plus dust) surface density. The dust surface density is about 100 times smaller. The temperature of a blackbody in radiative equilibrium with sunlight is described by Equation 2 with $T(R_0 = 10) = 88 \text{ K}$ and $q = 0.5$.

The values of disk parameters so derived are not particularly accurate, given the uncertainties in computing augmented masses from current masses and given the likelihood that the orbits of the planets were not always where we now find them. Still, the above give a reasonable starting guess for the structure of the disk. It is natural to think that observations of disks around young stars should provide independent constraints on likely disk parameters. Unfortunately, most existing data generally lack angular resolution high enough for the disk spatial parameters to be directly measured. Instead, the disk parameters are inferred from measurements of the spectral energy distribution using models in which the number of free parameters is larger than the number of observational constraints. *Assuming* $p = 3/2$, measurements give mean values $q = 0.6 \pm 0.1$ and $T(10 \text{ AU}) = 45 \pm 21 \text{ K}$ [4], which fit well with the nominal values. The dust mass inferred from disk observations averages $M_d = 4 \times 10^{-3} M_\odot$ ([4]; from 67 classical T-Tauri stars, likely analogs of the young Sun). The dust mass is really a lower limit to the mass in solids: particles much larger than the millimeter wavelengths of observation contribute little to the measured radiation and go undetected. Augmented to cosmic composition, the implied average disk mass is $\sim 0.4 M_\odot$. This is substantially larger than MMSN but the scatter in disk masses is large, as are the uncertainties, and there are presumably observational biases against the measurement of lower disk masses.

Constraints on Disk Timescales and Environment

The most important observational constraints on timescales in the protoplanetary disk are provided by measurements of the products of radioactive decay

of short-lived elements in meteorites. The latter are rocks derived by shattering collisions amongst the asteroids and delivered to Earth by gravitational scattering after their orbits become planet-crossing. Minerals in many meteorites incorporate the decay products of short-lived nuclei, showing that the minerals formed on timescales comparable to the half-lives of the decaying elements. The quintessential example is provided by ^{26}Al , which β -decays into ^{26}Mg with a half-life $t_{1/2} = 0.7$ Myr [90]. When ^{26}Mg is found incorporated within the mineral structure of a meteorite, we may conclude that ^{26}Al was originally present. To be captured in abundance, ^{26}Al must have been incorporated into the meteorites within a few half-lives of its formation. Element formation occurs naturally in the explosion of massive stars as supernovae but the significance of ^{26}Al has sometimes been questioned because it can be also formed by spallation reactions with particles accelerated to energies $> \text{MeV}$ [91]. Such particles might have been emitted by the magnetically super-active young sun. Recent measurements of ^{60}Ni , which is produced by the decay of ^{60}Fe with a half-life of 1.5 Myr [115], do not suffer this ambiguity since there is no route to its production through spallation. We conclude with confidence that macroscopic solid bodies formed in the asteroid belt on timescales of a few Myr.

Other timescale constraints come from observations of circumstellar matter in disks around nearby Solar-mass stars. These observations show that circumstellar gas has a lifetime that is less than 10 Myr ([159]; [10]) and potentially just a few Myr. Dust emission from stars also declines rapidly with age (Figure 3). The initial decline is probably due to growth into particles that are much larger than the wavelength of observation (typically ~ 1 mm). There is evidence for thermal excess above the emission from the stellar photospheres in stars as old as ~ 0.5 Gyr and this dust is probably produced in recent times by collisions among unseen bodies in the circumstellar disks, or released by unseen comets. The general decline in the dustiness of nearby stars is occasionally punctuated by objects with surprising dust emission excess. This could be showing that the stars are, for some reason, intrinsically more dusty than others of similar age. An alternative explanation is that the dust has been recently created, perhaps by impact and shattering of massive planetesimals in the unseen circumstellar disks ([129]).

Two pieces of evidence suggest that the Sun formed in a star cluster.

First, some of the short-lived radionuclides (notably ^{60}Fe) must have been produced, in an exploding star, only shortly before their incorporation into minerals and meteorite parent bodies (asteroids) otherwise they would have already decayed to insignificance. Supernovae are very rare (the galactic rate is only \sim one per 50 years) and typically distant so that the likelihood of having one occur nearly simultaneously with the formation of solid bodies in the disk is small. The simplest interpretation is that the Sun was part of a cluster of stars in which nearby high mass members exploded upon reaching the ends of their stable main-sequence lifetimes. An estimate of the cluster population can be made based on the dual requirements that the cluster must

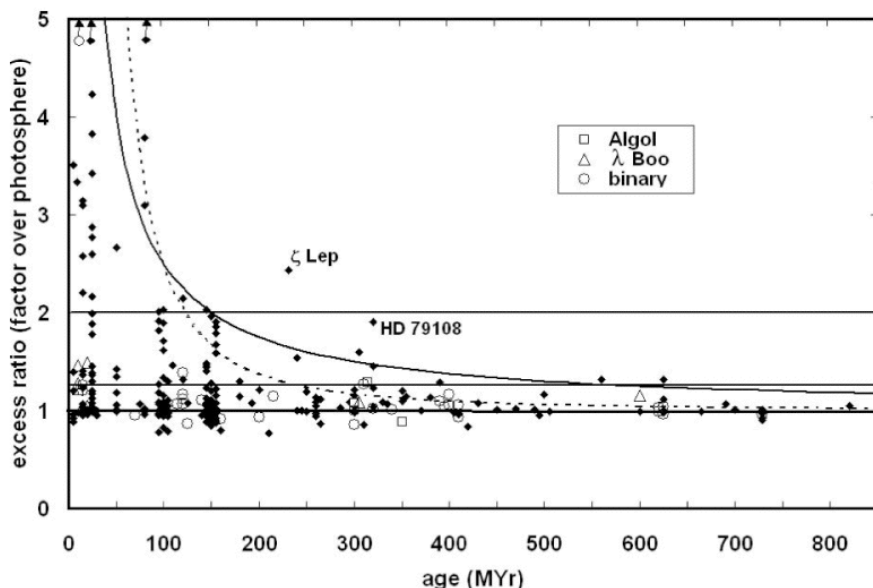


Fig. 3. Dust emission from nearby stars at $24\ \mu\text{m}$ wavelength expressed as a ratio to the flux density expected from the photosphere alone. Values >1 indicate excess emission, most likely from circumstellar dust heated by starlight. The emission generally declines with stellar age but, at any given age, there is a range of thermal excesses, with occasional dramatic spikes, as at $\zeta\ \text{Lep}$ and HD 79108. The solid curve shows a $1/(\text{time})$ dependence. Ages of the stars are estimated from cluster membership and from models of their spectra, and are accurate to about a factor of two. One interpretation of the spikes is that dust is impulsively created by collisions between massive bodies. Figure reproduced from [129].

have been populated enough to contain a massive star capable of reaching supernova status but yet not so populated that gravitational perturbations would have noticeably disturbed the orbits of the planets. A cluster containing $\sim 2000 \pm 1100$ stars seems capable of meeting both conditions [2].

Second, the truncated outer edge of the classical Kuiper belt and the excited dynamical structure of the belt in general suggests to some that the protoplanetary disk might have been tidally truncated by a passing star [66]; [113]. Numerical simulations show that to truncate or seriously disturb the disk down to radius r [AU] implies a stellar impact parameter $\sim 3r$. The classical belt ends near 50 AU, requiring a Solar mass star to pass ~ 150 AU from the Sun. In its current environment, the sun and stars are separated by ~ 1 pc (200,000 AU) and the probability of two stars passing within 150 AU in the 4.6 Gyr age is negligible. Again, a plausible inference is that the mean distance between the Sun and nearby stars was once much smaller: the Sun was in a cluster.

2.2 The Three Domains

It is useful to consider the Solar system as divided into three domains, based on the compositions, masses and radial distances of its constituents. These are as follows:

The Domain of the Terrestrial Planets

The primary objects are Mercury, Venus, Earth and Mars, but the asteroids in the main-belt between Mars and Jupiter are also included (the largest asteroid is (1) Ceres; see Table 1). These objects are all distinguished by refractory (non-volatile) compositions dominated by metals [principally iron (Fe) and nickel (Ni)] and compounds of silicon (Si), oxygen (O), magnesium (Mg) and aluminium (Al). The bulk densities are high ($\rho = 3930 \text{ kg m}^{-3}$ for Mars up to 5515 kg m^{-3} for Earth, the latter slightly enhanced by self-compression due to gravity), reflecting the lack of volatiles. Densities of many asteroids are smaller, apparently as a result of porous internal structures created by impact fragmentation and reassembly of these bodies since their formation. The densities of stony meteorites, small fragments from the asteroid belt, are $\rho \sim 3000 \text{ kg m}^{-3}$.

Table 1. Terrestrial Planets

Object	Mass/ M_{\oplus}	Radius/ R_{\oplus}	$\rho \text{ [kg m}^{-3}\text{]}$	a [AU]	e	i [deg]
Mercury	0.06	0.38	5430	0.387	0.206	7.0
Venus	0.81	0.95	5424	0.723	0.007	3.4
Earth	1	1	5520	1.000	0.017	0.0
Mars	0.11	0.53	3930	1.523	0.093	1.8
Ceres	1.6×10^{-4}	0.08	2080	2.766	0.078	10.6

All these bodies appear to have formed by “binary accretion”, the step-by-step growth occurring when two bodies collide and stick, starting from tiny dust particles in the original nebula about the Sun and reaching up to the sizes of the Earth and Venus. Indeed, the N-body models that are used to study the dynamics and growth of bodies in the outer Solar system have been honed to their highest levels of perfection in the study of Terrestrial planet growth. Still, new data continue to surprise and unnerve us. For example, N-body accretion models show that Earth grew to its final mass on a timescale $\sim 100 \text{ Myr}$ to 200 Myr ([18]; [127]) and this long timescale has remained more or less unchanged for the past several decades, since detailed estimates were first made by G. Wetherill. It stands in contrast to new isotopic data from the Hafnium-Tungsten (Hf-W) decay ([67]). Hafnium decays to Tungsten, $^{182}\text{Hf} \rightarrow ^{182}\text{W}$, with a 9 Myr half life. The quantity of ^{182}W in the Earth’s mantle (relative to the core) provides a measure of the amount of the unstable Hf

isotope at the epoch of core formation, and so sets the timescale for Earth’s differentiation. The W-Hf data show that the Earth accreted the bulk of its mass within 30 Myr, while major asteroids such as Vesta formed in an even shorter 3 Myr ([67]). This is a half to one order of magnitude discrepancy with the N-body models and remains unexplained.

The relevance to us is that models can give very plausible but wholly incorrect solutions. Without the benefit of independent constraints from the isotopes, we would remain completely unaware that the N-body terrestrial planet growth models are too slow. In the outer Solar system (where independent constraints on the models from isotopes or other sources are unavailable) it is easy to see that we are skating on very, very thin ice.

The Domain of the Giant Planets

Gas Giants

Jupiter and Saturn (Figures 4 and 5), in addition to being two orders of magnitude more massive than the Terrestrial planets (see Table 2), have very different, much more volatile-rich compositions. Jupiter and Saturn are mass-wise dominated by hydrogen (H_2) and helium (He) and are known as “gas giants”.

Table 2. Giant Planets

Object	Mass/ M_\oplus	Radius/ R_\oplus	ρ [kg m^{-3}]	a [AU]	e	i [deg]
Jupiter	316	11.21	1330	5.203	0.048	1.3
Saturn	95	9.45	700	9.537	0.054	2.5
Uranus	14.5	4.01	1300	19.191	0.047	0.8
Neptune	16.6	3.88	2300	30.068	0.009	1.8

The formation of the giant planets is imperfectly understood. Prevailing ideas suggest that, in the Solar system, the gas giant planets formed by a process of nucleated instability, a bit like a rain drop forming by condensation of water molecules on a refractory aerosol. The model was developed by Mizuno and others ([111], [122]). Briefly, solid bodies collide and grow by binary accretion in the protoplanetary disk, much as they did in the domain of the Terrestrial planets. Upon reaching a critical mass, generally estimated to be $\sim 10 M_\oplus$, the core precipitates the infall of surrounding nebular gas, producing a hydrodynamic flow that results in very rapid mass growth of the planet. As the planet mass undergoes a runaway growth, tidal torques exerted by the planet on the protoplanetary disk open a “gap” around the orbit of the planet. Subsequent mass in-flow to the planet continues at a reduced rate.

Growth by nucleated instability clearly involves two distinct timescales. First the core must grow to critical mass. Second, the nebular gas must be

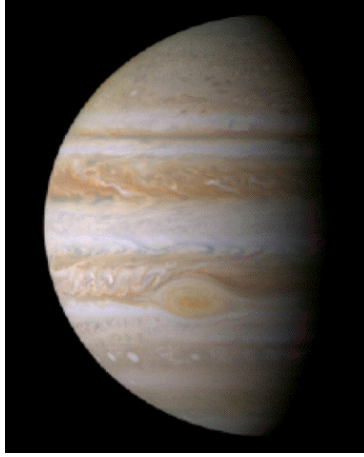


Fig. 4. Gas giant Jupiter from the Galileo spacecraft, showing its banded cloud structure and the Great Red Spot. Image from NASA.

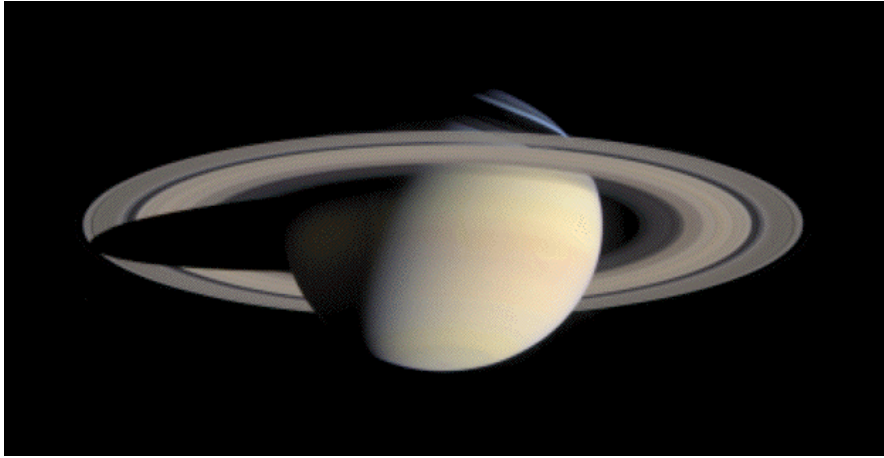


Fig. 5. Gas giant Saturn from the Cassini spacecraft. Courtesy NASA.

accreted by the core. Core growth, which occurs by binary accretion as for the terrestrial planets, is the slower process. It is the principal cause of concern with the nucleated instability model and so has been the subject of much attention. The key issue is that the core must grow on a timescale that is short compared to the timescale for the dissipation of the gas nebula. Observations of young stars with dust disks generally fail to reveal attendant gas, leading to the inference that the gas is quickly removed, probably on timescales of a few Myr for sun-like stars and almost certainly on timescales <10 Myr ([10]). This sets an upper limit to the core growth times and is a primary challenge to the core accretion model. One way in which core growth might have been

accelerated is through an increase in the disk column density just beyond the snow-line, owing to the extra mass in solids added by the freeze-out of nebular water vapor ([20]). Million year growth times at the orbit of Jupiter are not hard to obtain from current models, but more work is needed to induce Uranus and, especially, Neptune to grow on cosmically reasonable timescales.

A different giant planet growth scenario has been proposed in which the “slow step” of core accretion is side-stepped. In this model, the protoplanetary disk is supposed to have been intrinsically unstable to collapse under its own gravity. Disk instabilities clearly favor higher than MMSN disks (models typically assume disk masses ~ 10 times the MMSN in order to produce spontaneous collapse), but even MMSN models have been reported to be susceptible to collapse under some circumstances ([8]). Formation of giant planets by spontaneous collapse does not suffer the timescale problem of the nucleated instability model (because there is no need to wait for a nucleus to form) but there are other problems related to the long-term stability of the collapsing planet. Investigators differ on this issue. The differences are not fully understood, but might relate to the accuracy with which cooling processes are represented ([14]).

Neither core accretion nor nebula collapse predicted the over-abundance of heavy elements measured in Jupiter by the Galileo entry probe ([119], see Figure 6). In fact, pure collapse models implicitly contradict it because gravitational instabilities provide no way to selectively accrete elements according to their molecular weight. Pressure gradient forces might help to concentrate solids near growing planets ([56]) and one might conjecture that Jupiter’s heavy elements were accreted by the capture of ice-rich planetesimals in the extended atmosphere of the newly formed planet. There are problems with providing enough planetesimals to deliver the mass of Jupiter’s metal excess above Solar composition. This process further fails to explain N and Ar, which are over-abundant in Jupiter by factors of 3 or 4 (Figure 6) but which are too volatile to be carried by asteroids or the known comets in any appreciable abundance. The suggestion advanced by Owen et al. ([119]) is that Jupiter’s core grew by the accretion of ultra-cold (~ 30 K) planetesimals, in which N, Ar and other volatiles were efficiently trapped (probably by adsorption within amorphous water ice). But 30 K is too cold to fit the protoplanetary disk at 5 AU (c.f. Equation 2, which gives $T = 125$ K at this distance). A convincing resolution of this puzzle has yet to be identified.

Ice Giants

Compared to Jupiter and Saturn, Uranus (Figure 7) and Neptune (Figure 8) are an order of magnitude less massive and also compositionally distinct, being depleted in H_2 and He. The bulk of their mass is contained in heavier elements that form ices at low temperatures, such as C, N and O. Uranus and Neptune are known as “ice giants” for this reason. The difficulty in forming Uranus and Neptune on any reasonable timescale has motivated a number of novel,

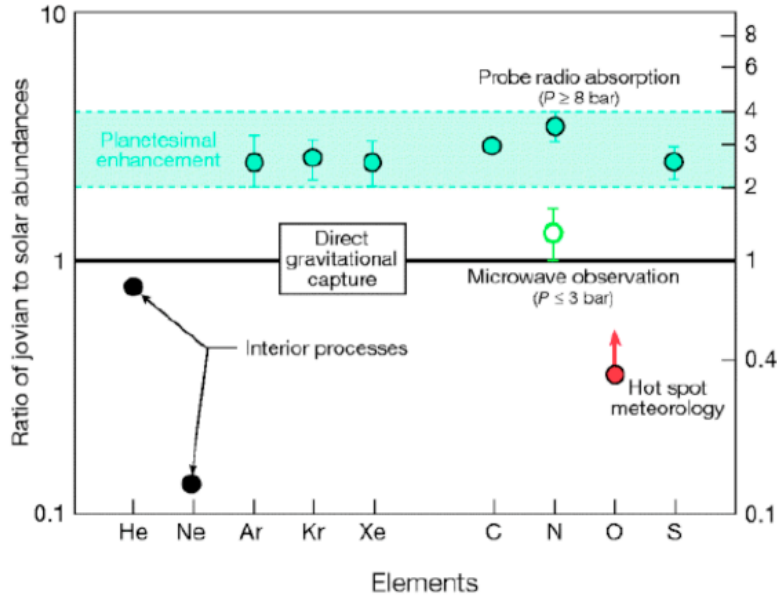


Fig. 6. Metal abundances in Jupiter relative to those in the Sun, as measured by the Galileo entry probe. Helium and Neon are low in abundance because they are partly dissolved in the metallic hydrogen core. Oxygen is low, probably because the probe entered Jupiter’s atmosphere at an (unrepresentative) hot-spot location, where conditions were atypically dry. The other measured elements are over-abundant relative to their Solar proportions. From [119].

alternative suggestions. For example, in one well-publicized model, Uranus and Neptune are envisioned to have formed between Jupiter and Saturn, were then scattered outwards by mutual perturbations and, finally, their orbits were circularized by friction with an assumed massive disk ([147]). To make all this happen, the authors placed the giant planets initially at 6.0, 7.4, 9.0 and 11.1 AU and assumed that they were initially each of $10 M_{\oplus}$, with an additional $95 M_{\oplus}$ of planetesimals between 12 AU and the assumed edge of the protoplanetary disk at 60 AU. In common with almost all other N-body Solar system simulations, they neglected collective interactions in the $95 M_{\oplus}$ disk (these might be expected to generate waves that could be important in the redistribution of angular momentum in the disk ([153])). Dynamical effects of the few $\times 10^4 M_{\oplus}$ of nebular gas (which must also have been present in order to keep the overall disk composition in approximately cosmic proportions) were also neglected, except that some of this gas was used to feed the runaway growth of the gas giants. The authors assert that their scenario for Uranus and Neptune formation is insensitive to the above assumptions and, indeed, it is easy to imagine that the first core to experience runaway mass growth should exert a strong gravitational influence on other cores nearby, perhaps

scattering them outwards. On the other hand, the initial conditions may have been very different from the ones envisioned in [147]. Worst of all, it is not clear to me what new observations can be taken to test it.



Fig. 7. Ice giant Uranus from the Voyager 2 spacecraft. Courtesy NASA.

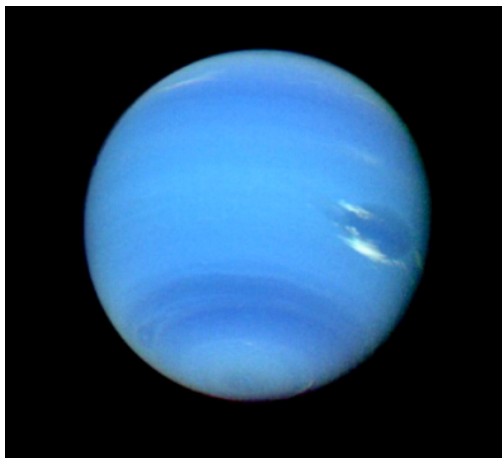


Fig. 8. Ice giant Neptune from the Voyager 2 spacecraft. Courtesy NASA.

An equally fascinating but rather different scenario for rapid ice giant formation assumes that these planets started out as gas giants and were then eroded down to their observed masses by intense fluxes of ionizing radiation from a nearby, massive star ([9]). According to this model, the future ice

giants are selectively depleted in mass relative to the surviving gas giants because they are more distant from the sun. Photoionized hydrogen (whose temperature is $\sim 10^4\text{K}$ and thermal velocity $\sim 10\text{ km s}^{-1}$) escapes more rapidly from heliocentric orbit at the distances of Uranus and Neptune than at Jupiter and Saturn, leaving the former two planets unprotected from the radiation while the latter two are heavily shielded. Again, the authors do not suggest observational tests of this model, although non-thermal loss of gases from planetary atmospheres often leads to isotopic fractionation effects that might be expected in this extreme case.

The Domain of the Comets

There are several useful definitions of what it is to be a comet, not all of them mutually consistent. The different definitions are used concurrently, sometimes without a clear understanding of the differences between them. The three different classification schemes are idealized in Figure 9.

Observationally, a comet is any object showing a gravitationally unbound atmosphere, known as a “coma” (from the Greek for “hair”). The coma is a low surface brightness region surrounding the central, mass-dominant nucleus. It owes its brightness to a combination of sunlight resonantly scattered from molecules and molecular fragments (radicals) and light scattered from tiny dust particles entrained in the outflowing gas. The visibility of the coma depends on the instrumental sensitivity and angular resolution. For this reason, objects which are discovered by survey telescopes as “asteroids” (i.e. bodies having no atmospheres) are commonly reclassified as comets based on the subsequent detection of comae by observers using more sensitive telescopes. Moreover, the strength of the coma diminishes rapidly with heliocentric distance, falling to invisibility beyond the orbit of Jupiter except in a few unusual cases. On longer timescales, cometary activity can evolve in response to evolutionary process on the surface, in a crust or “mantle” that throttles the release of escaping gas. What appears as a comet now might look completely asteroidal to observers of the 22nd century. Obviously, this observational definition of comet-hood is not at all a perfect one.

Compositionally, a comet may be defined as a small body in which a substantial part of the mass is contained in ice. Practically, we may expect all objects which condensed beyond the “snow-line” to contain bulk water ice. The snow-line is now near the orbit of Jupiter; all small bodies from the Jovian Trojans outward are likely to be compositional comets by this reasoning, whether or not they show comae. In the past, the snow-line may have been closer to the sun, meaning that ice could be present in many of the main-belt asteroids. These bodies are compositionally comets. Unfortunately, we have no meaningful way to estimate the bulk composition of a body without drilling into it, and this definition of comet-hood is consequently hard to apply.

Dynamically, a comet is any body with a Tisserand parameter measured with respect to Jupiter, $T_J \leq 3$ (the main-belt asteroids have $T_J > 3$). The

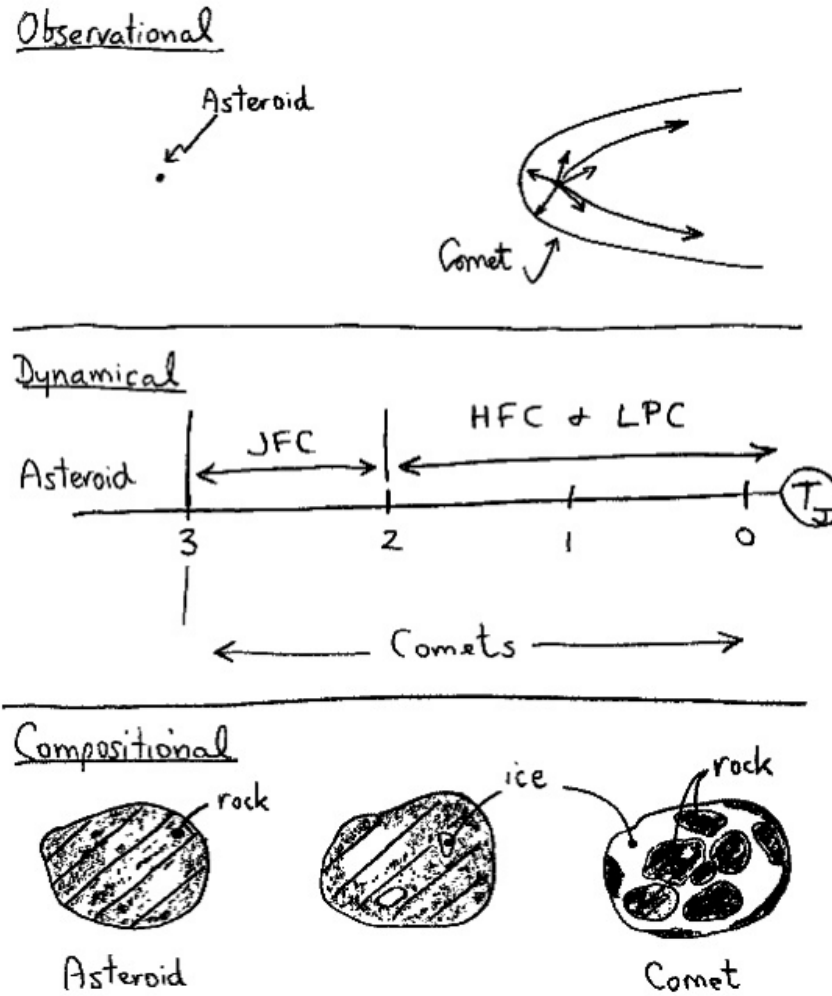


Fig. 9. Schematic diagram showing three different criteria for distinguishing comets from asteroids. Observationally, a comet is any body showing a coma (unbound atmosphere) at any point in its orbit. Dynamically, the distinction is made based on some model parameter, typically the Tisserand parameter, T_J . JFC, HFC and LPC denote Jupiter-Family Comets, Halley-Family Comets and Long-Period Comets. The Main-Belt Comets (MBCs) are located with the asteroids, in the middle panel of the figure. Compositionally, the distinction is based on the presence or absence of bulk ice in the body. The different definitions lead to the same classification in most cases, but there are growing numbers of bodies which are “cometary” by one definition but not the others.

Tisserand parameter is a constant of the motion in the restricted, circular three-body approximation, defined by

$$T_J = \frac{a_J}{a} + 2 \left[(1 - e^2) \frac{a}{a_J} \right]^{1/2} \cos(i) \quad (3)$$

where a_J is the semimajor axis of Jupiter's orbit (assumed circular), a , e and i are the semimajor axis, eccentricity and inclination of the small body orbit. Bodies with $T_J \leq 3$ strongly interact with the planet, indicating a short dynamical lifetime and a source elsewhere. Those with $T_J > 3$ are effectively decoupled from the planet. This definition, although seemingly clean-cut, also suffers from ambiguity. Some main-belt asteroids can be scattered onto orbits with $T_J \leq 3$. A few comets (the most famous is 2P/Encke) have $T_J > 3$ (although only slightly so), making them dynamically asteroidal.

The timescale for the loss of volatiles from a body is just $\tau_{dv} \sim M/(dM/dt)$, where M is the mass and dM/dt the rate of loss of mass. Whipple (1950) and authors since have assumed that mass loss is predominantly by sublimation, at a rate which can be calculated from the assumption of radiative equilibrium on the nucleus. There is growing evidence that the mass loss in at least some comets may be dominated by disintegration of the nucleus, in which mass is shed in macroscopic blocks or chunks rather than molecule-by-molecule as in the process of sublimation. Neglecting this possibility for the moment, we write the energy balance equation for a sublimating ice patch as

$$\frac{L_\odot}{4\pi R^2} (1 - A) \cos(\theta) = \epsilon \sigma T^4 + L(T) \frac{dm}{dt} + f_c + f_g. \quad (4)$$

Here, L_\odot is the luminosity of the Sun, R is the heliocentric distance, A and ϵ are the albedo and the emissivity of the surface, θ is the angle between the direction to the Sun and the surface normal, $L(T)$ is the latent heat of sublimation of the ice at temperature T , dm/dt is the mass loss rate per unit area and f_c represents the conducted energy flux from the surface while f_g is the flux of energy carried by gas flow into the nucleus. A few things should be noted. The quantity $L_\odot/(4\pi R^2)$ is the flux of sunlight falling on the projected surface. When evaluated at $R = 1$ AU this quantity is called the Solar Constant, F_\odot , and has the value $F_\odot = 1360 \text{ W m}^{-2}$. The first term on the right-hand side represents the power per unit area lost by radiation into space. The second term is the power per unit area consumed by sublimation. Physically this power is used to break the bonds connecting molecules together in the solid phase. The last term in the equation accounts for thermal conduction and can be either positive or negative, depending on the temperature gradient in the upper layers of the nucleus.

For a non-volatile ($L \rightarrow \infty$) black-body ($A = 0$, $\epsilon = 1$) material oriented perpendicular to the Sun ($\theta = 0$) and neglecting thermal conduction, the temperature is just

$$T = \left[\frac{F_{\odot}}{\sigma R_{AU}^2} \right]^{1/4} \sim \frac{393}{R_{AU}^{1/2}}. \quad (5)$$

This corresponds to the temperature at the sub-Solar point on a perfectly absorbing body. The average temperature on a spherical isothermal object will be reduced by a factor $4^{1/4}$, since the average value of $\cos(\theta)$ over the sunlit hemisphere is $1/4$, giving $T \sim 278/R_{AU}^{1/2}$.

For a sublimating surface, Equation (4) cannot be solved without prior knowledge of the temperature dependence of the latent heat. The Clausius-Clapeyron equation (for the slope of the solid-gas phase boundary in pressure vs. temperature space) can be used or, more directly, measurements of the thermal pressure exerted by sublimating water ice as a function of temperature can be employed. For illustrative purposes, we here consider an extreme approximation.

When close to the Sun (say for $R_{AU} < 1$ AU) water ice, the dominant cometary volatile, uses so much energy to sublimate that we may write

$$\frac{L_{\odot}}{4\pi R^2} (1 - A) \cos(\theta) \sim L(T) \frac{dm}{dt}. \quad (6)$$

as a rough approximation to Equation 4. Then we see that the characteristic mass loss rate per unit area (again with $\theta = 0$) is just

$$\frac{dm}{dt} \sim \frac{F_{\odot}}{L(T) R_{AU}^2} \quad (7)$$

and we have assumed for simplicity that the surface is perfectly absorbing, $A = 0$. Substituting $F_{\odot} = 1360 \text{ W m}^{-2}$ and $L(T) = 2 \times 10^6 \text{ J kg}^{-1}$ (for water ice), we have $dm/dt \sim 7 \times 10^{-4} / R_{AU}^2 [\text{kg s}^{-1} \text{ m}^{-2}]$.

The rate at which the sublimation surface recedes into the body of the nucleus is just

$$\frac{d\ell}{dt} = \rho^{-1} \frac{dm}{dt} \quad (8)$$

where ρ is the bulk density. With $\rho \sim 1000 \text{ kg m}^{-3}$, we estimate $d\ell/dt \sim 0.7 \mu\text{m s}^{-1}$ at $R_{AU} = 1$ AU. The sublimation lifetime of a nucleus of radius r_n is then

$$\tau_{dv} \sim \frac{r_n}{d\ell/dt} \sim \frac{\rho r_n}{dm/dt} \quad (9)$$

and, with the standard values as above, we obtain

$$\tau_{dv} \sim 50 \left(\frac{r_n}{1 \text{ km}} \right) [\text{yr}_1] \quad (10)$$

In this equation, the unit of time is denoted yr_1 to emphasize that it is the number of years of equivalent exposure to Sunlight at 1 AU.

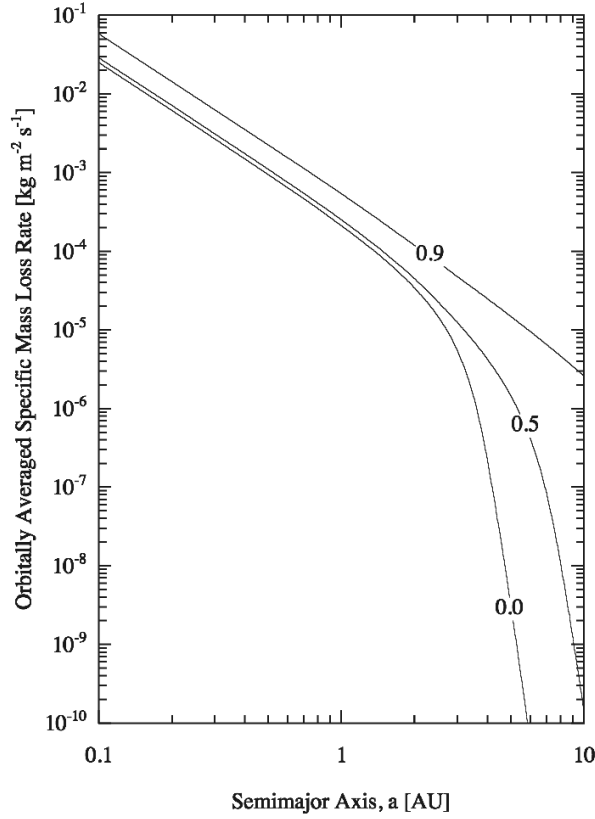


Fig. 10. Sublimation rate as a function of semimajor axis for comets having orbital eccentricities as marked. At a given semi-major axis, the sublimation rate averaged around the orbit increases with orbital eccentricity. This is because the enhanced sublimation near perihelion in an eccentric orbit more than compensates for the long period of inactivity surrounding aphelion. From [79].

Of course, no real comets circle the Sun in the orbit of the Earth. Instead they follow eccentric orbits with larger semimajor axes and are hot enough to sublime only when they dip in to perihelion. Still, the approximation described above nicely illustrates the fact that sublimation can potentially limit the active lifetimes of the comets to very small values, certainly values that are tiny compared to the 4.6 Gyr age of the Solar system.

Less approximate solutions of the energy balance equation are plotted in Figure 10. There I show the average value of dm/dt computed around the orbits of comets having eccentricities $e = 0, 0.5$ and 0.9 , as a function of

the semi-major axis. At a given semi-major axis, the net effect of non-zero eccentricity is to increase the orbitally-averaged mass loss rate relative to the circular orbit approximation, because sublimation grows fast enough near perihelion to overwhelm the long period of inactivity as the comet sails out to and back from aphelion. Figure 10 shows that, for a typical short-period comet having $a = 4$ AU and $e = 0.3$, the orbitally averaged mass loss rate is $dm/dt \sim 10^{-7}$ [kg s⁻¹ m⁻²], giving $d\ell/dt \sim 10^{-10}$ m s⁻¹ and $\tau_{dv} \sim 3 \times 10^5$ [yr].

All of the above is simplistic and intended merely to make a point, namely that sublimation can destroy nuclei quickly. We will have more to say about this later. For now, we use it to assert that the active comets must be derived from inactive source regions, if they are (as we believe) as old as the Solar system.

Source Regions

Three distinct source regions of the comets are now recognized. One, the Oort Cloud, was identified half a century ago ([118]) and is well known as the source of the long-period comets. The second, the Kuiper belt, was discovered in 1992 ([73]) and has played a major role in revamping our understanding of the Solar system. It is the source of the Jupiter Family Comets. The third, the Main-belt source, was discovered after the Saas Fee workshop ([63]), and is being written about here for the first time. Comets in this region are unique in being activated not by increased insolation resulting from inward dynamical evolution but by the transient exposure of near-surface ices, probably by collisions with other main-belt objects. Relations between the source regions and various small-body populations in the Solar system are summarized in Figure 11.

Oort Cloud Source

The Oort Cloud was identified from observations of long-period comets, whose orbits appear randomly (isotropically) distributed over the sky and whose semi-major axes are clustered at large values. The key observation made by Oort was that the orbital energies of many long-period comets (which Oort expressed by the inverse semimajor axes of their orbits) are smaller than the characteristic value of the energy change resulting from gravitational perturbations exerted by Jupiter in a single passage ([118]). He concluded that comets were falling into the planetary region from large (but finite) distances, and that many of the long-period comets had not been through the planetary region before, for otherwise they would already have been scattered out of the narrow (bound) energy peak in which they sit. This basic conclusion remains unchanged, to the undying credit of Mr. Oort. Likewise, available data, much improved in quantity and quality since Oort's time, continue to show that the Cloud is closely spherical in shape, albeit with a characteristic diameter

Tremaine ([156]) in a careful analysis some 50 yrs later). Like Oort ([118]) all researchers have assumed that the disagreement between the data and the model is best explained by fading or disintegration of the in-coming comets. However, the nature and reality of the fading remain unidentified. The low rate of detection of weakly active or completely inactive long-period comets has been interpreted as evidence that objects from the Oort cloud do not merely run out of gas but physically disintegrate ([94]). This conclusion rests on a poorly known relation between the brightness of active long-period comets and the sizes of their underlying nuclei. For example, if the nuclei are much smaller than assumed in [94] then they might escape detection without disintegrating.

The population and mass of the Oort Cloud are also uncertain. The population is derived from measurements of the rate of arrival of new comets from the Oort Cloud coupled with models of the rate of erosion of the cloud by external perturbers. Oort considered passing stars to be the main external perturbers. The asymmetric tide of the Milky Way is now thought to be a larger perturber ([58]). In addition, the rate of arrival of new comets is subject to observational biases that are difficult to quantify. Until recently, published population estimates relied on the work of visual observers ([43], [64]), most of whose survey techniques and other details went unpublished. A recent attempt to use data from the LINEAR survey (whose parameters are better, but still not completely, known) gives $\sim 5 \times 10^{11}$ comets with absolute magnitude $H \leq 11$ ([51]), about 10 times smaller than estimated previously.

Lastly, the relation of the Halley family comets to the Oort Cloud is unclear. These objects have distinctly non-random distributions of inclinations (with some retrograde members but many more prograde ones) and orbital periods, by definition, < 200 yr. The most likely source is the inner Oort cloud, but the location and population of this region remain poorly constrained.

Kuiper Belt Source

The Kuiper belt became real with the discovery of 1992 QB1 ([73]). Before that time, its only observed member was Pluto, misleadingly given planetary status for a host of mostly socio-scientific reasons. In fact, if Pluto had been accurately interpreted in 1930, our study of the structure of the Solar system could have advanced by many decades over the actual case (e.g. many bright KBOs have been “precovered” in photographic plates taken in the 1950s. They were not *discovered* using plates because astronomers did not think to look for them until after the discovery of 1992 QB1). Indeed, at least one astronomer correctly recognized in 1930 that Pluto must be just one of many trans-Neptunian objects ([92]) based on the dubiousness of the proposition that Tombaugh had been lucky enough to find the only one so soon after starting his survey. This reasoned position was drowned out by the assertion that Pluto must be the long-sought “Planet X”, predicted by Percival Lowell on the basis of a model of (what turned out to be unreal) deviations in the motion of Uranus. Still, everything is obvious in hindsight, and it is too easy

to see what should have been done knowing what we know, and too difficult to reconstruct the full state of confusion that reigned only a few decades ago. For example, Edgeworth in 1943 ([40]) speculated about “clusters” in the trans-Plutonian region (clusters were his idea for the structure of comets) while Kuiper (for whom the belt is somewhat ironically named) in 1951 ([85]) considered that this region should be empty, having been cleared of objects by strong perturbations from “massive” Pluto. Later, Fernandez in 1980 ([45]) reasoned that a flat disk source was needed to explain the inclination distribution of the short-period comets. Prior to this time, most researchers had been happy with the contention that short-period comets were somehow dynamically evolved versions of long-period comets. Later still, in 1988, Duncan and collaborators ([38]) showed, using numerical methods, the correctness of Fernandez’ argument.

The dynamics of the Kuiper Belt are extensively and masterfully discussed in the Saas Fee lectures by Alessandro Morbidelli.

Main Belt Source

Main belt comets (MBCs) have orbits in the main asteroid belt between Mars and Jupiter. At the time of writing, three MBCs have been identified ([63]; see Figure 12). The best known is asteroid 7968 also known as 133P/Elst-Pizarro, first observed to be accompanied by a dust trail in 1996. Initially interpreted as an impact-produced dust cloud ([148]), the reappearance of the trail near perihelion in 2003 showed that another explanation is required ([62]). The newest examples are comet P/2005 U1 (Read) and (118401) 1999 RE70, both of which show persistent dust emission over timescales of months. These three objects have similar semimajor axes located beyond 3 AU, in the outer regions of the main-belt. Their orbital inclinations are also all small, but the similar a and i are at least in part results of observational bias, since our surveys have targeted exactly these types of object. The MBC orbits are decoupled from both Mars and Jupiter, and appear to be dynamically stable on billion year timescales, like those of the main-belt asteroids which occupy exactly the same region of orbital element space (Figure 13).

Could the MBCs be comets captured from other regions, for example from the Jupiter family comet (JFC) or long-period comet (LPC) populations? As observers we are open to this possibility, but dynamical simulations of the motions of comets suggest that this is very unlikely. In fact, pure dynamical calculations completely fail to inject comets into MBC-like orbits even when the perturbations of the Terrestrial planets are included ([47]; [95]). Some work has been done on the effects of non-gravitational accelerations (caused by asymmetric mass loss from cometary nuclei) but again, MBC-like objects are not produced. Failing some dramatic revision of the dynamics, we are forced to the conclusion that the MBCs are what they appear to be: asteroids that outgas like comets.

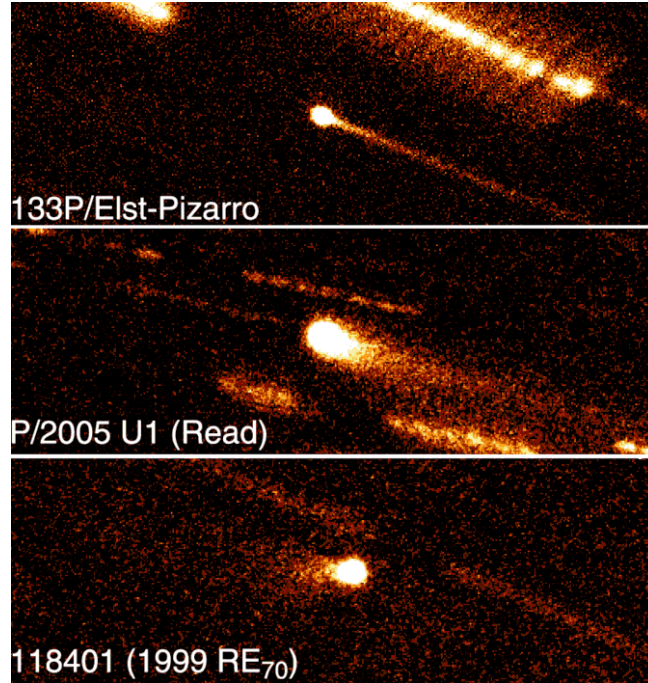


Fig. 12. Three main-belt comets (MBCs) in deep CCD images from Mauna Kea. These objects emit dust like comets but have orbits which are like those of outer main-belt asteroids. Background stars and galaxies appear trailed owing to the non-sidereal motions of the MBCs. From [63].

Several lines of argument indicate that the mass loss from MBCs is driven by sublimation, probably of near-surface water ice. First, mass loss from 133P has been observed at consecutive perihelia but not in between. This is exactly as expected for sublimation-driven activity. The sunward “nose” of the coma of P/2005 U1 (Read) is well resolved, with an apex scale of several arcseconds. This implies that the particles are ejected from the nucleus at considerable speed ($>100 \text{ m s}^{-1}$), as expected for water ice sublimating at $\sim 3 \text{ AU}$. Other explanations for mass loss seem less viable. The nucleus of 133P is rapidly rotating and it is possible that centripetal effects assist the launching of dust particles from its surface. However, centripetal effects alone cannot explain the observation that activity is confined to perihelion. Neither do we find evidence for rapid rotation in P/2005 U1 (Read) or 1999 RE70: these objects spin so slowly that rotation can play no role in the mass loss. Electrostatic levitation of dust grains has been observed in the terminator regions of the moon, where the derived velocities of the dust grains are $\sim 1 \text{ m s}^{-1}$. Such low speeds are incompatible with the extended coma of P/2005 U1 (Read) and, furthermore, it is hard to see how electrostatic ejection of grains could be episodic (as on

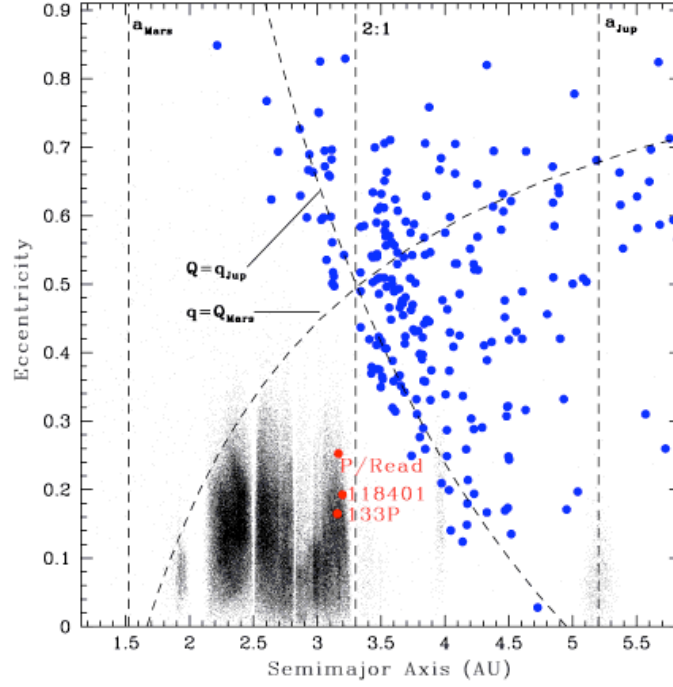


Fig. 13. Semimajor axis vs. orbital eccentricity for asteroids (small black dots), Jupiter family comets (blue circles) and the known MBCs (red circles). Vertical dashed lines mark the semimajor axes of the orbits of Mars and Jupiter and the 2:1 mean-motion resonance with Jupiter, which practically defines the outer edge of the main belt. Curved dashed lines show the locus of orbits which are just Mars and Jupiter crossing. Objects below these two curves cross neither Mars nor Jupiter, like essentially all of the main-belt asteroids. The MBCs fall within the domain occupied by stable main-belt asteroids and far from the periodic comets. From [63].

133P), or why it would be confined to only three of several hundred asteroids examined in detail by our on-going survey.

For these reasons, it appears that the MBCs are really comets in a special population where the source reservoir and the current locations are one and the same. Unlike the long and short-period comets, the MBCs are not activated by being brought from cold storage locations into the hot inner Solar system. Instead, we suspect that they are activated collisionally. For example, the mass loss from P/2005 U1 (Read) corresponds to sublimation from an exposed patch of dirty water ice having a diameter of only ~ 20 meters. Such a patch could be exposed by the impact of a meter-scale boulder into the nucleus surface. The mass loss rate at 3 AU is about $10^{-5} \text{ kg s}^{-1} \text{ m}^{-2}$

(Figure 10) and, with density $\rho \sim 1000 \text{ kg m}^{-3}$, the surface recession rate is $d\ell/dt \sim 10^{-8} \text{ m s}^{-1}$. An ice patch 20 meters in diameter would sublime to a depth equal to its diameter on timescale $\tau \sim 2 \times 10^9$ second (50 yr), thereafter declining into inactivity from self-shadowing. Triggering collisions involving the impact of 1-meter scale boulders should not be overly rare: we expect to find many MBCs in planned all-sky surveys such as Pan STARRS.

Is ice in the asteroid belt surprising? It should not be. Some meteorites show textural and geochemical evidence that they have been aqueously altered, probably by being bathed in liquid water at temperatures not far above the triple point ([84]). This evidence includes the presence of clay minerals and serpentines that most naturally form with water, as well as carbonates and mineral deposits in veins that cross-cut other structures in the meteorites (showing that the vein materials were emplaced after formation). Spectrally, about half of the outer belt asteroids show absorption features attributed to water of hydration in minerals (not free ice, but water bound chemically within other materials such as clays; [16]). At both smaller and larger distances, the prevalence of these hydration features decreases. One interpretation that fits the available data is that, at smaller distances the asteroids were too hot for liquid water to have survived while at larger distances the ice was so cold as to never be melted, foreclosing the possibility of hydration reactions that could produce water of hydration bands ([82]).

The greatest excitement behind the MBCs lies in the potential relation between these objects and the oceans (and, through water, life). Earth probably formed too hot to trap much water and so it is widely believed that a separate source is required. Possible sources include the comets (but the measured deuterium/hydrogen (D/H) ratios in the three that have been measured seems higher than in the oceans [109]) or watery asteroids like the MBCs ([112]). MBCs are so close to Earth that we should soon be able to visit them with a mass spectrometer, to measure their D/H (and other isotope ratios, including $^{16}\text{O}/^{17}\text{O}/^{18}\text{O}$) abundances, and so to make a direct comparison with the oceans.

3 Cometary Nuclei

The nucleus is the fundamental component of any comet because it contains most of the mass. Unfortunately, it also the hardest to study, because most of the cross-section is carried by dust and gas ejected from the nucleus and not by the nucleus itself. As a result, physical studies of comets have, until recent times, been biased towards the study of gas and dust released from the nucleus by its sublimation. These are the subjects of Heike Rauer's lectures in this Saas Fee workshop. A great deal of important information about comets has been gleaned, for example, from the study of molecular fragments from dissociated parent molecules. In this section, though, I want to focus on what we have learned about the nucleus itself.

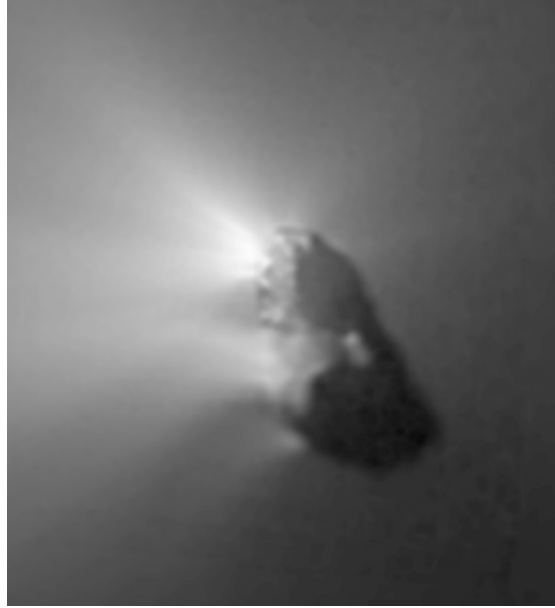


Fig. 14. Nucleus of 1P/Halley imaged from the ESA Giotto spacecraft. This classic image was the first to show a nucleus at high spatial resolution. While various surface features can be discerned, it is obvious that important structure lurks beneath the resolution of the data. Dust jets are seen to emanate primarily from the sun-facing side of the nucleus. Courtesy Giotto camera PI H. U. Keller and ESA.

The first well-established detections of nuclei were achieved from the ground in 1984, quickly followed by close-up images of the nucleus of 1P/Halley in 1986 (Figure 14). Before that time, direct observation of the nucleus was held by many to be impossible because of contamination of the nuclear signal by scattering from nearby dust and gas. A common misperception is that the nucleus is invisible from the ground because it is shielded from view by near-nucleus dust. This is almost never the case for a very simple reason: dust is ejected from the nucleus by the drag forces exerted on it by sublimated ice. If the coma were to become optically thick, the source of heat driving the sublimation would be shut down, reducing the dust opacity. Feedback, then, stabilizes the optical depth along a line of sight from the nucleus to the Sun, to be smaller than unity. Transient exceptions to this feedback control can be imagined, and might occur, but few or none of the observed properties of comets require large broadband optical depths in order to be understood. Cometary comae are, to a good level of approximation, optically thin. Since 1986, the nuclei of comets Borrelly (Figure 15), Wild 2 (Figure 16) and Tempel 1 (Figure 17) have been imaged by spacecraft.



Fig. 15. Nucleus of P/Borrelly imaged from NASA’s Deep Space 1 spacecraft. The effective radius is ~ 2.2 km and surface albedo ~ 0.03 . Note the lobed structure of the nucleus (perhaps caused by a composite structure consisting of two major bodies in contact) and the smooth “pond” material above the waist. Courtesy NASA.

Nucleus Size

Cometary nuclei subtend minuscule angular diameters (milliarcseconds) and are unresolved in optical ground-based data. No occultation of a field star by a nucleus has ever been observed: most nucleus sizes must be inferred by indirect means. The size of the cometary nucleus can be inferred from the “classical” technique first used by Dave Allen [3] in which simultaneous optical (scattered) and infrared (thermally emitted) flux densities are compared. This method is so important to the study of small bodies that it is worth describing in more detail: in essence it is very simple. Photons from the Sun strike a body and are either reflected or absorbed. The fraction reflected is called the “Bond albedo”, A . Photons not reflected are absorbed, raising the temperature of the body and producing thermally emitted photons at longer wavelength. The fraction of the incident photons that is absorbed is $(1 - A)$. The optical flux density scattered from a body is proportional to the product $C_e p$, where C_e is the cross-section, while the thermally emitted flux density is proportional to $C_e(1 - p)$. Here, p is the “geometric albedo”, which is related to the Bond albedo by $A = pq$, where q is a measure of the angular dependence of the scattering function called the “phase function”. Provided q is known, measurements at optical and thermal wavelengths permit us to solve for the two unknowns C_e and



Fig. 16. Nucleus of P/Wild 2 imaged from the NASA Stardust spacecraft. The effective radius is ~ 2.1 km and surface albedo ~ 0.03 . Note the remarkably smooth shape of the nucleus, which resembles that of a rotational figure of equilibrium. Courtesy Don Brownlee and NASA.

p (c.f. Figure 18). The measurements should be simultaneous because small bodies are usually not spherical, causing C_e to vary with time.

Examined closely, the Allen size method is more complicated. The scattered radiation is both anisotropic and wavelength dependent, introducing two extra parameters. The phase function q is not in general known and has only been measured for a few bodies that can be observed over a very wide range of phase angles. Real surface materials will have (wavelength dependent) thermal emissivities < 1 , introducing another parameter. Most seriously of all, heat absorbed on the day-side of a rotating body can be carried by rotation onto the night-side, meaning that the emitted flux density depends on the heat-retaining capacity of the surface layers (traditionally characterized by the “thermal inertia parameter” $I = k\rho c_p$, where k is the thermal conductivity, ρ is the bulk density and c_p is the specific heat capacity, or by the “thermal diffusivity” defined by $\kappa = k/\rho c_p$). The magnitude of this interaction between the rotation and the thermal emission introduces more parameters, for the thermal constants of the surface, and for the magnitude and orientation of the rotation vector relative to the line of sight. What looked like a conceptually simple method is in fact horribly complicated: the number of unknown parameters in the model generally exceeds the number of observational constraints.

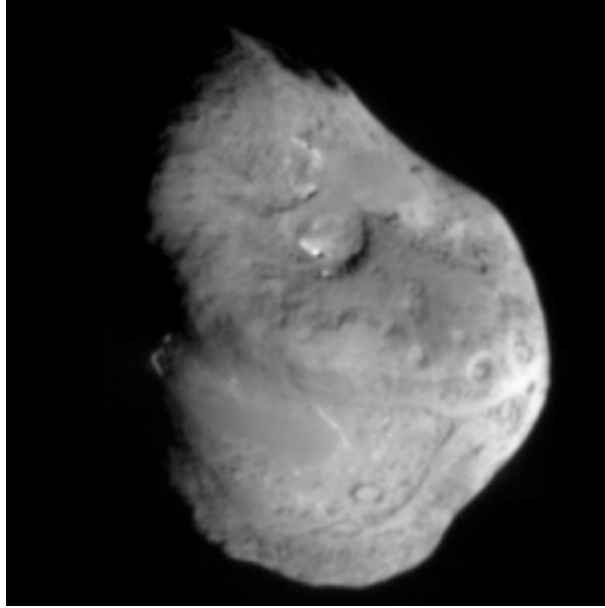


Fig. 17. Nucleus of P/Tempel 1 imaged from the NASA Deep Impact spacecraft. The effective radius is ~ 3.1 km and surface albedo ~ 0.05 . Note the craters, the left-right gash across the nucleus and the two regions of smooth terrain apparently occupying lowland positions. Courtesy Mike A'Hearn and NASA.

What saves the Allen method is the empirical finding that assumed values for a great many of the unknown parameters can nevertheless give object cross-sections and albedos of useful accuracy. The “Standard Thermal Model” (STM) has arisen as a way to bundle the many assumptions in such a way that they are not too visible to the user and so not too frightening! In STM, the thermal emission is assumed to emanate from a spherical body in which the surface temperature is set by instantaneous equilibrium with sunlight and where the effects of rotation are unimportant. This could mean that the surface heat retention is very small, so that heat is lost before rotation carries it away from the day-side, or it could mean that the rotation vector points exactly at the Sun, so that rotation does not change the surface heating pattern. Even with these and other assumptions for the emissivity (generally ~ 0.9) and the angular dependence of the scattering, STM must include a fudge factor called η , the “beaming parameter”, that is supposed to represent the angular dependence of the emission from the surface caused by surface roughness and topographic effects. The value of η in STM is often taken to be $\eta = 0.756$ ([89]) but in fact it is very uncertain and recent work suggests that $\eta = 1$ may apply. For our purposes, the point is that the interpretation of thermal emission data in terms of object size and albedo depends on poorly specified parameters such as η .

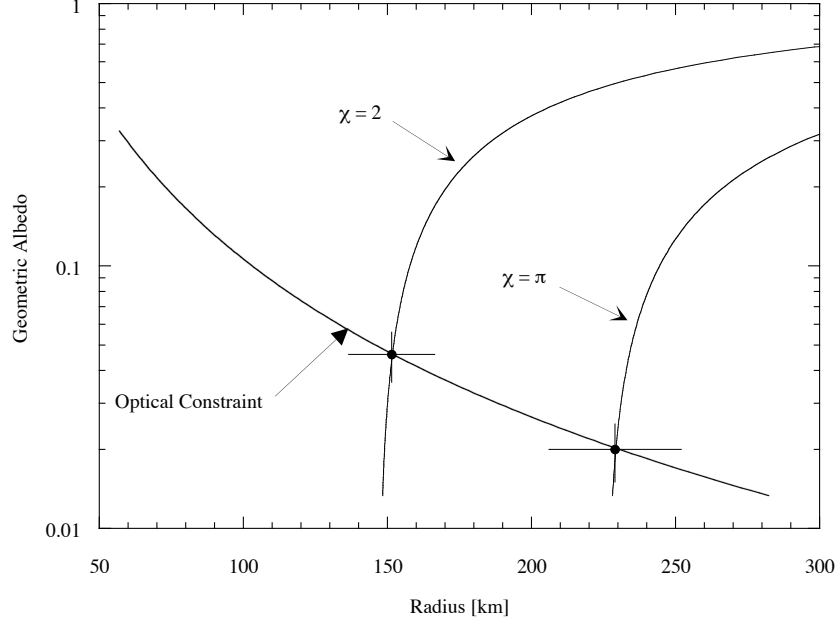


Fig. 18. Example of the thermal-optical method of determining the size and albedo of an object. The optical data place a constraint on the product $p_R r^2$, where p_R is the geometric albedo and r is the effective radius. The thermal data place, through a model of the surface temperature distribution, a constraint on $(1 - p_R)r^2$. The two curves labeled $\chi = 2$ and $\chi = \pi$ refer to the STM and ILM surface temperature approximations. The dots mark plausible solutions for these two models. Both yield low geometric albedos for this object. From [74].

A counterpart to the STM is the “Isothermal-Latitude Model” (ILM) which is best thought of as applying to a spherical body with the Sun in its equator and a rotation period so short that the temperature is independent of azimuth and a function only of latitude. The ILM model has lower mean surface temperatures than the STM and so requires a larger C_e (and smaller p) to generate a given thermal emission signal.

The strength of the Allen method is that it is widely applicable and seems mostly to give diameters accurate to $\sim 5\%$ or 10% when appropriately “tuned” by the selection of the uncertain parameters. It has been used to measure the cross-sections and albedos of about a dozen comets, as listed in Table 3.

The size distribution of the cometary nuclei has been measured by different groups with different investigators reaching different conclusions. Usually the size distribution is represented as a power law with index q

$$n(r)dr = \Gamma r^{-q}dr \quad (11)$$

Table 3. Well-Measured Cometary Nuclei

Object	r_e ^(a)	p ^(b)	P ^(c)	b/a ^(d)
1P/Halley	5.5	0.04±0.02	52.8,177.6	2.0
2P/Encke	2.4	0.05±0.02	11?	2.6
9P/Tempel 1	3.1	0.05±0.02	41.0	1.4
10P/Tempel 2	5.3	0.022±0.005	9.0	1.7
19P/Borrelly	2.2	0.03	25.0	2.5
22P/Kopff	1.7	0.042±0.006	12.3	1.7
28P/Neujmin 1	10.7	0.03±0.01	12.75	1.5
49P/Arend-Rigaux	4.2	0.04±0.04	13.47	1.6
81P/Wild 2	2.1	0.03±0.01	12?	1.7
107P/Wilson-Harrington	1.7	0.05±0.01	6.1	1.2
C/1995 O1 (Hale-Bopp)	37	0.04±0.03	11.34	2.6
C/2001 OG108 (LONEOS)	8.9	0.030±0.005	57.19	1.3

a: Effective radius [km]. *b*: Visual albedo. *c*: Rotation period [hr]. *d*: Axis ratio.

where Γ is a normalization constant. Reported values are $q = 3.6^{+0.3}_{-0.2}$ ([46]), $q = 2.6 \pm 0.03$ ([155]), $q = 2.6 \pm 0.3$ to 2.9 ± 0.3 ([88]), $q = 2.45 \pm 0.05$ for the radius range 1 to 10 km and $q = 1.91 \pm 0.06$ for radius between 2 to 5 km ([108]) and $q = 3.7 \pm 0.3$ also for radius between 2 to 5 km ([142]).

What do these values mean and why are they so different? The large scatter amongst the measurements has been attributed by Tancredi et al. ([142]) as the result of poor sample definition and, in some cases, the use of inaccurate nuclear magnitudes. When only JFC nuclei are considered, they obtain an index $q \sim 3.7$ regardless of which data-set is used. While many of the measurements were taken apparently in the desire to make a comparison with the size distribution of Kuiper belt objects, this is a difficult comparison to make. First of all the well-measured Kuiper belt objects (for which the size distribution index is $q = 4.0^{+0.6}_{-0.5}$ ([149])) are one to two orders of magnitude larger than the measured cometary nuclei. There is no reason why a single power law should hold from the largest KBOs down to km-sized cometary nuclei, particularly if it is true that the smaller objects are collisional products while the larger KBOs are survivors from a primordial population ([44]). Indeed, initial observations of fainter, smaller KBOs show that the size distribution flattens below ~ 100 km diameter, with an index $q \sim 2.6$ in this region ([7]).

More seriously, the measured cometary nuclei are a highly unrepresentative sample of cometary nuclei as a whole. The measured nuclei tend to be those of comets in which mass loss has almost certainly changed the nucleus shape (see below) and size. In fact, if the sublimation lifetime increases with the nucleus size (see Equation 8 or 9), then small nuclei should be destroyed faster than large ones, leading to a net flattening of the size distribution relative to the distribution in the initial (pre-heated) population. It is hard to see how the nucleus size distribution can tell us much of fundamental value about the

comets so long as our measurements are confined to the relatively evolved nuclei of comets with perihelia in the terrestrial planet domain.

Nucleus Colors

Accurate determinations of the colors are available for a small number of cometary nuclei (Table 4). This value is consistent with the mean colors of various other inner- and middle-Solar system small-body populations, including the Jovian Trojans, the nuclei of dead JFCs and the Damocloids (likely nuclei of dead Halley-family comets, see Figure 19). However, the optical colors of comets are not consistent with those of KBOs or Centaurs, in the sense that the ultrared matter ($S' > 25\%/1000\text{\AA}$) found on many of these objects is completely absent on the nuclei (Table 4). We will return to this observation in our discussion of the effects of a surface mantle.

Table 4. Mean Optical Reflectivity Gradients [From [79]]

Object	S'_{min}	S'_{max}	S'_{med}	$S' \pm \sigma$	Number
Damocloids	5	17	13	11.9 ± 1.0	12
Active JFC Nuclei	-5	22	11	11.6 ± 2.3	11
Inactive JFC Nuclei	-6	18	6	7.2 ± 2.0	12
Trojans	3	25	9	9.6 ± 0.9	32
Centaurs	0	43	19	20.3 ± 2.8	22
KBOs	-10	48	21	21.1 ± 1.4	83

Nucleus Shape and Rotation

The shapes and rotational states of cometary nuclei (and asteroids) can be determined from their rotational lightcurves (temporal variations in the scattered light). The main measurable parameters are the lightcurve period and the range. Two things are immediately worth mentioning. First, the relation between the lightcurve period and the underlying rotational period may not be obvious, a-priori. If the lightcurve is caused by albedo spots then the two are likely to be equal. If the lightcurve is caused by variations in the projected cross-section owing to aspherical shape, then the lightcurve period is likely to be half the rotational period. Where sufficient data exist to discriminate between these possibilities, the lightcurves are almost always found to be caused by aspherical shape more than by albedo spots. Second, the range of the lightcurve is routinely but inaccurately described in the literature as the amplitude (formally the amplitude is half the range). The measured range sets only a lower limit to the nucleus axis ratio, since in general the rotational axis will not be aligned perpendicular to the line of sight. Repeated measurements under different geometries are needed to remove these effects of projection.

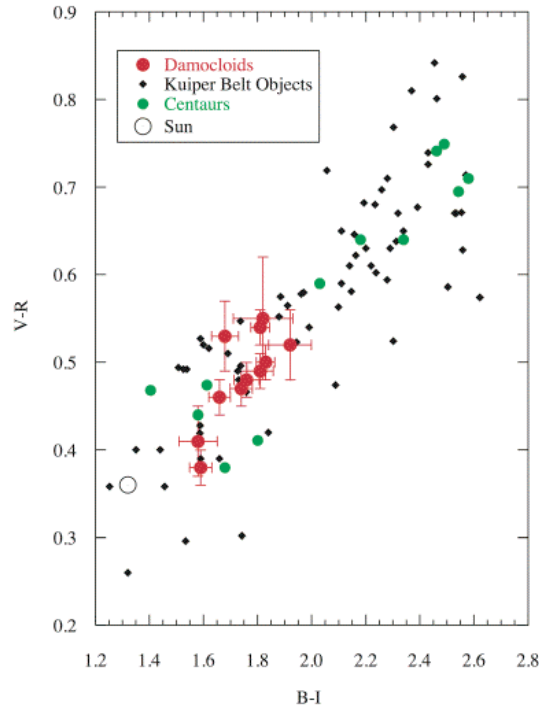


Fig. 19. Color-color plane showing the Damocloids, KBOs and Centaurs. While the KBOs and Centaurs show a wide range of surface colors (and, presumably, compositions) the Damocloid surfaces are entirely lacking in ultrared matter (spectral gradient $>25\%/1000\text{\AA}$, corresponding to the upper right in this color-color diagram). From [79].

Figure 20 shows a range vs. period plot for those comets thought to be well-measured. Added to the plot are curves computed for two models. First, I show curves for prolate bodies in rotation about a minor axis, computed under the assumption that gravity at the tip of the spheroid exactly equals the centripetal acceleration there ([68]). Secondly, I assume that the nuclei are figures of rotational equilibrium and plot curves taken from Chandrasekhar's (in)famous book ([17]) in which the shapes of strengthless bodies are computed as a function of their density and angular momentum. The Figure shows that the nuclei do not need to be very dense (in general, the critical densities are $<1000 \text{ kg m}^{-3}$) in order to be stable against centripetal effects, regardless of which model is used. It is not known whether the nuclei behave at all like strengthless bodies, but the consensus view (influenced very strongly by the split comets, see [6]) is that this is likely to be a good approximation.

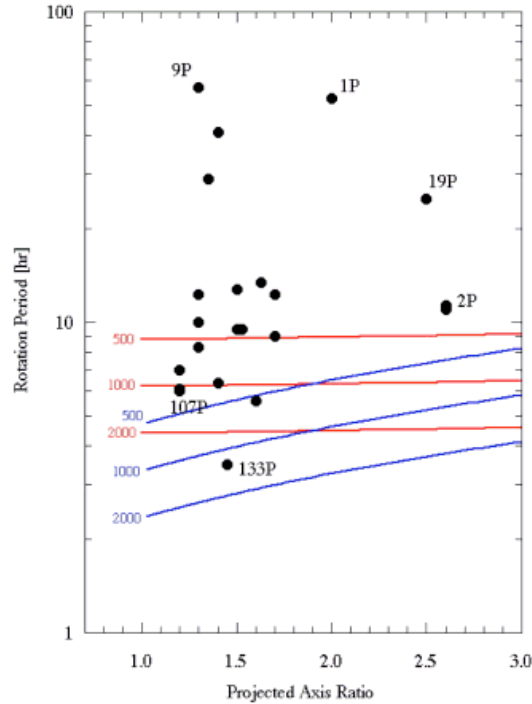


Fig. 20. Rotation period vs. axis ratio (derived from lightcurve range) for cometary nuclei. Prolate spheroid curves (blue) were computed as described in the text. The equilibrium spheroids (red) were computed by Pedro Lacerda. The densities of the models are given in the figure and can be interpreted as limits to the nucleus densities under the assumption that the nuclei are strengthless.

Nucleus Density

There are no good measurements of the densities of cometary nuclei, but there are many strong opinions held by planetary scientists about what those densities are! Perhaps because of preconceived ideas about the way in which comets formed, most planetary scientists believe that the nuclei are less dense than water. This might be true, but we do not know.

Several indirect methods have been invoked to measure the densities of the cometary nuclei.

- The range vs. period plot was first used to argue that the densities must be low ([68]). A prolate ellipsoid nucleus model was used to estimate the density from the period and the lightcurve range. There is no particular reason to assume that the nuclei are well described by prolate ellipsoids and, as can be seen from Figure 20, an alternate assumption gives sub-

stantially lower densities for a given period, range pair.

- D/Shoemaker-Levy 9 was disrupted while passing close to Jupiter (Figure 21). Measurements of the spreading rate of the “string of pearls” comet after disruption, when interpreted as the product of tidal stresses acting on an aggregate body of negligible tensile strength, give a relatively robust estimate of the density $\rho = 600 \text{ kg m}^{-3}$ ([6]).

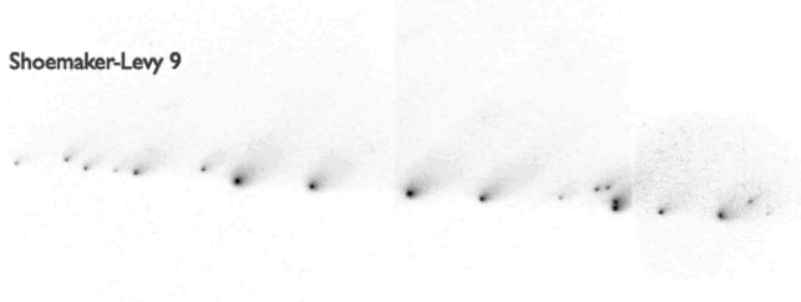


Fig. 21. Multiple components of the nucleus of D/Shoemaker-Levy 9 imaged from the Hubble Space Telescope by Weaver et al. Each component sports a stubby tail, created by radiation pressure sweeping of emitted dust. Photometry shows that the emission was largely impulsive and occurred at the moment of break-up of the nucleus as it passed Jupiter (minimum distance 93,500 km or about 1.31 R_J).

- Asymmetrical outgassing exerts a “rocket” acceleration on the nucleus of magnitude

$$\alpha_n = f_r \frac{V}{M_n} \frac{dM_n}{dt} \quad (12)$$

where M_n is the nucleus mass, V is the bulk speed of the material launched from the nucleus by sublimation and f_r is a dimensionless constant. The value of f_r depends on the angular distribution of the momentum flux in material launched from the nucleus. For a nucleus which ejects matter in a perfectly collimated beam $f_r = 1$ while for isotropic ejection $f_r = 0$. Consider a 1 km radius comet (mass $\sim 4 \times 10^{12}$ kg) ejecting mass at 10^3 kg s^{-1} in a collimated beam ($f_r = 1$) while at 1 AU from the Sun. The rocket acceleration is $\alpha_n \sim 3 \times 10^{-7} \text{ m s}^{-2}$, or about 10^{-5} times the Solar gravity at this distance. Although small, the long action time allows the rocket acceleration to produce measurable deviations from Keplerian motion. To use Equation (12) to determine nucleus density, the acceleration α_n must first be measured from astrometry of the comet. Spectroscopy gives V from the Doppler shift of lines resonantly scattered from escaping gas and dM_n/dt can be estimated from the strengths of molecular emission lines.

Then, given a value of f_r , this equation gives the nucleus mass. Coupled with an estimate of the nucleus volume, the density can be determined. This method has been used to estimate the densities of 81P/Wild 2 ($\rho < 600$ to 800 kg m^{-3} ; [30]), 67P/Churyumov-Gerasimenko ($\rho < 600 \text{ kg m}^{-3}$; [29]), 19P/Borrelly ($100 < \rho < 300 \text{ kg m}^{-3}$; [28]). The low densities are interesting and in accord with the value obtained for D/Shoemaker-Levy 9 by a different method but, given the large amount of modeling needed to estimate f_r , I suspect that this method can give almost any density the user wants.

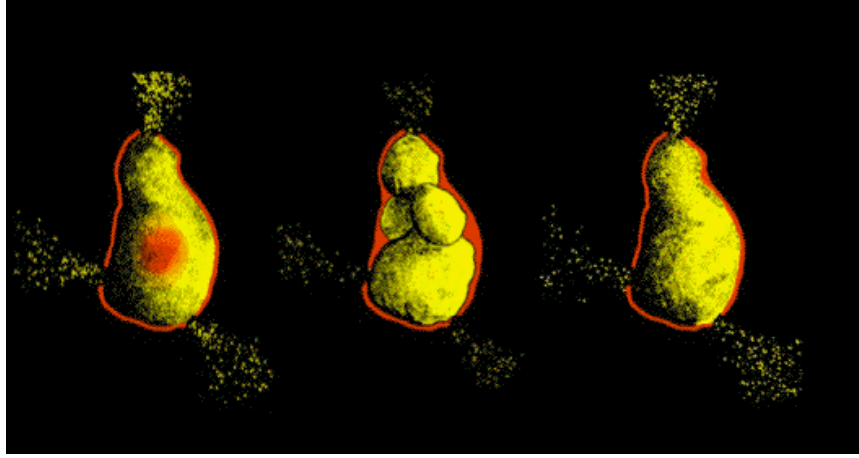


Fig. 22. Schematic of possible internal structure of the cometary nucleus. On the left, a differentiated nucleus in which the material properties (strength, composition) vary radially as a result of past heating, concentrated at the core. This model seems unlikely, given the high volatile contents and low tensile strengths of comets. However, some of the larger nuclei could have experienced non-negligible internal heating from radioactive decays (enough to mobilize interior volatiles). In the middle, a multi-component (sometimes called “rubble pile”) nucleus in which sub-elements in the body are loosely bound by gravity. This is probably closest to the real structure inside cometary nuclei. On the right, a monolithic nucleus with structural integrity over its whole diameter. Very small comets (like asteroids of $<100 \text{ m}$ scale), could be like this. The red skin on each object symbolizes the non-volatile mantle.

Still, accepting for the moment that the densities are $<1000 \text{ kg m}^{-3}$ and that the strengths are small, it is interesting to speculate about the possible internal structures of the nuclei. Most probably the nuclei are porous dirt-ice mixtures with a broken internal structure consisting of blocks each much smaller than the aggregate size (middle panel of Figure 22).

3.1 Mantles

Observations show that the surfaces of cometary nuclei are largely non-volatile, consisting of refractory matter generally described as a “mantle” (crust might be a better word, and certainly less confusing given the stratigraphic relationship between the Earth’s mantle and crust). Evidence for the existence of mantles includes

- Images from the ground and from space show that the mass loss from comets occurs from only a fraction of the total surface, suggesting that surface volatiles are not widely distributed (note: this says nothing about the distribution of volatiles inside the nucleus). Specifically, the mass loss occurs in jets and the total rate of production of water is less than would be expected if the whole nucleus were covered in water ice. The derived fractional “active areas” range from $\sim 0.01\%$ to $\sim 10\%$ ([1]).
- Spectral maps of comet 9P/Tempel 1 obtained from the NASA Deep Impact spacecraft show evidence for water only in a few locations occupying about 0.5% of the total surface ([138]).
- Temperatures of some nuclei are higher than can be sustained by a sublimating ice surface. Examples include 1P/Halley (peak temperature > 360 K [41]), C/1996 B2 (Hyakutake) (320 K; [98]) and 9P/Tempel 1 (peak temperatures ~ 330 K [138]).

The physical properties of the mantles remain poorly determined. This is a more serious problem for cometary science than it at first sounds, because almost everything we know about the comets is either controlled or at least strongly modulated by the mantles. Likewise, the physics behind mantle formation and destruction is not well known.

Figure 23 compares the colors of objects within each of several small-body populations. Color is parametrized by the normalized reflectivity gradient, S' [%/1000Å], essentially the slope of the spectrum of the object after division by the spectrum of the Sun. Several features in Figure 23 deserve comment.

a) The nuclei of comets, both dead and alive, show a spread in color that matches that observed in the Trojans but which is distinct from the KBO color distribution. A few *blue* nuclei are known. We will argue below (Section 3.1) that these are most likely surfaces covered by rubble mantles.

b) The Trojans (which are often but incorrectly described as consisting of very red D-type asteroids) in fact show a wide range of surface colors, down to neutral ($S' = 0$), and they are much less red than the majority of KBOs.

c) Very red material is found only on the surfaces of the KBOs and the Centaurs. Specifically, if we define ultrared matter as having $S' \geq 25\%/1000\text{\AA}$ ([71]), then the Figure shows that ultrared matter is absent in the inner Solar system populations including the Jovian Trojans, the nuclei of active and inactive Jupiter family comets and the Damocloids (not shown here, but see Figure 19; [79]). Since the progression of objects from top to

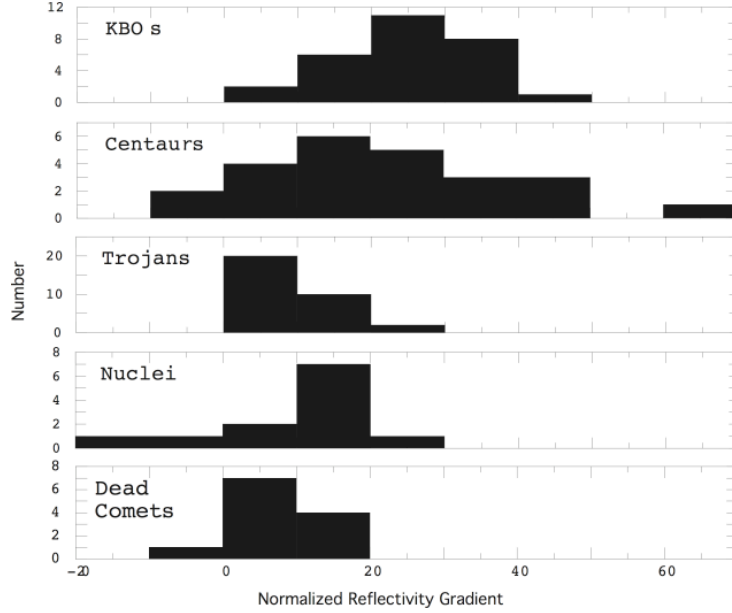


Fig. 23. Histogram showing the normalized reflectivity gradients measured in various small-body populations. Negative (positive) spectral gradients indicate blue (red) reflection spectra, relative to the Sun, which by definition has a spectral reflectivity gradient of zero. Material with $S' \geq 25 \text{ \%}/1000\text{\AA}$ is defined as ultrared matter. Figure from [79].

bottom in Figure 23 represents (except for the Trojans) a dynamical progression from the Kuiper belt source inwards, a plausible conclusion is that the ultrared matter cannot survive in the inner Solar system. One guess is that the ultrared objects are coated in organic matter that has been irradiated by long-term exposure to cosmic rays and other particles, creating an “irradiation mantle” (Section 3.1).

Rubble Mantles

A rubble mantle consists of refractory, particulate debris that is left behind on the surface of the nucleus by the sublimating gases. Particles bigger than a certain critical size, a_c , are too heavy to be launched against the gravitational attraction to the nucleus and remain behind. Assuming a spherical nucleus of radius r_n and density ρ_n , the surface gravitational force is just

$$g_n = \frac{16}{9} \pi^2 G \rho_n \rho_d r_n a^3 \quad (13)$$

where $G = 6.6 \times 10^{-11} \text{ N kg}^{-2} \text{ m}^2$ is the gravitational constant and ρ_d and a are the density and radius of the dust grain. The gas drag force is a complicated

function of the grain parameters (shape, roughness) and of the ratio of the grain size, a , compared to the mean free path in the gas, λ_{mfp} . In the case where the grain size, $a \ll \lambda_{mfp}$, it is reasonable to consider the momentum of impacting gas molecules as being added one at a time, giving the classical drag force expression

$$F_d = C_d \pi a^2 \mu m_H N_1 \Delta V^2 \quad (14)$$

in which C_d is the (dimensionless) drag coefficient, μ is the molecular weight of the sublimating gas ($\mu = 18$ for water), $m_H = 1.67 \times 10^{-27}$ kg is the mass of the hydrogen atom, N_1 [m⁻³] is the concentration of the gas at the nucleus surface and ΔV is the velocity of the gas relative to the grain. We calculate N_1 from the thermal equilibrium equation for sublimating ice. The velocity difference ΔV is roughly the bulk speed of the gas as it leaves the nucleus, which data, physics and models show is of order, V_s , the sound speed in the gas at the temperature of the sublimating surface. Balancing gravitational force on a spherical grain with the gas drag then gives

$$a_c = \frac{\mu m_H N_1 V_s^2}{G \rho_n \rho_d r_n} \quad (15)$$

for the critical size above which a grain cannot be accelerated to the escape speed from the nucleus and so which must fall back to the surface. We have ignored numerical constants in this expression and, given our state of ignorance, set $C_d = 1$. Noting that

$$\frac{dm}{dt} = \mu m_H N_1 V_s \quad (16)$$

we can rewrite this expression as

$$a_c = \left[\frac{V_s}{G \rho_n \rho_d r_n} \right] \frac{dm}{dt} \quad (17)$$

where dm/dt is obtained by solution of Equation (3). Substitution gives us an immediate estimate of a_c . Consider a water ice nucleus 1 AU from the Sun and with radius $r_n = 5$ km. The sublimation rate per unit area is $dm/dt \sim 10^{-4}$ kg m⁻² s⁻¹ (Figure 10). If we take $V_s \sim 500 R_{AU}^{-1/2}$ [m s⁻¹] as a first-order approximation to the gas speed at heliocentric distance R_{AU} [AU] and further take $\rho_n = \rho_d = 1000$ kg m⁻³, we obtain $a_c \sim 0.1$ m. Decimeter-sized bodies can be launched by gas drag against the gravitational attraction to the nucleus. This critical size decreases dramatically with increasing heliocentric distance owing to the rapid decline in the specific sublimation rate as the nucleus temperature drops. Beyond $R_{AU} \sim 5$ or 6 AU, we find $a_c < 0.1 \mu\text{m}$, and the particles that can escape the gravity of the nucleus are those that are too small to efficiently scatter optical photons (with wavelengths $\lambda \sim 0.5 \mu\text{m}$), rendering them unobservable. The magnitude of a_c is plotted in Figure 24 as a function of nucleus size and heliocentric distance.

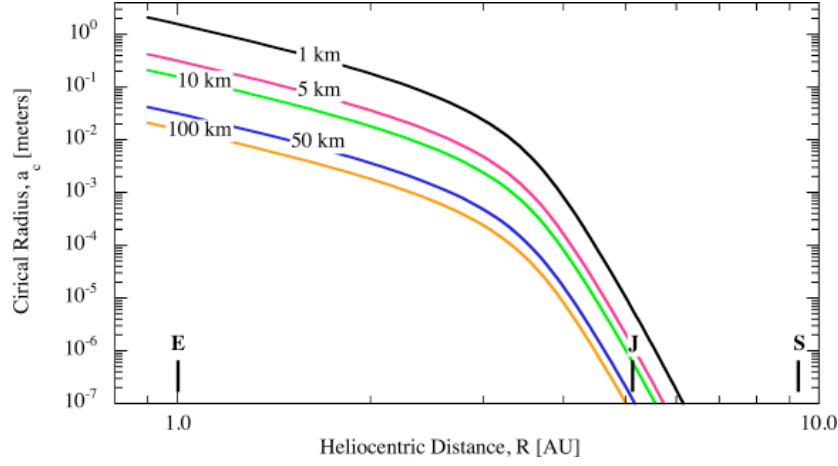


Fig. 24. Solution to Equation 17 computed for dark (albedo 0.04) sublimating water ice nuclei as a function of heliocentric distance for nucleus radii from 1 km to 100 km (as marked). The semimajor axes of the orbits of Earth, Jupiter and Saturn are marked for reference. Particles larger than the wavelength of visible light, $\lambda \sim 0.5 \mu\text{m}$, can be ejected all the way out to Jupiter’s orbit but not much beyond.

There are many weaknesses in this simple calculation and many papers have been written to refine it since Whipple’s (1950) classic exposition. Still, the essential point is that very large particles, *if they exist in the nucleus*, cannot be easily launched by gas drag into interplanetary space, and will remain on the surface where they will impede the heating of surface ice and so diminish the sublimation gas flux. This is the rubble mantle (Figure 25).

It is interesting to consider some consequences of this simple model. First, how thick must such a mantle be? The physical condition for the mantle to seriously impede the heating of ice is that the mantle thickness must rival or exceed the diurnal thermal skin depth. The latter is a measure of the depth to which heat can be carried from the surface by conduction, and is given by $L_D \sim (\kappa P_{rot})^{1/2}$, where κ is the thermal diffusivity and P_{rot} is the rotation period of the nucleus. With $\kappa = 10^{-7} \text{ m}^2 \text{ s}^{-1}$ (appropriate for the porous dielectric materials likely to comprise the mantle matter) and $P_{rot} = 10 \text{ hr}$ (typical of the well-observed cometary nuclei; see Table 3) the skin depth is only $L_D \sim 0.06 \text{ m}$ (6 cm!) and the mantle need not be very thick in order to impede the gas production.

The timescale for such a mantle to form is

$$\tau_M \sim \frac{\rho_n L_D}{f_M dm/dt} \quad (18)$$

where f_M is the fraction of the solid mass that cannot be ejected by gas drag because it is contained in bodies with $a > a_c$. For a power-law distribution in which the number of solid particles with sizes in the range a to $a + da$ is

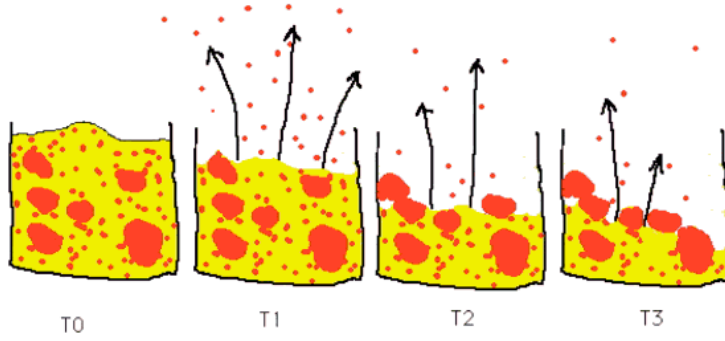


Fig. 25. Schematic cross-section in a cometary nucleus showing the formation of a rubble mantle. At initial time T0, the nucleus consists of a mix of “rocks” (red) and ices (yellow). The nucleus is heated from above by sunlight, leading to the sublimation of the ices. Gas drag forces expel smaller rocks into the coma while larger solid particles are left behind. Movement of the sublimation surface into the nucleus exposes more rocks, including large ones that eventually clog the surface, creating a thermally insulating, non-volatile rubble mantle. Any mantle thicker than the diurnal skin depth (~ 5 cm) can inhibit sublimation. The interval from T0 to T3 is a function of nucleus size and the pattern of insolation on the nucleus, but can be shorter than the orbit period for comets in the inner Solar system.

given by $n(a)da = \Gamma a^{-q}da$ (Γ and q are constants), the fraction f_M is easily calculated from

$$f_M = \frac{\int_{a_c}^{a_+} a^{3-q} da}{\int_{a_-}^{a_+} a^{3-q} da} \quad (19)$$

where a_- and a_+ are the minimum and maximum sizes in the dust size distribution. This integral takes a particularly simple form when $q = 4$, and this happens to be not too different from the size distribution measured in the coma of 1P/Halley by the dust detectors of the Giotto spacecraft, at least for sizes near $100 \mu\text{m}$ ([87]). Then

$$f_M = \frac{\ln(a_+/a_c)}{\ln(a_+/a_-)} \quad (20)$$

provided $a_+ \geq a_c$, and $f_M = 0$ otherwise. The size of the largest “particle” in the cometary nucleus is unknown, but studies of bolides show that comets eject bodies of decimeter and larger sizes when near the sun. We take $a_+ = 0.1$ m and, based on observations of tiny dust particles in 1P/Halley, set $a_- = 10^{-8}$ m.

Combining Equations (16 - 19) and using Equation (3) to calculate dm/dt , we obtain an estimate of the mantling time, τ_M , and the results are plotted in Figure 26. Two volatiles have been used to estimate the timescales, water

and carbon monoxide, the main difference being that the latent heats of sublimation of these materials are in the ratio of about 10:1. I further show curves computed for two values of the nucleus radius (at constant assumed density 1000 kg m^{-3}) to indicate the effect of size. Several features of Figure 26 are worthy of note.

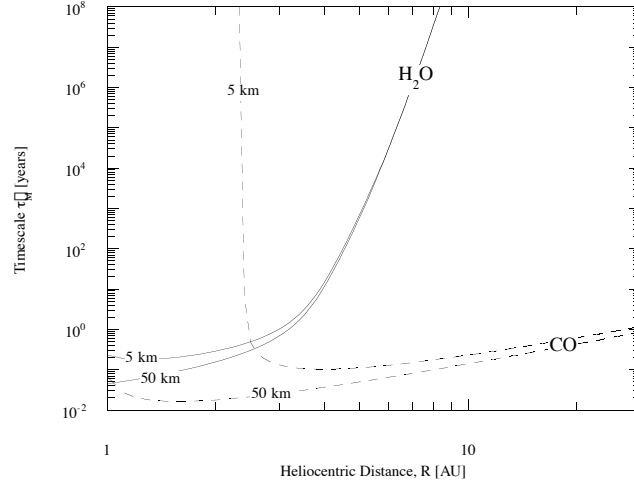


Fig. 26. Timescale for mantle formation from a simple model (Equation 18) as described in the text. Curves are shown for two volatiles (CO and H_2O) and two nucleus radii (5 km and 50 km), with assumed density of 1000 kg m^{-3} . From [71].

- The mantling timescales for water are less than 1 yr for heliocentric distances $\leq 3 \text{ AU}$, for nuclei of both 5 km and 50 km radius. This very short timescale means that rubble mantles can potentially grow within a single orbit. A patch of ice exposed to the Solar insolation would, in this model, seal itself against continued sublimation on a timescale of a year. If true, we should think of the mantle as a dynamic structure that can adapt to changes in the insolation.
- Mantling of the water nuclei slows with increasing heliocentric distance. At distances $R_{\text{AU}} \geq 6$, the mantling time exceeds the $\sim 0.5 \text{ Myr}$ dynamical lifetime of the Jupiter family comets ([96]). Rubble mantles should not form at larger distances if formed only by the sublimation of water ice.
- Cometary activity powered by CO sublimation extends to much lower temperatures and larger heliocentric distances than for water. Indeed, CO is so volatile that it sublimates strongly across the entire planetary region

of the Solar system. The mantling time due to CO is therefore very short even out to the orbits of the KBOs. One conclusion is that CO should not be found on the surfaces of the KBOs (unless held there by gravity on the largest objects). Another is that the past presence of CO in the Kuiper Belt would have led to rapid and complete encrustation of these bodies by rubble mantles.

- Figure 26 shows that the mantling times rise at the smallest heliocentric distances. This is most obvious for the CO, 5 km radius model, which rises towards infinity at about 2.5 AU. The physical reason for this is that when sublimation is very strong, the gas drag forces are able to eject even the largest solid bodies in the distribution (i.e., $a_c > a_+$), and no mantles can form.

This simple model illustrates many of the key features of the rubble mantle. It needs to be only centimeters thick in order to protect nucleus ice from the heat of the Sun. It can form very quickly. A mantle formed at large heliocentric distance can be unstable to ejection at smaller distances. Mantles on large nuclei are more stable than on small nuclei. Depending on the size distributions in the refractory particles, very tiny nuclei might be unable to retain rubble mantles at all. Considerations like these have induced some researchers to consider models which couple mantle development with orbital evolution, particularly with the drop in perihelion distance which has occurred to most observed comets. The results are very interesting, and parallel to the qualitative ones presented here ([128]).

The given picture of rubble mantle development is highly simplistic, however. For example, the role of centripetal acceleration has been ignored. An elongated nucleus in rotation about its short axis will experience net reduction in gravity towards the tips that could render the rubble mantle unstable, producing bald spots. The existence of even a small tensile strength would overwhelm the significance of the tiny nuclear gravity, and could give the mantle properties quite different from those inferred above. Lastly, and most importantly, what we have presented is no more than a hideous cartoon compared to the complex surface structures imaged by spacecraft on the nuclei of comets (Figures 14, 15, 16 and 17). Making a deeper connection between the properties of the mantle and the surface morphology will require mechanical and other data from a surface lander. Perhaps ESA's Rosetta will do the job?

Irradiation Mantles

An entirely different type of mantle has long been postulated for the surfaces of cometary nuclei. This mantle is formed by the long-term bombardment of ices on the nucleus surface by energetic particles from the sun, the Solar wind and galactic cosmic rays and is generally known as the “irradiation mantle” (see Figure 27). Ironically, there is no specific evidence for irradiation mantles on the nuclei of comets. Instead, if they exist anywhere, they are most likely

to be found on the exposed surfaces of the Kuiper belt objects. The reason for this is simple: rubble mantle formation timescales are *much* shorter than the timescales for radiation damage, given the known fluxes of energetic particles.

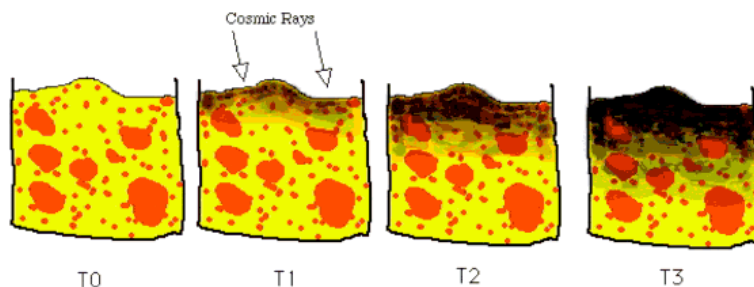


Fig. 27. Schematic cross-section in a cometary nucleus showing the formation of an irradiation mantle. At initial time T0, the nucleus consists of a mix of “rocks” (red) and ices (yellow). Cosmic rays bombard the surface layers, breaking bonds in the ice molecules, allowing the formation of radicals, the preferential escape of hydrogen and the formation of a carbon-rich, low albedo “irradiation mantle”. The thickness of the layer is of order 1-meter (for bulk density 1000 kg m^{-3}). The interval from T0 to T3 is uncertain, but probably ~ 100 Myr for complete processing. A thinner surface layer (affected only by low energy particles) could form on a shorter timescale.

Energetic particles dissipate their energy in a complicated cascade of interactions that results in breaking the covalent bonds that hold common molecules together. New bonds can form, producing molecules that were not present in the initial mix. Hydrogen liberated from parent molecules in this way is small enough and sufficiently volatile to be able to escape, leaving behind C, N and O to form complex molecules with whatever hydrogen remains. Experiments show that the result is a chemically complex mixture of organics, both aliphatic (carbon chain molecules) and aromatic (carbon ring molecules), in some cases polymerized to a very high molecular weight ($\mu > 100$'s). High molecular weight corresponds to low volatility and the resulting irradiation mantle is stable against sublimation relative to the common ices. The mantle is also of low albedo, a reflection (pun intended) of the high carbon content. In fact, the molecular and chemical nature of this type of material is poorly defined. Related complex organic materials called “Tholins” are sometimes used as analogs, but these are produced by spark discharge in low pressure gases and they may not be an appropriate analog for the mantle material. “Kerogens”, high molecular weight hydrocarbons found in terrestrial oil shales, may be a good analog, although these are not produced by irradiation.

The depth to which material can be damaged by energetic particles is a function of the particle energy. In the planetary region, the largest fluxes are for low energy particles in the Solar wind (energy ~ 1 to 10 keV) and these

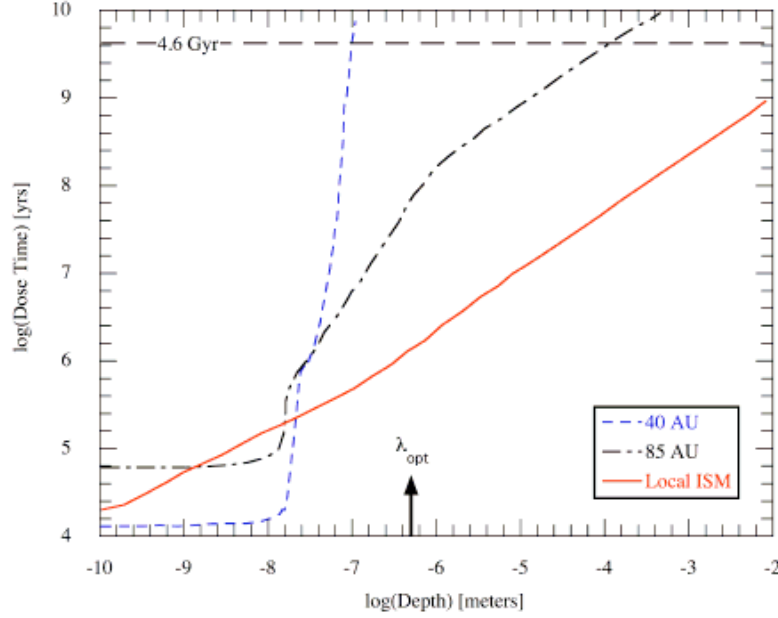


Fig. 28. Timescale for delivering 100 eV per $\mu = 16$ atom as a function of depth in water ice (density 1000 kg m^{-3}) at three heliocentric distances. The dashed horizontal line at the top marks the age of the Solar system. The wavelength of visible photons is marked by λ_{opt} at the bottom. Damaged layers thicker than λ_{opt} are likely to have significant effect on the reflected light spectrum. Replotted from [23].

particles have very small penetration depths in ice. Much more energetic particles (MeV to GeV and beyond) are found in the cosmic rays but at relatively low fluxes. Damage occurs fastest at the surface but, given billions of years should extend to column densities $\sim 1000 \text{ kg m}^{-2}$ (1 meter in ice of density 1000 kg m^{-3}). Calculations of the timescale for delivery of 100 eV per oxygen atom are shown in Figure 28, for heliocentric distances of 40 AU, 85 AU and “ ∞ AU” (corresponding to the local interstellar medium). This energy dose is chosen because it corresponds to heavy damage to the exposed material. Major differences exist between these locations both because the flux of low energy particles from the Solar wind declines with the inverse square of the distance and because the magnetic interaction of the wind with the interstellar medium results in a gradient in the flux of energetic particles. The Figure shows that the Kuiper belt objects at ~ 40 AU in fact exist in a relatively benign radiation environment. Solar wind particles quickly irradiate a surface skin $\sim 100 \text{ \AA}$ thick (in 10^4 yr) but damage to $0.1 \text{ }\mu\text{m}$ takes a considerable fraction of the age of the Solar system. At 85 AU, close to the recently detected

termination shock (where the Solar wind decelerates as it impacts the heliopause from the inside) the flux of energetic particles is increased and total damage occurs to depths of $\sim 10^{-4}$ m on billion-year timescales. In the open interstellar medium, the damage can reach depths in ice ~ 1 m on the same timescale.

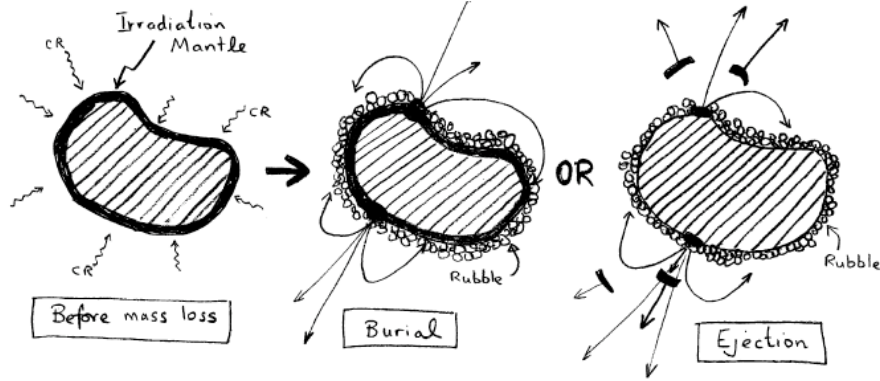


Fig. 29. Possible styles for the destruction of irradiation mantle. On the left, billions of years of exposure to energetic particles on a frigid surface has created an irradiation mantle (black) on a nucleus that is otherwise pristine (shaded). At the onset of sublimation-driven mass-loss, the irradiation mantle could be buried (middle) or cracked and ejected by gas drag (right), the exposed surface of the nucleus being replaced by a rubble mantle consisting of excavated, unirradiated matter in both cases.

What does all this mean? First of all, the timescales for irradiation damage (Figure 28) are vastly longer than those for the production of a rubble mantle (Figure 26). I conclude that irradiation mantles should not be found on any object whose past life has allowed the possibility of mass loss and, so, of rubble mantle formation. Objects in the outer Solar system are too cold to sublimate water, and so remain as candidates for irradiation mantling. Perhaps the ultrared matter ($S' \geq 25\%/1000\text{\AA}$) that appears to be a unique feature of the KBOs and of some Centaurs, is irradiated mantle material. Consistent with this inference is the observation that ultrared matter does not survive approach to the sun within the orbit of Jupiter ([71],[79]), corresponding to the heliocentric distance inside which water begins to sublimate and the timescale for rubble mantle formation becomes short (Figure 26). The mode of destruction of the irradiation mantle is not clear, however. The mantle could still be present but buried beneath a recently deposited rubble mantle consisting of (less red) debris excavated from beneath the ~ 1 m thick irradiated layer. Or it could be ejected by gas drag at the onset of strong sublimation inside ~ 5 AU (Figure 29).

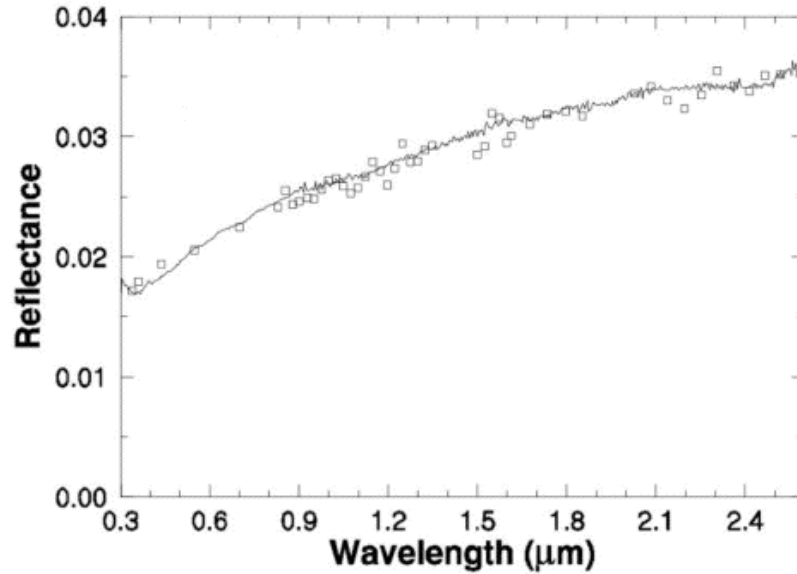


Fig. 30. Comparison of the optical-near infrared reflection spectrum of D-type asteroid (368) Haidea (points) with the Tagish Lake meteorite (line), showing a nearly perfect match. Figure from [60].

4 Kuiper Belt

Several of the important properties of the Kuiper Belt, established over the past 14 years by painstaking observational work around the world, have been summarized in the section of this book by Alessandro Morbidelli. I will avoid duplication and instead focus on aspects of the Kuiper Belt that are less thoroughly covered elsewhere in this volume.

4.1 Kuiper Belt Physical Properties: Colors and Albedos

Ideally we would use spectra to determine the surface compositions of KBOs and other Solar system bodies. The faintness of most such objects makes this ideal unreachable and, instead, broadband colors are often used as a proxy for the spectra and so for surface composition. Problems with this approach are numerous. Colors cannot, in general, be used to determine compositions. Colors are influenced by composition, but also by wavelength-dependent scattering effects in particulate regoliths, and by viewing geometry. On the other hand, colors can be used to classify objects into groups. The Holy Grail of colorimetric work on the KBOs has been for some years to find correlations between the colors and other properties such as size and orbital character

([102]; [143]; [144], [75]; [57]; [106]; [145]). Correlations like this might provide illuminating clues about the KBOs and their histories.

The use of color to learn about KBOs has been, to say the least, an uphill battle. The first property to be measured was color diversity; the KBOs exhibit a range of surface optical colors that is large compared to the uncertainties of measurement. In fact, color diversity has emerged as the only physical property to be confirmed by every subsequent study. Later, color diversity at optical wavelengths was found to extend into the near infrared ([74]; [27]; [106]; [31]; [33]). Moreover, the optical and infrared colors are correlated, which indicates that a single coloring agent is responsible for the wavelength dependence of the reflectivity across the wavelength range from B-band ($0.45 \mu\text{m}$) to J-band ($1.2 \mu\text{m}$) and perhaps beyond to K-band ($2.2 \mu\text{m}$).

The physical significance of color diversity is unclear. One possibility is that the different colors reflect intrinsically different compositions. This might be the case, but it is difficult to understand why the compositions of the measured KBOs would be so varied. After all, the measured objects are located in a comparatively narrow band between about 30 AU and 50 AU, where the radiation equilibrium temperatures vary from ~ 40 K to ~ 50 K. This very small temperature range could scarcely effect the compositions of the KBOs enough to cause major color differences.

For this reason, a second model was proposed to explain the color dispersion. In this “resurfacing model”, the hemispherically averaged color of a KBO is time-dependent, and determined by a competition between collisional resurfacing and cosmic ray processing. For example, suppose that cosmic ray processing causes an exposed surface to become redder on timescale τ_{cr} . This process competes with impact-driven resurfacing, in which impacts excavate “fresh” material from beneath the irradiated layer. If the excavated matter has a different (neutral?) color, the instantaneous, hemispheric average color will vary stochastically between extremes set by fully radiation-processed matter and fresh, excavated material. Substantial color fluctuations are possible when the timescale for resurfacing, τ_{coll} is $\sim \tau_{cr}$.

Attractive though it at first seems, several predictions of the resurfacing model have not been confirmed by observations. The model predicts that rotational color variations on KBOs should be nearly as large as the color differences which exist between KBOs of a given size. This is not observed. The model also predicts that the range of colors observed should vary with KBO size, since the timescale for collisional resurfacing varies with object size while τ_{cr} does not. Again, this violates the observations. The model has been extended by the addition of color variations owing to possible outgassing effects ([31]) but the problems remain. Collisional resurfacing is unlikely to be responsible for the color dispersion of the KBOs, although it could conceivably be a contributing factor.

Tegler and Romanishin reported that the colors of KBOs were not just dispersed over a wide range, but were *bimodally* distributed ([143]). They continued to find bimodal color distributions with larger samples ([144], [145])

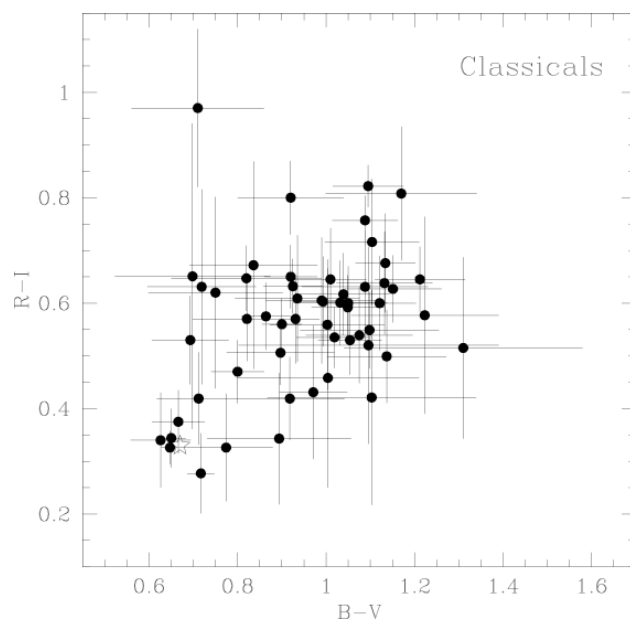


Fig. 31. Color-color diagram for classical KBOs. From [32].

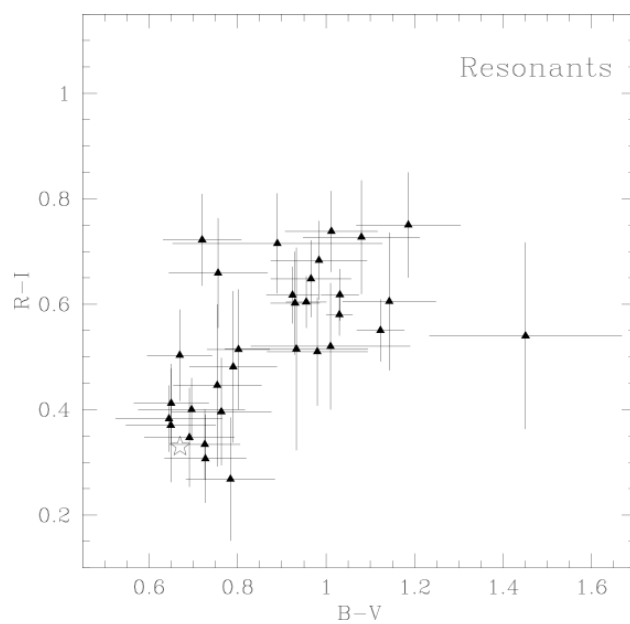


Fig. 32. Color-color diagram for resonant KBOs. From [32].

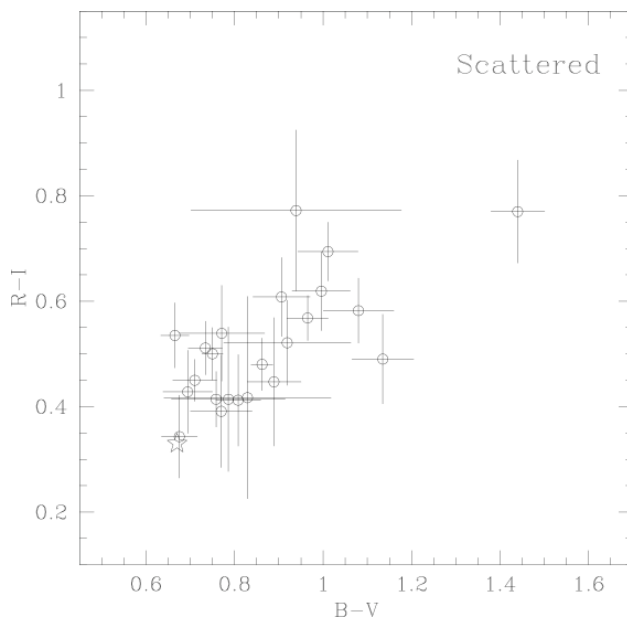


Fig. 33. Color-color diagram for scattered KBOs. From [32].

but failed to receive observational support for this finding from independent observers ([75], [57]; [31]; [36]). The colors of the KBOs available at the time of writing (2006 March) are distinctly unimodal (see Figures 31, 32 and 33). Recently, Peixinho et al. ([120]) reported that, while the KBO colors are indeed unimodally distributed, the Centaurs appear bimodal (see the next Section). This is more than an academic distinction: a bimodal color distribution would have placed strong constraints on the nature of the KBOs, had it been real.

Few of the long-sought correlations between colors and other physical and dynamical properties have turned out to be observationally robust. The correlation that seems most likely to be real is between color and perihelion distance ([144]) or, equivalently, between color and inclination ([150]) amongst the classical KBOs. The perihelion vs. inclination ambiguity arises because these quantities are loosely related amongst the Classical objects. Doressoundiram ([35]) finds that the color vs. perihelion distance correlation is slightly stronger than the color vs. inclination correlation. Trujillo and Brown ([150]) find that classical objects with small inclinations are redder, on average, than those with high inclinations. The latter observation has been factored into dynamical models by R. Gomes ([53]). He asserts that the high inclination (“hot”) Classical KBOs were scattered outwards while the low inclination (“cold”) Classical KBOs were formed exterior to Neptune, where they now reside ([53]). Whether or not this is true, the central mystery that is unaddressed by dynamical models is why the cold and hot populations would have different colors. As measured by the B-I color index, the color vs. inclination (or color vs. per-

ihelion) correlation appears secure at the 3σ or 4σ confidence level. However, the correlation is absent when V-R or V-I color indices are used ([136]). One possibility is that the color correlation is forced by the B data (for example, there could be a B-band absorber whose distribution is correlated with inclination or perihelion distance but which would have no effect on color indices at wavelengths longer than B). As new observations are collected it will be interesting to see whether or not the reported correlation will survive. No convincing explanation for the correlation, if real, has been suggested.

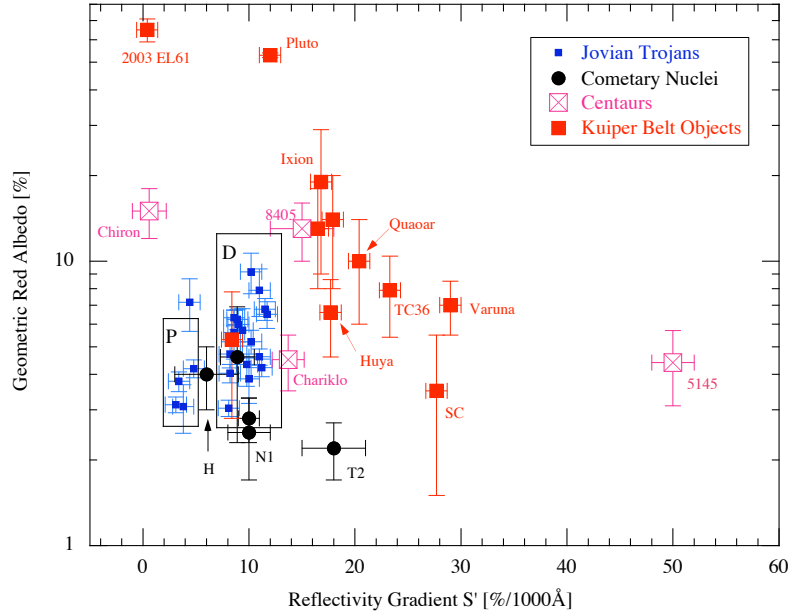


Fig. 34. Color-albedo plane for cometary nuclei, Jovian Trojans, Centaurs and KBOs. Identities of particular objects are abbreviated for clarity: SC = 1993 SC, TC36 = 1999 TC36, H = 1P/Halley, N1 = 28P/Neujmin 1 and T2 = 10P/Tempel 2. Boxes mark the nominal positions of the P- and D-type asteroids. Data compiled from [48], [80] and [25].

About a dozen KBOs possess both color and albedo determinations ([25]). These are plotted in Figure 34 together with corresponding data for the nuclei of comets, the Jovian Trojans, and Centaurs ([80]). There it is seen that the wide dispersion of colors of the Centaurs and KBOs is matched by a wide dispersion in the albedos, with the large objects 2003 EL61 and Pluto defining one extreme. By comparison, the nuclei of the comets and the Jovian Trojans are confined to a small fraction of the color-albedo plane, with surfaces that

are on average less red and darker than the KBOs and Centaurs. The diagram reinforces the conclusion that the surfaces of the comets and of the Trojans, while similar to each other, are not the same as the surfaces of the Centaurs and KBOs. If this difference reflects an evolutionary trend, then the fact that the Centaur and KBOs overlap in Figure 34 shows that the modification occurs after the Centaur phase. Most likely it is associated with the onset of sublimation on bodies whose perihelia have approached or crossed the orbit of Jupiter (the rough boundary outside which water does not appreciably sublimate; [71]). The very high albedos of EL61, Pluto and perhaps some other objects are clearly associated with the presence of surface ice and the cleanliness of this ice suggests that it has been recently emplaced, probably by frost deposition from an atmosphere. None of the Trojans or cometary nuclei possess surface ice in quantities sufficient to influence the albedo, because they are too hot (surface ice would quickly sublimate). However, the simple removal of ice cannot explain why the surfaces of many low albedo KBOs and Centaurs are so much redder than any seen in the comet or Trojan populations (c.f. Figure 23). Some form of instability of the ultrared matter is required.

4.2 Kuiper Belt Physical Properties: Spectra

Only ~ 10 KBOs are bright enough for useful spectra to be obtained. The spectra fall into three basic classes.

The Water Worlds (Figure 35). KBOs (50000) Quaoar ([76]), 2003 EL61 ([141]) and others show strong absorptions at $2.0\ \mu\text{m}$ and $1.5\ \mu\text{m}$ that are diagnostic of water ice. Water ice is stable against sublimation at Kuiper belt distances and temperatures, and it is appropriate to think of it as “bed rock” for other, more volatile species. The ice on Quaoar and 2003 EL61 is known to be crystalline as it shows a narrow band at $1.65\ \mu\text{m}$ that is absent in the spectrum of amorphous ice. This is a puzzle, because ice at the $\sim 40\ \text{K}$ to $50\ \text{K}$ surface temperatures of the KBOs should be indefinitely stable in the amorphous form. Why should the ice instead be crystalline?

Crystallinity indicates that the ice has been raised above the critical temperature for transformation (roughly $100\ \text{K}$ or $110\ \text{K}$) at some point in its history. This heating could have occurred in the deep interiors of the KBOs provided that there is a way for heated ice at depth to be emplaced onto the surface. One way for this to occur is through the past action of cryovolcanism; liquid water (or slush) might have erupted onto the surfaces of these KBOs when they were still internally hot from the decay of trapped radioactive nuclei. Conceivably, heating by micrometeorites is responsible, although this possibility is difficult to test given that we do not know the flux of impacting dust particles within the Kuiper belt. A more serious problem is that crystalline ice exposed to the unimpeded bombardment of energetic particles from the Solar wind and the cosmic rays should be transformed back towards the amorphous state, as the bonds in the crystalline lattice are systematically

demolished. The timescale for this process is uncertain but probably short (~ 1 Myr to 10 Myr). Hence, it appears that these KBOs must be resurfaced on a geologically very short timescale in order for the ice to have escaped back-conversion to the amorphous form. Again, the mechanisms for resurfacing are unknown. Comet-like outgassing (perhaps with CO playing the role of “volatile”) is a possibility, but some effect related to micrometeorite “gardening” of the regolith, as is seen in the rocky fragmental layer on the surface of the Moon, seems more likely. The optically active surface layers may be continually churned together with buried crystalline ice that is protected from irradiation.

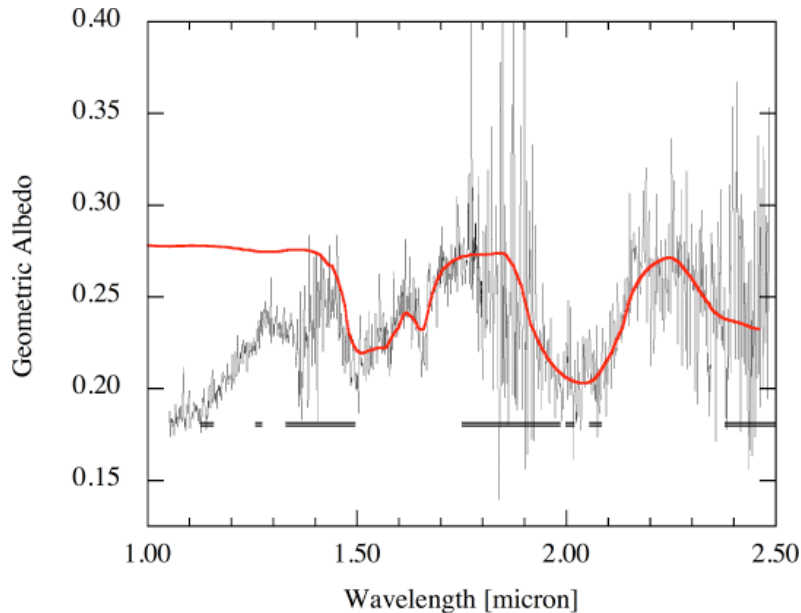


Fig. 35. Near infrared reflection spectrum of (50000) Quaoar. The red line is a crystalline water ice spectrum over-plotted (not fitted) to the data. Note the feature at $1.65 \mu\text{m}$ that is diagnostic of crystalline ice. From [77].

The issue of the crystalline state of water ice in small bodies deserves further exploration. Ice in comets is rarely directly detected, but in comets C/Hale-Bopp ([26]) and C/2002 T7 (LINEAR) ([83]) the absence of the $1.65 \mu\text{m}$ band shows that the ice is amorphous. Both objects are long period comets and it is possible that the amorphous nature of the ice is a result of energetic particle bombardment, rather than primordial in nature. The outgassing activity of some comets at heliocentric distances beyond the ~ 5 AU water sublimation zone (e.g. Figure 24) is often interpreted as evidence for internal heating by the (exothermic) amorphous \rightarrow crystalline phase transition ([125]). An

interesting question to be addressed observationally is the state of the ice in Jupiter family comets: is this ice crystalline as in the large KBOs or amorphous, as in the two measured long-period comets?

The Methanoids (Figure 36). KBOs Pluto, 2003 UB313 ([151]) and 2005 FY9 ([99]) show evidence for surface methane, with distinct bands in the near infrared spectral region. (Triton, likely to be a large KBO captured by Neptune, also shows a methane-rich spectrum).

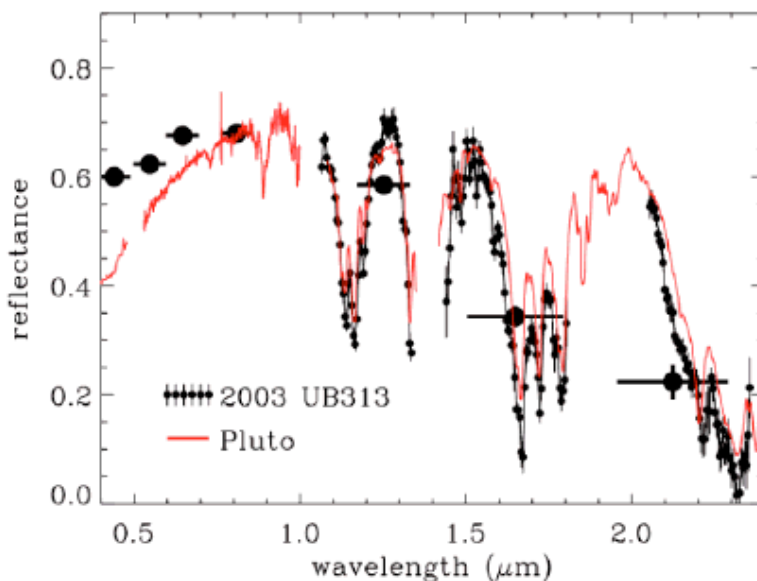


Fig. 36. Optical and near infrared reflection spectra of large KBOs Pluto (red) and 2003 UB313 (black). The principal absorptions in both spectra are due to methane. From [13].

Methane is interesting from two perspectives. First of all, methane is unstable to sublimation on long timescales at the distances and temperatures of most Kuiper belt objects. This instability has been explored in detail for Pluto, where it is found that the escape of methane is limited by the flux of energetic (EUV) Solar radiation ([65]), but can still exceed several kilometers equivalent thickness over the age of the Solar system. The escape from smaller bodies will be dramatically faster, perhaps explaining why the known Methanoids are large (but not explaining why ~ 1200 km diameter Quaoar is methane-free). Second, the origin of the methane is problematic. Low temperatures and pressures in the Solar nebula are thought to favor the incorporation of carbon atoms in the oxidized form as CO and CO₂, rather than in the re-

duced form of CH_4 ([124]). Therefore, it seems unlikely that the methane was delivered to these bodies from the nebula. One possibility is that CH_4 arrived as a clathrate (a physical cage in crystalline water ice in which sufficiently small “guest molecules” can be trapped). In my mind, it seems more likely that the CH_4 is produced in the interiors of these bodies, probably from hydrogen released by serpentinization followed by Fischer-Tropsch reactions, and then outgassed on to the surface. The lack of methane on small KBOs could then reflect a lack of production, since only bodies large and hot enough to sustain liquid water can experience serpentinization.

Featureless Class. Objects in this class have sloped but otherwise featureless near infrared spectra. Obviously, all spectra are featureless when observed at sufficiently low signal-to-noise ratio, so here “featureless” is probably a relative term and many objects labeled as such will resolve into the other classes once better spectra are secured. By analogy with the featureless spectra of many mantled objects already observed at decent signal-to-noise ratios, including the nuclei of dead comets and the Jovian Trojans (e.g. [101]; [37]; [42]), however, it is likely that a subset of the featureless objects will remain so even under more intense scrutiny.

4.3 Kuiper Belt Physical Properties: Shapes, Spins

The shapes and spins of Kuiper belt objects are studied from their rotational lightcurves ([132]; [86]). The most informative way to present these data is in a plot showing the photometric range as a function of the rotational frequency (rotations per day), as here in Figure 37, from [132]. The range-frequency plane is divided into three regions, based on the original prescription of Leone et al. ([93]).

Region A shows lightcurves of small range and any period, for which the lightcurve could be affected by surface albedo variations and for which, in any case, the interpretation is likely to be highly ambiguous. Strictly, *any* lightcurve can be produced by a surface albedo distribution of arbitrary complexity. However, studies of the lightcurves of hundreds of asteroids show few examples where albedo variations are important, perhaps because regolith transport is efficient and albedo differences are quickly smeared out by the redistribution of dust. Those examples are confined to rotational ranges $\Delta m \sim 0.1$ to 0.2 mag. To be conservative, in Figure 37 we have marked Region A as extending up to $\Delta m = 0.3$ mag. The most notable exception to this empirical rule is Saturn’s 1460 km diameter satellite Iapetus, which shows a hemispherical albedo asymmetry, with the leading hemisphere being ~ 6 times darker than the trailing. However, the Iapetus albedo asymmetry is a consequence of its synchronous rotation about the planet (which leads to hemispherically asymmetric fluxes of incident charged particles from Saturn’s magnetosphere and of Saturn-orbiting dust particles), a circumstance which is not replicated in the KBOs.

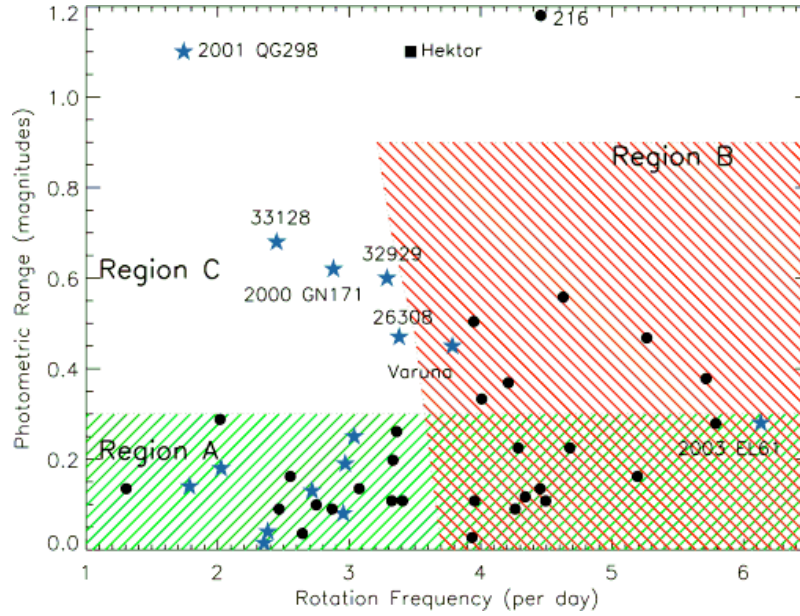


Fig. 37. Rotational range vs. frequency (rotations per day), modified by Scott Sheppard from [132]. Black dots denote large main-belt asteroids (diameters >200 km) while KBOs are marked as blue stars. Note that Hektor is a Jovian Trojan while 33128 is a Centaur.

Region B shows objects rotating sufficiently rapidly that centripetal distortion of the shape constitutes a likely explanation of the lightcurve. The region is marked for an assumed density $\rho = 1000 \text{ kg m}^{-3}$ and calculated from the figures of equilibrium by Chandrasekhar ([17]). Higher (lower) densities would push the left boundary of Region B to the right (left). The implicit assumption is that the tensile strengths are zero and, while this is unlikely to be exactly correct it is a reasonable approximation for bodies which have been internally fractured by past collisions. Two KBOs fall in Region B; (20000) Varuna ($\rho \sim 1000 \text{ kg m}^{-3}$, [72]), and 2003 EL61 ($\rho \sim 2600$ to 3340 kg m^{-3} , [126]).

Region C shows locations in the range vs. frequency plot where close and contact binaries would plot. A binary consisting of two spheres viewed equatorially would have $\Delta m = 2.5 \log(2) = 0.7$ mag. Mutual gravitational deformation would elongate the components, raising Δm to 0.9 mag. ([93]). Objects with $\Delta m > 0.9$ mag are not explainable as rotationally deformed single bodies and contact binaries are preferred. In the whole Solar system, very few objects have been found with such large photometric range. The main examples are Trojan (624) Hektor, which is believed to be a 150 km scale binary, 200 km main belt asteroid (216) Kleopatra and ~ 260 km KBO 2001 QG298 ([132]). The inferred abundance (admittedly from a single detection) of contact or very close binaries in the Kuiper belt is at least 10% to 20% ([132]).

To give a short summary, rotational studies of KBOs have revealed a number of interesting cases for rotational deformation (Varuna and 2003 EL61) and close or contact binaries (the best case remains 2001 QG298 but other KBOs in Region C of Figure 37, like 2000 GN171, are candidates for contact binaries observed non-equatorially). The appearance of these examples in a still-small ($N \sim 40$) observational sample is evidence that rotationally deformed and contact-binary structures must be common in the Kuiper belt. Preliminary evidence suggests that the shape distributions of KBOs larger and smaller than 400 km diameter are not the same ([86]). If confirmed by future work, this observation might find a natural explanation in terms of collisional effects at small sizes and self-gravity at larger sizes.

4.4 Kuiper Belt Physical Properties: Multiple Objects

About 20 examples of multiple KBOs have been reported as of early 2006 (many are not yet properly published, appearing only in electronic circulars). Multiple KBOs in Table 5 have been collected from [32] and [117] and from a few recent electronic publications. The objects are binaries except for Pluto (three satellites known) and 2003 EL61 (two satellites known), but this is no doubt an effect of observational selection against small, faint companions and a larger fraction of the KBOs must have multiple satellites. The largest satellite of the first-known (but mis-labeled) KBO Pluto has been known for decades, but it still surprising to see how many KBOs observed at high angular resolution are double. What can we learn from the binaries?

Firstly, binaries are present in the Classical, Scattered and Resonant KBO populations. Systematic observations of 81 KBOs spread across these classes reveal 9 binaries at the resolution (and magnitude difference) accessible to the Hubble Space Telescope and its NICMOS camera, giving an average binary fraction of $11^{+5}_{-2}\%$ ([137]). Given that binaries of very small separation and those having a large magnitude difference between the components cannot be detected, this must be taken as a strong lower limit to the binary fraction.

Secondly, low inclination ($i < 5^\circ$) Classical KBOs have a binary fraction $22^{+10}_{-5}\%$ ([137]), which is different from the average value at the $\sim 2\sigma$ level. The mean value for all KBOs other than the $i < 5^\circ$ Classicals is $5.5^{+4}_{-2}\%$, which is different enough from $22^{+10}_{-5}\%$ to be interesting. The difference, if real, could be a hint that the diverse dynamical histories of the bodies have had an effect on the survival of binaries. For example, perhaps whatever excited the orbital inclinations and eccentricities of KBOs also acted to split a fraction of the binaries.

Thirdly, the binaries appear to be of different types. Pluto (and probably 2003 UB313 and others) have short orbital periods and orbital eccentricities $e \sim 0$. Together these strongly suggest the effects of tidal damping. Close binaries like these might be produced by glancing impacts between large precursors ([15]). The number density of large KBOs is presently far too low to account for such collisions. If this is the correct explanation, the collisionally

Table 5. Multiple KBOs

Object	a [km] ^a	e ^b	i[deg] ^c	Type ^d	θ [arcsec] ^e	P[days] ^f	Δ mag
Pluto				3:2			
Charon	19,400	0.00	96	—	0.9	6.4	1.3
S/2005 P1	64,700?	—	—	—	2.2	38.3?	9.0
S/2005 P2	49,400?	—	—	—	1.7	25.5?	9.4
1995 TL ₈	—	—	—	Sca	—	—	—
(58534) 1997 CQ ₂₉	8,010(80)	0.45	—	Clas	0.2	312(3)	0.3
(26308) 1998 SM ₁₆₅	11,310(11)	—	—	2:1	0.2	130	1.9
1998 WW ₃₁	22,300	0.82	42	Clas	1.2	574	0.4
(79360) 1999 CS ₂₉	—	—	—	Clas	—	—	—
1999 OJ ₄	—	—	—	Clas	—	—	—
1999 RZ ₂₅₃	4,660(170)	0.46	—	Clas	—	46	—
(47171) 1999 TC ₃₆	7,640(460)	—	—	3:2	0.4	50.5	1.9
2000 CF ₁₀₅	—	—	—	Clas	0.8	—	0.9
2000 CQ ₁₁₄	—	—	—	Clas	—	—	—
2000 CM ₁₀₅	—	—	—	Clas	—	—	—
2000 CM ₁₁₄	—	—	—	Clas	0.07	—	0.5
2000 OJ ₆₇	—	—	—	Clas	—	—	—
2000 YW ₁₃₄	—	—	—	Sca	—	—	—
2001 QC ₂₉₈	3,690(70)	—	—	Clas	0.17	19.2	N/A
(88611) 2001 QT ₂₉₇	27,300(340)	0.24	—	Clas	0.6	—	0.5
2001 QW ₃₂₂	—	—	—	Clas	4.0	—	0.4
2002 CR ₄₆	—	—	—	Sca	0.11	—	1.2
2003 EL ₆₁	—	—	—	Sca	—	—	—
S/2005 (2003 EL ₆₁) 1	49,500(400)	0.050(0.003)	234.8(0.3)	—	1.3	49.12±0.03	3.3
S/2005 (2003 EL ₆₁) 2	39,300?	—	—	—	1.0	34.1?	4.5
2003 QY ₉₀	—	—	—	Clas	—	—	—
2003 UB ₃₁₃	36,000	—	—	Sca	0.5	14	4.2
2003 UN ₂₈₄	—	—	—	Clas	—	—	—
2005 EO ₃₀₄	—	—	—	Clas	—	—	—

a: semi-major axis of the binary system

b: eccentricity

c: inclination

d: Dynamical type: 3:2, 2:1 = resonant, Clas = Classical, Sca = Scattered

e: Angular separation

f: Orbital period

produced binaries must be relics from an earlier time at which the number density in the belt was much (probably two to three orders of magnitude) higher than now ([71]; [15]).

Where measured, most KBOs have periods from months to years and the eccentricities of the orbits are in the range $0.2 \leq e \leq 0.8$ (see Table 5). These wider, more eccentric binaries are unlike the binaries expected to be produced by glancing, massive impacts and other explanations must be sought. Several have already been proposed, including binary formation through dynamical friction ([52]), three-body interactions ([52]; [154]) and exchange reactions ([50]). These models are all good in the sense that they make observationally testable predictions. The exchange model predicts binary eccentricities larger than observed, and can probably be ruled out, at least in its simplest form. Three-body interactions should produce mainly weakly bound binaries. It is not yet clear if the distribution of semimajor axes of the known binaries is in-

compatible with three-body captures, but this seems likely (Table 5). Capture by dynamical friction (exerted on large, growing bodies by the “sea” of smaller bodies surrounding them and now dissipated) is expected to produce a large binary fraction (as observed) with a high abundance of tight binaries (maybe consistent with the data). Continued action of dynamical friction should lead the binary components to spiral together, making contact binaries (one, 2001 QG298, is already suspected) but it is not clear that observed eccentricities $0.2 \leq e \leq 0.8$ can be explained. At this early stage, I do not know if the proposed models fail because they are completely wrong, or because they tell only part of the story. Binaries could form by dynamical friction, for example, and then be excited by external agents after the source of dynamical friction had dissipated. Long-term (4 Gyr) survival of the KBO binaries appears to be possible, but the existing pairs may constitute only a fraction of those initially present, with the softest binaries having all been disrupted ([121]).

4.5 Kuiper Belt Physical Properties: Densities

Densities have been discussed here and there throughout this chapter. For convenience, I have summarized them graphically in Figure 38, where they are plotted as a function of the object diameters. The densities of cometary nuclei plotted in the Figure have been estimated from a variety of techniques as discussed in Section 3. Densities of KBOs are estimated from binary motions and size estimates (Pluto, Charon and 1999 TC36, [135]), from lightcurves interpreted as rotational deformation of the shape ((20000) Varuna ([71]) and 2003 EL61 ([126])) and from a contact binary model (2001 QG298, [132]). The densities of the planetary satellites are obtained nearly directly from gravitational perturbations on the motions of spacecraft, except that the densities of small Saturnian satellites including Pandora and Prometheus are estimated from a more complicated model of these satellites’ interaction with nearby rings.

What does Figure 38 show? The most obvious feature is a general trend towards higher densities at larger diameters, adequately described by the power law relation $\rho = 340 D^{0.2}$ (with ρ in kg m^{-3} and D in km). This trend is apparent within the various populations (i.e. the planetary satellites and the KBOs independently show this trend) and, although there is considerable scatter in the densities of bodies at any particular diameter, the trend appears to be real.

The mean density of a composite body consisting of rock and ice is

$$\bar{\rho} = \rho_i f_i + \rho_r f_r \quad (21)$$

where ρ_i and ρ_r are the densities of ice and rock and f_i and f_r are the fractional volumes occupied by ice and rock, respectively. The latter are related by

$$f_i + f_r + f_v \equiv 1 \quad (22)$$

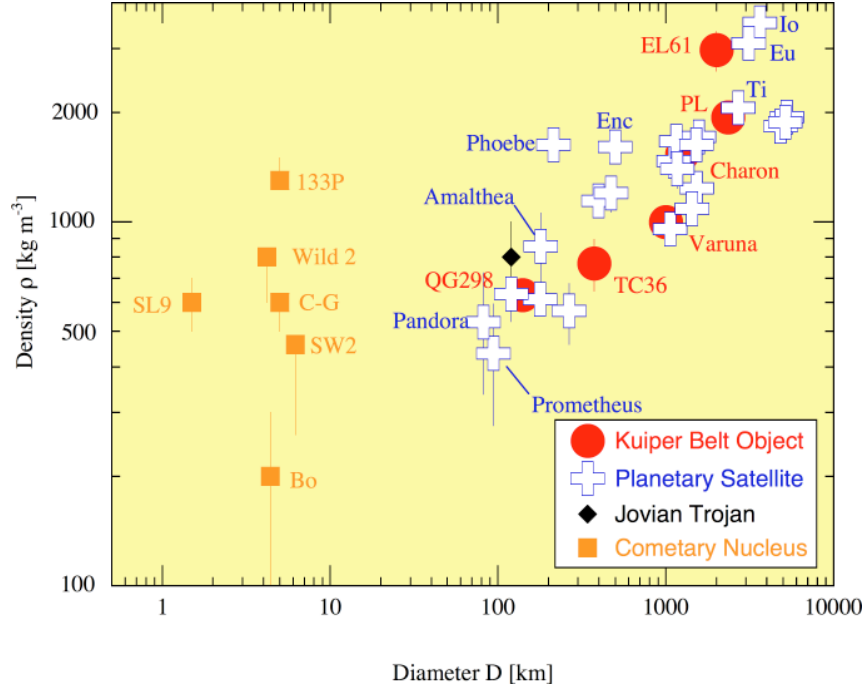


Fig. 38. Densities of KBOs, cometary nuclei, planetary satellites and Jovian Trojan Patroclus. Abbreviations in the plot are **Comets** Bo = 19P/Borrelly ([28]), C-G = 67P/Churyumov-Gerasimenko ([29]), SL9 = D/Shoemaker-Levy 9 ([6]), SW2 = 31P/Schwassmann-Wachmann 2, Wild 2 = 81P/Wild 2 ([30]), 133P = 133P/Elst-Pizarro ([62]) **Kuiper Belt Objects** EL61 = 2003 EL61 ([126]), PL = Pluto, TC36 = 1999 TC36 ([?]), QG298 = 2001 QG298 ([132]; [139]) **Planetary Satellites** Enc = Enceladus, Ti = Titan, Eu = Europa. These densities are culled from the NASA-JPL site at <http://ssd.jpl.nasa.gov/>, mostly based on data from the Voyager, Galileo and Cassini missions. The single Trojan is (617) Patroclus ([105]). Plotted error bars are 1σ uncertainties. Single-sided errors below or above the points indicate either upper limits and lower limits to the density, respectively.

in which f_v is the fractional void space, also known as “porosity”. In the context of Figure 38, much of the trend in the bulk density is likely to be related to size-dependent variations in f_v . This is because self-compression of ice and rock is not very important across most of the plotted diameter range (the central hydrostatic pressure in a body of radius r and average density ρ is $P_c \sim G\rho^2 r^2$. With $\rho = 1000 \text{ kg m}^{-3}$ and $r = 500 \text{ km}$, $P_c \sim 20 \text{ MPa}$ ($1 \text{ MPa} = 10^6 \text{ N m}^{-2}$), or roughly 200 bars, but densification via the collapse of void space is likely. Laboratory experiments with ice at 77 K show brittle failure at comparable pressures ([39]) and suggest that part of the density-radius correlation may result from self-compression, particularly by the closing of void-space in porous bodies ([71]; [107]).

Any object less dense than pure water ice ($\rho \sim 1000 \text{ kg m}^{-3}$) must be porous. This includes most of the comets in Figure 38 (but not 133P, the one MBC for which we possess a density constraint) and several of the co-orbital satellites of Saturn (Pandora and Prometheus both have $\rho \sim 500 \text{ kg m}^{-3}$). More surprisingly, Jupiter’s innermost satellite Amalthea ($\sim 160 \text{ km}$ in diameter) has $\rho = 800 \pm 200 \text{ kg m}^{-3}$ ([5]) and so is likely porous and ice-rich. This is a big surprise, given that prior to the density determination Amalthea was always described as one of the most refractory, high-temperature products of Jupiter’s long-gone accretion disk. The evidence for porosity is strong and independent infrared spectral observations ([140]) show a deep hydration feature that supports a watery constitution.

Porosity can be due to large, empty spaces (“macroporosity”) or to open structure on a small scale “microporosity”, and everything in between. Microporosity in stony meteorites averages 10% and can reach 30% in some samples ([12]). Macroporosity can be produced by past impacts that have cracked and even dissociated bodies leading to their re-assembly as a collection of irregularly shaped blocks with considerable internal void space. Evidence for this is seen even in the main-asteroid belt (e.g. rocky asteroid (253) Mathilde has $\rho = 1300 \pm 200 \text{ kg m}^{-3}$; [157]; [11]). Porosity caused by collisional shattering and reassembly should become less important at larger diameters both because sufficiently energetic impacts are rare and because of closure of pore space at the higher hydrostatic pressures in large objects. I suspect that most of the slope in the ρ vs. D relation seen in Figure 38 is caused by systematic decrease in the porosity as D grows larger. The equation of state for self-gravitating ice bodies ([100]) is too flat to fit the trend apparent in Figure 38.

4.6 Centaurs

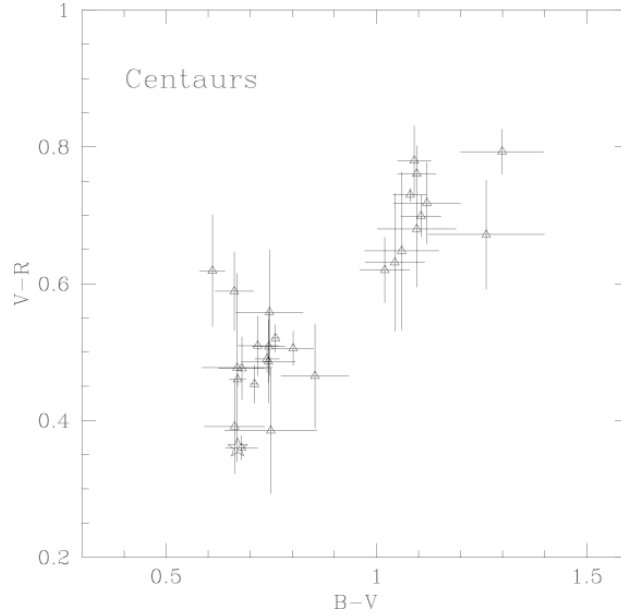
The Centaurs are bodies strongly interacting with the giant planets. Several definitions exist. When defined as non-Trojan bodies having both perihelia and semimajor axes between the orbits of Jupiter, $a_J = 5.2 \text{ AU}$, and Neptune, $a_N = 30 \text{ AU}$, there are about 87 known examples of Centaurs as of early 2006. Of these, five or six display comae and so are double-designated as comets (the most famous and prototypical example is (2060) Chiron; Table 6). A detailed classification scheme has recently been proposed ([61]).

The differential size distribution of the Centaurs is consistent with a power law having an index $q \sim 4$ and this is compatible with the size distribution measured for KBOs of similar size ([130]). The known Centaurs tend to be intermediate in size between the nuclei of well-studied comets (typically a few to 10 km diameter) and the well-studied KBOs (mostly $\sim 100 \text{ km}$ to $\sim 1000 \text{ km}$ diameter). The latter is simply an effect of selection: the Centaurs are intermediate in distance between the perihelia of the well-studied comets and the Kuiper belt.

In terms of their albedos and surface colors, the Centaurs resemble the KBOs more than any other Solar system population (Figure 34). This is con-

Table 6. The Known Cometary Centaurs

Object	Perihelion [AU]	Semimajor Axis [AU]	Eccentricity	Inclination [deg]	T_J
C/2001 M10	5.30	26.66	0.80	28.0	2.59
29P/SW1	5.72	5.99	0.04	9.4	2.98
39P/Oterma	6.83	7.25	0.24	1.9	3.01
2060 Chiron	8.45	13.62	0.38	6.9	3.36
C/2001 T4	8.56	13.92	0.38	15.4	3.29
(60558) 2000 EC98	5.83	10.73	0.46	4.3	3.03

**Fig. 39.** Color-color diagram for Centaurs showing evidence for bimodality. From [32].

sistent with the recent extraction of the Centaurs from the Kuiper Belt. Dynamical models (see the parallel Saas Fee review by Alessandro Morbidelli) show that the Centaurs have dynamical lifetimes limited by scattering from (and occasional impacts into) the giant planets. Their most usual fate is to be ejected to the interstellar medium, on a median timescale $\sim 10^8$ yr, but some survive entanglement with Jupiter and are kicked into orbits with perihelia $q < 5$ AU, where they begin to sublimate strongly in the heat of the sun, and are thereafter labeled as comets.

One property that has been reported to differ between the Centaurs and KBOs is the distribution of optical colors ([120]; [146]). The available data suggest that the Centaur colors may be bimodally distributed (Figure 39)

whereas the KBOs, as noted in a previous section (Figures 31, 32 and 33), are not. It is tempting to imagine that this effect (which is formally statistically significant) could be caused by past or present activity on the Centaurs. However, a search for correlations between Centaur color and such likely indicators as perihelion distance, semimajor axis, nucleus size or current outgassing activity has revealed nothing of importance. The Centaur bimodality, if it is real, is unexplained.

4.7 Irregular Satellites

There are two, largely distinct types of planetary satellite, based on dynamical characteristics. The most familiar satellites have small eccentricities and inclinations, and orbit from a few to a few dozen planetary radii from their parents planets. These are the regular satellites, most thought to have formed by accretion within circumplanetary disks that were present around the planets during the formation epoch (the details of satellite formation in disks remain obscure, and are the subject of interesting speculation and on-going research). Other satellites, out-numbering the regulars by a considerable margin, follow eccentric and highly inclined orbits with large semimajor axes. These “irregular satellites” in fact sweep-out a considerable fraction of the Hill spheres of their planets. The Hill sphere is the region in which planetary gravity is dominant over Solar gravity and has radius (roughly the distance from the planet to the innermost Lagrange point) of

$$R_H = a \left(\frac{m_p}{3M_\odot} \right)^{1/3} \quad (23)$$

where a is the planet semimajor axis, m_p is the planet mass and M_\odot is the Solar mass (Table 7). The irregular satellites are interesting in the context of the Saas Fee workshop from the point of view of their possible origin. They cannot have been formed like the regular satellites within accretion disks surrounding the planets. Instead, they must have been captured. It is not known from where they were captured, but there are two broad possibilities. First, they might have had a local source. The irregular satellites could be planetesimals that were initially in heliocentric orbits and were captured by the planets as a result of their sudden growth (we will discuss the “standard models” for satellite capture in a moment). In this case the irregulars are interesting because they are surviving samples of the kinds of solid bodies most of which were accreted into the cores of the giant planets, or which were scattered out of the Solar system soon after the giant planets attained their final masses. A second possibility is that the irregular satellites are captured comets or, equivalently, captured KBOs. In this case the irregulars would take on new significance as (relatively) local examples of objects from the much more distant Kuiper Belt.

Research into the irregular satellites is in the midst of a sudden burst of new work, driven by the application of large-format CCD detectors to the

problem of their detection. Less than a dozen irregular satellites were discovered in the entire 20th Century. Most of these were chance detections made by observers using photographic plates and long exposures on large telescopes. Within the past \sim half-decade, nearly 100 new irregulars have been identified, most as the result of surveys conducted using a variety of telescopes and large cameras on Mauna Kea ([131], [133]; an updated summary of the data may be found at <http://www.ifa.hawaii.edu/~jewitt/irregulars.html>). These surveys continue, and more irregular satellites discoveries are anticipated, but we already are beginning to see new patterns in the distribution of the satellites that raise problems concerning the mechanisms of capture.

Table 7. Planetary Hill Spheres

Object	Mass/ M_{\oplus} ^(a)	a [AU] ^(b)	R_H [AU] ^(c)	θ_H [deg] ^(d)
Jupiter	316	5	0.35	5
Saturn	95	10	0.43	2.8
Uranus	15	20	0.47	1.4
Neptune	17	30	0.77	1.5

- (a) Planetary mass. (b) semimajor axis. (c) Hill sphere radius from Equation (23).
 (d) apparent angular radius of the Hill sphere from Earth.

The central problem of permanent capture is that a body which follows an orbit initially unbound to a planet must lose or otherwise redistribute some of its kinetic energy in order to become bound to the planet. For a long time, the standard model for the capture of the irregular satellites has been through the action of gas drag forces on heliocentric planetesimals passing through the bloated gaseous envelopes of the young giant planets. This model, which was developed in parallel with models for the formation of gas giant planets like Jupiter and Saturn, implies that the irregular satellites observed today are those objects which were neither too small (ablated and absorbed in the gaseous envelopes like meteors in the Earth's upper atmosphere) nor too large (passed through the envelopes with negligible deceleration to continue in heliocentric orbits). It also relies upon the sudden collapse of the extended envelopes in order to leave the satellites behind: continued friction would lead to all trapped bodies spiraling into the planets.

A problem with this gas-drag capture model is that the new surveys show that Uranus and Neptune possess irregular satellite systems of their own. In fact, when corrected for the magnitude-limited nature of the observational surveys to the best of our ability, the new surveys show that the gas giants and the ice giants possess about the same number of irregular satellites, measured down to a given satellite absolute magnitude or size. This is seen by comparing Figure 40 (the apparent magnitude distributions of the satellites of all four giant planets) with Figure 41 (same as Figure 40 but corrected for the varying

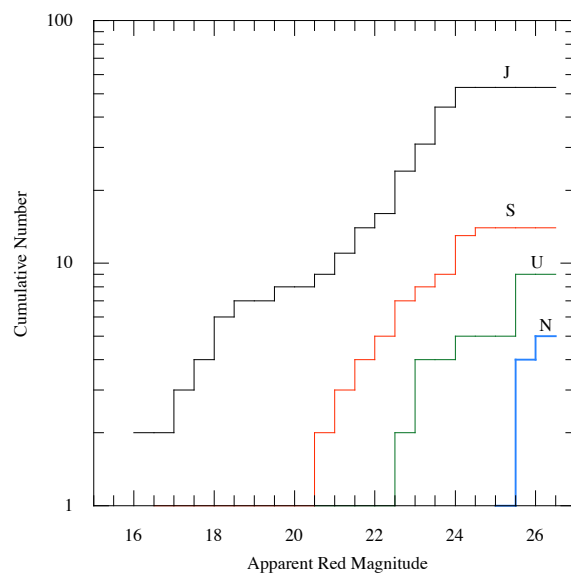


Fig. 40. Number of irregular satellites of each planet as a function of apparent magnitude. From [78].

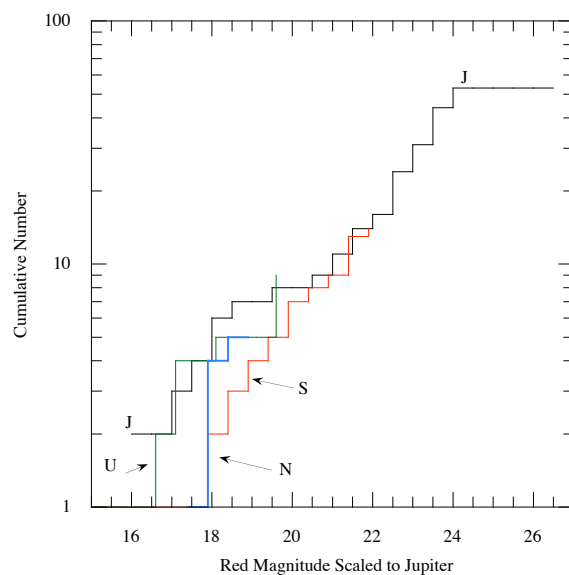


Fig. 41. Number of irregular satellites of each planet as a function of reduced magnitude (i.e. corrected for their differing heliocentric and geocentric distances using the inverse square law). From [78]

distances of the planets using the inverse-square law ([78]). Within the errors, the irregular satellite absolute brightness (size) distributions are the same. This is a remarkable and unexpected observational result. It is difficult to see how Uranus and Neptune, which are relatively gas-free ice giants, formed by processes quite different from those that produced the gas giants Jupiter and Saturn, could capture the irregular satellites by gas drag. At least, gas-drag capture has never been demonstrated for the ice giants in any publication of which I am aware. Taken as a whole, the uniform abundance of irregular satellites around the gas and ice giants argues against gas-drag capture.

What about other capture processes? A separate mechanism has been proposed in which runaway mass-growth of a planet leads to the permanent trapping of objects initially moving within the Hill sphere. This is called “pull-down capture” ([59]). Like gas-drag, pull-down capture works best for the gas giants, which had a runaway growth of mass as they attracted gas from the protoplanetary nebula in a hydrodynamic in-flow (Section 2). The ice giants, instead, grew slowly by successive collisions with solid bodies in the disk and they did not experience a runaway growth in mass. Therefore, it seems unlikely that pull-down capture can explain the irregular satellite systematics revealed in Figure 41.

This leaves the generic class of “three-body interactions” as possible explanation of the capture of the irregular satellites. Three-body capture is appealing because it separates the capture mechanism from the details of planet formation. All that is needed is a sufficient density of objects for three-body interactions (two small bodies within the Hill-sphere of a large one) to occur with high enough frequency to be relevant. Although suggested long ago ([22]), three-body captures have rarely been discussed in the context of the irregular satellites precisely because the densities of small bodies in the Solar system are so low that the frequency of interaction is negligible. Our changing perspective, in which the density of small bodies may have been hundreds or thousands of times larger than now, makes three-body processes more attractive.

Is there any evidence that the irregular satellites were captured from a local source as opposed to a Kuiper belt source, or vice versa? The color distribution of the irregular satellites is different from the color distribution in the Kuiper belt ([54], [55]) with the main difference being that the ultrared matter is absent on the satellites but common on both KBOs and Centaurs. This could indicate that the Kuiper belt is not the source of the irregular satellites, suggesting that sources local to each planet are more likely. Alternatively, there could be a delivery mechanism from the Kuiper belt which operates selectively to exclude the ultrared objects. At Jupiter, it is possible that the colors of the satellites have been modified by rubble mantle formation or by another process, as is inferred for the Trojans at the same heliocentric distance. The authors of the Nice, France model ([114]) are careful to note that objects captured by Jupiter as Trojans have mostly spent time at smaller heliocentric distances (by which they mean to say that the color differences

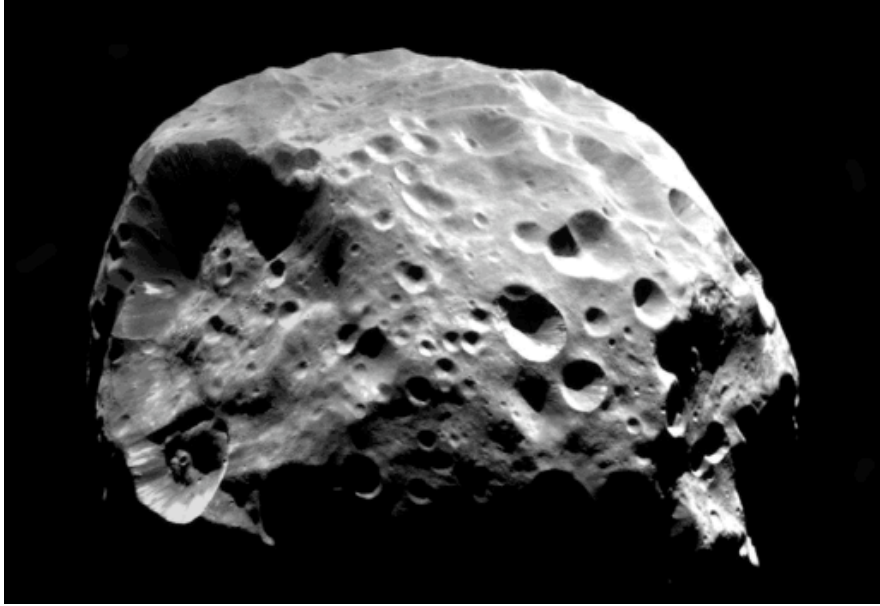


Fig. 42. Saturnian irregular satellite Phoebe, roughly 220 km in diameter and in possession of a magnificent impact crater almost half its size. Courtesy Cassini Imaging Team and NASA/JPL/SSI.

between Trojans and KBOs may be explained by past outgassing). The same argument could be made for the irregular satellites of Jupiter. Modification by mantling seems unlikely at Saturn, Uranus and Neptune, however, because of the lower temperatures at 10, 20 and 30 AU and the expected lack of sublimation driven activity at these distances.

The size distribution of the irregular satellites ($q \sim 2$; [78]) is flatter than the corresponding distribution of the KBOs ($q \sim 4$; [149]). This does not rule out an origin by the capture of KBOs, however, because the satellite size distribution could have been strongly modified either by the capture process or by size-dependent evolutionary effects ([116]).

Measurements of the density ($1630 \pm 33 \text{ kg m}^{-3}$) of Saturn's large irregular satellite Phoebe (Figure 42) have been claimed as evidence for Kuiper belt origin ([81]). The argument is that Phoebe is denser than most other Saturnian satellites and that the higher density more closely resembles the densities of Kuiper belt objects such as Pluto and Triton ($\rho \sim 1900 \text{ kg m}^{-3}$). This is a difficult argument to sustain, however, given that the densities of KBOs seem to vary over a wide range and that the Saturnian regular satellite Enceladus has a density ($1606 \pm 12 \text{ kg m}^{-3}$) essentially identical to that of Phoebe (but there is no suggestion that Enceladus is captured). I note without further comment that the *low* density of Jovian Trojan (617) Patroclus ($\rho = 800^{+200}_{-100} \text{ kg m}^{-3}$) has been asserted as evidence for its origin by capture from the

Kuiper Belt ([105]). The bottom line is that there is no simple link between density and formation location, and it seems impossible to me to use one to predict the other.

Measurements of diverse surface composition on Phoebe, including ices of water, trapped CO₂, and organics and cyanide compounds, suggest to some that this body was formed at a remote location and then captured ([21]). Again, the argument is an indirect one and, as the authors note, it is possible that the surface ice on Phoebe is in part a coating from the impact of a comet itself from distant regions.

4.8 Trojans

The origin of the Trojans has long been a source of mystery. Objects colliding near the Lagrangian L4 and L5 resonances have a small but finite probability of being captured there, particularly if they were already nearly co-moving with Jupiter ([158]; [104]; [19]). Icy asteroids near the growing Jupiter could also be pulled into trapped orbits by the mass growth of Jupiter ([49]; [104]). It has also been suggested that the Trojans might have originated at remote locations in the Solar system, and were captured through the action of out-gassing forces ([158]) or a chaotic disturbance that would have resulted if Jupiter and Saturn were once in 2:1 mean-motion resonance with each other ([114]).

In terms of what we know from observations the Trojans may have no connection at all to the Kuiper belt or they may be genetically closely related. The observational constraints are presently too weak for us to determine the origin of these intriguing bodies at any level above the conjectural. One reason for this sorry state of affairs is that most Trojans are twice as distant and so $2^4 = 16$ times fainter than main-belt asteroids of corresponding size. By comparison, the main-belt asteroids represent “low hanging fruit” to most observers and so they have received the lion’s share of the attention. This situation has only recently started to change. Indeed, until recently *only* Jovian Trojans were known. Now we are also aware of Trojans of Mars and of Neptune. Planned all-sky surveys should greatly improve our knowledge of the populations and size distributions of these bodies. In this section, we briefly review the known properties of the Trojans and compare them with the KBOs and other bodies.

Surveys show that the number of Jovian Trojans rivals the number of main-belt asteroids when measured down to a common limiting diameter ([134]; [70]). There are about 1.5×10^5 Trojans larger than 1 km in radius. They occupy two banana-shaped clouds in Jupiter’s orbit, leading and trailing the planet by $\pm 60^\circ$. Objects in the clouds librate around the L4 and L5 Lagrangian points in response to the combined gravitational attractions to the Sun and Jupiter (see [49] for a nice discussion of Trojan dynamics, from which the following is taken). In the idealized planar, restricted three-body (Sun-Jupiter-Trojan) approximation their equation of motion is

$$\frac{d^2\phi}{dt^2} = \left(\frac{27}{4}\right) \mu n_J^2 \phi = 0 \quad (24)$$

where ϕ is the angular separation between the Trojan and its Lagrangian point, t is time, μ is approximately the ratio of the mass of Jupiter to the mass of the Sun and n_J is the mean motion of Jupiter in its orbit. The solution to Equation 24 is

$$\phi = \frac{A}{2} \cos \left[\left(\frac{27}{4} \mu n_J^2 \right)^{1/2} t + B \right] \quad (25)$$

where A and B are constants representing the amplitude of the libration and the phase, respectively. With $\mu \sim 0.001$, $n_J \sim 0.52 \text{ yr}^{-1}$, the characteristic frequency is $\omega = (27\mu n_J^2/4)^{1/2} \sim 0.043 \text{ yr}^{-1}$, corresponding to a period $2\pi/\omega \sim 150 \text{ yrs}$, or about 10 times Jupiter's orbital period. The distribution of libration amplitudes, A , is very broad, with a mean near 30° ([134]; [110]). In addition to wide excursions about the Lagrangian points in the orbital plane of the planet, the Trojans also occupy a broad distribution of inclinations, with a bias-corrected mean of 14° ([69]) to 17° ([134]). As a consequence, the velocity dispersion amongst the Trojans ($\sim 5 \text{ km s}^{-1}$) rivals that amongst the main-belt asteroids. Collisions between Jovian Trojans are expected to be highly erosive.

Limited work on the long-term stability of Trojans at Jupiter suggests two loss mechanisms. There is a slow leak due to dynamical chaos ([97]), with a timescale that depends on A (larger A being less stable). The more significant loss process is due to ejection from the Lagrangian clouds following collisions or near-miss interactions. Kilometer sized and larger bodies are lost this way at a rate $\sim 10^3 \text{ Myr}^{-1}$ ([103]), meaning that the observed population of small objects should vanish in a few $\times 10^2 \text{ Myr}$. That small Trojans remain is presumably a result of a collisional cascade, with the small objects being both lost and continually supplied from the break-up of larger bodies. Ejected Trojans pursue orbits that are scattered by the planets, quickly becoming indistinguishable from the orbits of Jupiter family comets. Up to $\sim 10\%$ of the latter could be escaped Jovian Trojans ([103]; [70]): the contributions from the Trojan swarms of other planets are unknown.

Several physical properties of the Trojans have been measured. The size distribution is a broken power law ([134]; [70]). Objects with absolute magnitudes $V(1,1,0) < 9.5$ (corresponding to diameters $> 84 \text{ km}$, for albedo 0.04) are well described by a differential power law index $q = 5.5 \pm 0.9$ (Figure 43). Those with $11 \leq V(1,1,0) \leq 14$ (diameters $4.4 \leq D \leq 40 \text{ km}$, for the same albedo) instead have $q = 3.0 \pm 0.3$ ([70]). The index for the smaller objects is close to the nominal value expected for a system in collisional equilibrium ([34]), consistent with the idea that these smaller bodies are part of a collisionally-produced cascade. The steep slope of the large Trojans presumably reflects a "production function": at least, these big bodies seem unlikely to have been

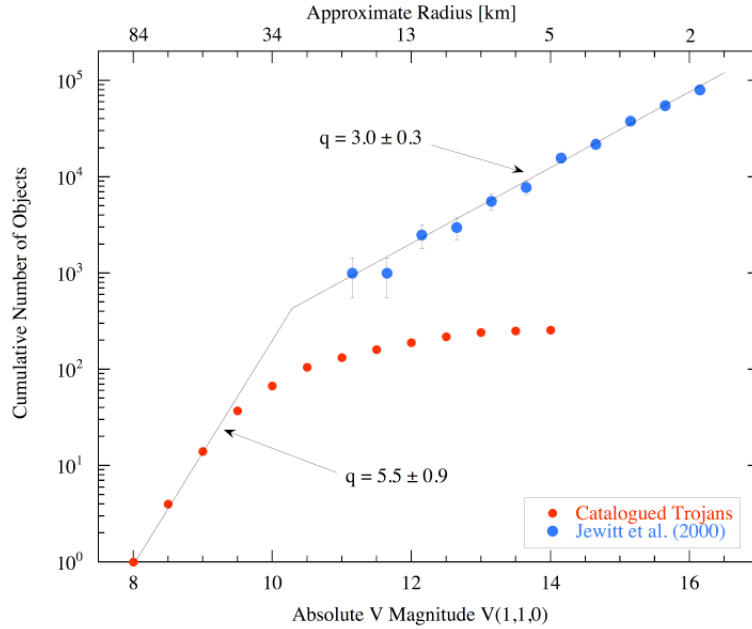


Fig. 43. Brightness distribution of the Jovian Trojans, showing the break in the size distribution. Red points show the numbered Trojans. The roll-over above $V(1,1,0) \sim 10$ is due to observational incompleteness. The blue points are from [70], scaled to correct for the small area of the Trojan swarms observed in that survey. The difference in slope between the large and small objects is independent of the scaling. The radius scale at the top is computed on the assumption that the Trojans all have albedo 0.04. From [70].

molded much by energetic collisions. For comparison, the $D \geq 100$ km KBOs occupy a distribution with $q \sim 4$ ([149]). Within the errors ($\sim 2\sigma$) this is compatible with the size distribution of the larger Trojans, as might be expected if the latter were captured from the Kuiper belt ([114]).

As already noted, the optical color distribution of the Trojans is different from that of the KBOs and Centaurs because the Trojans lack ultrared matter. This could mean that there is no relation between the Trojans and the KBOs or Centaurs, or it could mean that the surfaces of the Trojans have been modified in some way by their exposure to sunlight (as have the surfaces of the Jupiter family comet nuclei, which very likely do come from the Kuiper belt). We prefer the latter explanation, but it does not tell us anything about the source of the Trojans, because the surface modification process could operate regardless of the origin of the bodies. Any object formed beyond the snow-line (perhaps originally at ~ 3 AU or slightly closer) is expected to be icy and should evolve when heated to develop a surface mantle. In the same

vein, the albedo distribution of the Trojans is very narrow compared to that of the KBOs and Centaurs ([48]) but more similar to the nuclei of Jupiter family comets (Figure 34). This is probably also a result of surface modification on bodies that have been heated strongly by the sun but, again, we cannot use this information to specify the source of the Trojans. In terms of their spectra, the Trojans have steadfastly resisted every attempt to assess surface composition from observations taken in the optical and near infrared ([101]; [37]; [42]). The absence of features is consistent with the very dark surfaces of these bodies, and suggests (but does not require) an organic-rich compositional nature ([24]). Observations at thermal wavelengths have revealed features consistent with emission from silicates in three Trojans (624 Hektor, 911 Agamemnon and 1172 Aneas; [25]).

Lastly, the density of Trojan (617) Patroclus has been estimated from infrared observations ([48]) and from its dynamical system mass as $\rho = 800^{+200}_{-100}$ kg m⁻³ ([105]). Although the authors of [105] cite this low density as evidence that Patroclus is a captured KBO, in fact low density is only evidence for a high mass fraction of ice and/or vacuum (“porosity”), and cannot be diagnostic of the Trojan source. In fact, any object formed at any distance beyond the snow-line would be expected to have a high ice content and correspondingly low density. Simply put, “density is not destiny”.

An accurate summary is that the available physical data on Trojans, from their surface colors ([69]) to their albedos ([48]) to the one measured density ([105]) are similar to the corresponding quantities reported for the nuclei of comets but not similar to those of the KBOs. The measured Trojan properties very probably reflect refractory surface mantles left behind following ancient mass-loss, but we cannot uniquely determine the source of the Trojans from the physical data. An interesting exercise for the readers of this article is to think of observations that could be taken to uniquely determine the source of the Trojans. I, for my part, will be trying to do exactly the same.

Acknowledgements

I thank Nick Thomas for inviting me to Saas Fee and NASA and NSF for supporting my research on the comets and Kuiper Belt Objects for many years. Sean Andrews, Audrey Delsanti, Nader Haghighipour, Pedro Lacerda, Rita Mann, Scott Sheppard and Bin Yang kindly read the manuscript.

References

1. M. A'Hearn, R. Millis, D. Schleicher, D. Osip and P. Birch. The Ensemble Properties of Comets. *Icarus* 118, 223-270 (1995).
2. F. Adams and G. Laughlin. Constraints on the Birth Aggregate of the Solar System. *Icarus*, **150**, 151-162 (2001).
3. D. A. Allen. Infrared Diameter of Vesta. *Nature*, **227**, 158-159 (1970).
4. S. Andrews and J. Williams. Circumstellar Dust Disks in Taurus-Auriga. *Ap. J.* **631**, 1134-1160 (2005).
5. J. Anderson et al. Amalthea's Density is Less than that of Water. *Science* 308, 1291-1293 (2005).
6. E. Asphaug and W. Benz. Structure of Comet Shoemaker-Levy 9 Inferred from the Physics of Tidal Breakup. *Icarus*, 121, 225-248 (1996).
7. G. Bernstein et al. The Size Distribution of Trans-Neptunian Objects. *Astron. J.* 128, 1364-1390 (2004).
8. A. Boss. Gas Giant Protoplanet Formation: Disk Instability Models with Thermodynamics and Radiative Transfer. *Ap. J.* 563, 367-373 (2001).
9. A. Boss, G. Wetherill and N. Haghighipour. Rapid Formation of Ice Giant Planets. *Icarus* 156, 291-295 (2002).
10. C. Briceno et al. The CIDA-QUEST Large Scale Survey of Orion OB1: Evidence for Rapid Disk Dissipation in a Dispersed Stellar Population. *Science* 291, 93-97 (2001).
11. D. Britt and G. Consolmagno. The Structure of High Porosity Asteroids. *Icarus* 152, 134-139 (2001).
12. D. Britt and G. Consolmagno. Stony Meteorite Porosities and Densities. *Meteor. Plan. Sci* 38, 1161-1180 (2003).
13. M. Brown, C. Trujillo and D. Rabinowitz. Discovery of a Planetary-Sized Object in the Scattered Kuiper Belt. *Ap. J.* 635, L97-100 (2005).
14. K. Cai et al. Effects of Metallicity and Grain Size on Gravitational Instabilities in Protoplanetary Disks. *Ap. J.* 636, L149-152 (2006).
15. R. Canup. A Giant Impact Origin of Pluto-Charon. *Science* 307, 546-550 (2005).
16. J. Carvano, T. Mothe-Diniz and D. Lazzaro. Search for Relations Among a Sample of 460 Asteroids with Featureless Spectra. *Icarus* 161, 356-382.
17. S. Chandrasekhar. *Figures of Equilibrium* (New York: Dover).
18. J. Chambers. Making More Terrestrial Planets. *Icarus* 152, 205-224 (2001).
19. E. Chiang and Y. Lithwick. Neptune Trojans as a Test Bed for Planet Formation. *Ap. J.*, 628, 520-532 (2005).
20. F. Ciesla and J. Cuzzi. The Evolution of the Water Distribution in a Viscous Protoplanetary Disk. *Icarus* 181, 178-204 (2006).
21. R. Clark et al. Compositional maps of Saturn's moon Phoebe from imaging spectroscopy. *Nature* 435, 66-69 (2005).
22. G. Colombo and F. Frankin. On the Formation of the Outer Satellite Groups of Jupiter. *Icarus* 15, 186-189 (1971).
23. J. Cooper, E. Christian, J. Richardson and C. Wang. Proton Irradiation of Centaur, Kuiper Belt and Oort Cloud Objects. *Earth, Moon and Planets* 92, 261-177 (2003).
24. D. Cruikshank et al. Constraints on the Composition of Trojan Asteroid 624 Hektor. *Icarus* 153, 348-360 (2001).

25. D. Cruikshank, M. Barucci, J. Emery, Y. Fernandez, W. Grundy, K. Noll and J. Stansberry. Physical Properties of Trans-Neptunian Objects. *Protostars and Planets V* (eds. B. Reipurth, D. Jewitt and K. Keil), Univ. Az. Press, Tucson, p. 879-892. (2006).
26. J. Davies et al. The Detection of Water Ice in Comet Hale-Bopp. *Icarus* 127, 238-245 (1997).
27. J. Davies et al. Visible and Infrared Photometry of Fourteen KBOs. *Icarus* 146, 252-262 (2000).
28. B. Daviddson and P. Gutierrez. Estimating the Nucleus Density of 19P/Borrelly. *Icarus* 168, 392-408 (2004).
29. B. Daviddson and P. Gutierrez. Nucleus Properties of 67P/Churyumov Gerasimenko. *Icarus* 176, 453-477 (2005).
30. B. Daviddson and P. Gutierrez. Non-gravitational force modeling of Comet 81P/Wild 2. *Icarus* 180, 224-242 (2006).
31. A. Delsanti, O. Hainaut, E. Jourdeuil, K. Meech, H. Boehnhardt and L. Barrera. Simultaneous visible-near IR Photometric Study of Kuiper Belt Objects. *Astron. Ap.* 417, 1145-1158 (2004).
32. A. Delsanti and D. Jewitt. The Solar System beyond the Planets. In *Solar System Update* (eds. Ph. Blondel and J. Mason), Springer-Praxis, Berlin, p. 267-294 (2006).
33. Delsanti, A., Peixinho, N., Boehnhardt, H., Barucci, A., Merlin, F., Doressoundiram, A., and Davies, J. K. *Astron. J.* 131, 1851-1863 (2006).
34. J. Dohnanyi. Collisional Models of Asteroids and Their Debris. *J. Geophys. Res.* 74, 2531-2554 (1969).
35. A. Doressoundiram. Color Properties and Trends in Trans-Neptunian Objects. *Earth, Moon and Planets* 92, 131-144 (2003).
36. A. Doressoundiram et al. Meudon Multicolor Survey (2MS) of Centaurs and Trans-Neptunian Objects. *Icarus* 174, 90-104 (2005).
37. C. Dumas, T. Owen and M. Barucci. Near-Infrared Spectroscopy of Low-Albedo Surfaces in the Solar System. *Icarus* 133, 221-232 (1998).
38. M. Duncan, Quinn and S. Tremaine. The Origin of Short-Period Comets. *Ap. J.* 328, L69-73 (1988).
39. W. Durham, H. Heard and S. Kirby. Experimental Deformation of Polycrystalline H₂O Ice at High Pressure and Low Temperature. *J. geo. res.* 88, B377-392 (1983).
40. K. Edgeworth. *J. Br. Astron. Soc.*, 53, 181-188 (1943).
41. C. Emerich et al. Temperature and Size of the Nucleus of P/Halley Deduced from IKS Infrared Vega 1 Measurements. *Astron. Ap.*, 187, 839-842 (1987).
42. J. Emery and R. Brown. Constraints on the Surface Composition of Trojan Asteroids from Near Infrared (0.8-4.0 μ m) Spectroscopy. *Icarus* 164, 104-121 (2003).
43. E. Everhart. Intrinsic distributions of cometary perihelia and magnitudes. *Astron. J.*, 72, p. 1002 (1967).
44. P. Farinella and D. Davis. Short-period comets: Primordial bodies or collisional fragments? *Science* 273, 938-941 (1996).
45. J. Fernandez. On the Existence of a Comet Belt Beyond Neptune. *MNRAS* 192, 481-491 (1980).
46. J. Fernandez, G. Tancredi, H. Rickman and J. Licandro. The population, magnitudes, and sizes of Jupiter family comets. *Astron. Ap.*, 352, 327-340 (1999).

47. J. Fernandez, T. Gallardo and A. Brunini. *Icarus* 159, 358 (2002).
48. Y. Fernandez, S. Sheppard and D. Jewitt. Albedo Distribution of Jovian Trojan Asteroids. *Astron. J.* 126, 1563-1574 (2003).
49. H. Fleming and D. Hamilton. On the Origin of the Trojan Asteroids: Effects of Jupiter's Mass Accretion and Radial Migration. *Icarus*, 148, 479-493 (2000).
50. Y. Funato, J. Makino, P. Hut, E. Kokubo and D. Kinoshita. The Formation of Kuiper Belt Binaries Through Exchange Reactions. *Nature* 427, 518-520 (2004).
51. P. Francis. The Demographics of Long-Period Comets. *Ap. J.*, 635, 1348-1361 (2005).
52. P. Goldreich, Y. Lithwick and R. Sari. Formation of Kuiper Belt Binaries by Dynamical Friction and Three-Body Encounters. *Nature* 420, 643-646 (2002).
53. R. Gomes. The Origin of the Kuiper Belt High Inclination Population. *Icarus* 161, 404-418 (2003).
54. T. Grav, M. Holman, B. Gladman and K. Aksnes. Photometric Survey of the Irregular Satellites. *Icarus* 166, 33-45 (2003).
55. T. Grav, M. Holman and W. Fraser. Photometry of Irregular Satellites of Uranus and Neptune. *Ap. J.* 613, L77-80 (2004).
56. N. Haghighipour and A. Boss. On Gas Drag-Induced Rapid Migration of Solids in a Nonuniform Solar Nebula. *Ap. J.* 598, 1301-1311 (2003).
57. O. Hainaut and A. Delsanti. Colors of Minor Bodies in the Outer Solar System: A Statistical Analysis. *Astron. Ap.* 389, 641-664 (2002).
58. J. Heisler and S. Tremaine. The influence of the galactic tidal field on the Oort comet cloud. *Icarus*, 65, 13-26 (1986).
59. T. Heppenheimer and C. Porco. New Contributions to the Problem of Capture. *Icarus*, 30, 385-401 (1977).
60. T. Hiroi, M. Zolensky and C. Pieters. The Tagish Lake Meteorite: Possible Sample from a D-Type Asteroid. *Science* 293, 2234-2236 (2001).
61. J. Horner, N. Evans, M. Bailey and D. Asher. The Populations of Comet-Like Bodies in the Solar System. *MNRAS* 343, 1057-1066 (2003).
62. H. Hsieh, D. Jewitt and Y. Fernandez. The Strange Case of 133P/Elst-Pizarro. *Astron. J.* 127, 2997-3017 (2004).
63. H. Hsieh and D. Jewitt. A Population of Comets in the Main Asteroid Belt. *Science*, 312, 561-563 (2006).
64. D. Hughes. The magnitude distribution, perihelion distribution and flux of long-period comets. *MNRAS*, 326, 515-523 (2001).
65. D. Hunten and A. Watson. Stability of Pluto's Atmosphere. *Icarus* 51, 665-667 (1982).
66. S. Ida, J. Larwood and A. Burkert. Evidence for Early Stellar Encounters in the Orbital Distribution of Edgeworth-Kuiper Belt Objects. *Ap. J.* **528**, 351-356 (2000).
67. S. Jacobsen. The Hf-W Isotopic System and the Origin of the Earth and Moon. *Ann. Rev. Earth and Planet Sci.* 33, 531-570 (2005).
68. D. Jewitt and K. Meech. Optical Properties of Cometary Nuclei and a Preliminary Comparison with Asteroids. *Ap. J.* 328, 974-986 (1988).
69. D. Jewitt and J. Luu. CCD Spectra of Asteroids II. The Trojans as Spectral Analogs of the Cometary Nuclei. *Astron. J.* 100, 933-944 (1990).
70. D. Jewitt, C. Trujillo and J. Luu. Population and Size Distribution of Small Jovian Trojans. *Astron. J.* 120, 1140-1147 (2000).

71. D. Jewitt. From Kuiper Belt Object to Cometary Nucleus: The Missing Ultra-Red Matter. *Astron. J.*, 123, 1039-1049 (2002).
72. D. Jewitt and S. Sheppard. Physical Properties of Trans-Neptunian Object (20000) Varuna. *Astron. J.* 123, 2110-2120 (2002).
73. D. Jewitt and J. Luu. Discovery of the Candidate Kuiper Belt Object 1992 QB1. *Nature*, 362, 730-732 (1992).
74. D. Jewitt and J. Luu. Optical-Infrared Spectral Diversity in the Kuiper Belt. *Astron. J.* 115, 1667-1670 (1998).
75. D. Jewitt and J. Luu. Colors and Spectra of Kuiper Belt Objects. *Astron. J.*, 122, 2099-2114 (2001).
76. D. Jewitt and J. Luu. Crystalline Water Ice on Kuiper Belt Object (50000) Quaoar. *Nature* 432, 731-733 (2004).
77. D. Jewitt, S. Sheppard and C. Porco. In *Jupiter* (eds. F. Bagenal et al), Cambridge Univ. Press. (2004).
78. D. Jewitt and S. Sheppard. Irregular Satellites in the Context of Giant Planet Formation. *Space Sci. Rev.* 116, 441-456 (2005).
79. D. Jewitt. A First Look at the Damocloids. *Astron. J.*, 129, 530-538 (2005).
80. D. Jewitt. From the Cradle to the Grave: The Rise and Demise of the Comets. In *Comets II* (eds. M. Festou, H. Weaver and H. U. Keller), Univ. Az. Press, Tucson, pp. 659-676 (2005).
81. T. Johnson and J. Lunine. Saturn's Moon Phoebe as a Captured Body from the Outer Solar System. *Nature*, 435, 69-71 (2005).
82. T. Jones, L. Lebofsky, J. Lewis and M. Marley. The Composition and Origin of the C-, P- and D-Asteroids. *Icarus* 88, 172-192 (1990).
83. H. Kawakita et al. Evidence of Icy Grains in Comet C/2002 T7 (LINEAR at 3.52 AU. *Ap. J.* 601, L191-194 (2004).
84. K. Keil. Thermal Alteration of Asteroids: Evidence from Meteorites. *Planet. Space Sci.* 48, 887-903 (2000).
85. G. Kuiper. On the Origin of the Solar System, in *Astrophysics*, (ed. J. Hynek), McGraw-Hill, New-York, pp. 357-424 (1951).
86. P. Lacerda and J. Luu. Analysis of the Rotational Properties of Kuiper Belt Objects. *Astron. J.* 131, 2314-2326. (2006).
87. P. Lamy, E. Grun and J. Perrin. *Astron. Ap.*, 187, 767.
88. P. Lamy, I. Toth, Y. Fernandez and H. Weaver. In *Comets II*, (eds. M. Festou, H. Weaver and H. Keller), Univ. Az. Press, Tucson, pp. 223-264 (2005).
89. L. Lebofsky et al. A Refined Standard Model for Asteroids Based on Observations of 1 Ceres and 2 Pallas. *Icarus* 68, 239-251 (1986).
90. T. Lee, F. Shu, Shang, H., Glassgold, A. and Rehm, K. *Ap. J.*, **506**, 898-912 (1976)
91. T. Lee, D. Papanastassiou and G. Wasserburg. *Geophys. Res. Lett.* **3**, 41-44 (1998)
92. F. Leonard. The New Planet Pluto. *Leaflet Astron. Soc. Pac.* No 30, Aug, 121-124 (1930).
93. G. Leone, P. Farinella, P. Paolicchi and V. Zappala. Equilibrium Models of Binary Asteroids. *Astron. Ap.* 140, 265-272 (1984).
94. H. Levison et al. The Mass Disruption of Oort Cloud Comets. *Science*, 296, 2212-2215 (2002).
95. H. Levison, D. Terrell, P. Wiegert, L. Dones and M. Duncan. On the Origin of the Unusual Orbit of Comets 2P/Encke. *Icarus*, 182, 161-168 (2006).

96. H. Levison and M. Duncan. Long Term Dynamical Behavior of Short-Period Comets. *Icarus*, 108, 18-36 (1994).
97. H. Levison, G. Shoemaker and C. Shoemaker. The Dispersal of the Trojan Asteroid Swarm. *Nature* 385, 42-44 (1997).
98. C. Lisse et al. The Nucleus of Comet Hyakutake (C/1996 B2). *Icarus* 140, 189-204 (1999).
99. J. Licandro, N. Pinilla-Alona, M. Pedani, E. Oliva, G. Tozzi and W. Grundy. Methane Ice Rich Surface of Large TNO 2005 FY9. *Astron. Ap.* 445, L35-38 (2006).
100. M. Lupo and J. Lewis. Mass-Radius Relationships in Icy Satellites. *Icarus* 40, 157-170 (1979).
101. J. Luu, D. Jewitt and E. Cloutis. Near-Infrared Spectroscopy of Primitive Solar System Objects. *Icarus* 109, 133-144 (1994).
102. J. Luu and D. Jewitt. Color Diversity Among the Centaurs and Kuiper Belt Objects. *Astron. J.* 112, 2310- (1996).
103. F. Marzari, P. Farinella and V. Vanzani. Are Trojan Collisional Families a Source for Short-Period Comets? *Astron. Ap.* 299, 267 (1995).
104. F. Marzari, P. Tricarico and H. Scholl. Clues to the Origin of Jupiter's Trojans: The Libration Amplitude Distribution. *Icarus* 162, 453-459 (2003).
105. F. Marchis et al. *Nature* 439, 565-567 (2006).
106. N. McBride et al. (2003). Visible and Infrared Photometry of Kuiper Belt Objects: Searching for Evidence of Trends. *Icarus* 161, 501-510 (2003).
107. W. McKinnon. On the Initial Thermal Evolution of Kuiper Belt Objects. *Proc. Asteroids, Comets, Meteor, ESA SP-500, Noordwijk, Netherlands* pp. 29-38 (2002).
108. K. Meech, O. Hainaut and B. Marsden. Comet nucleus size distributions from HST and Keck telescopes. *Icarus* 170, 463-491 (2004).
109. R. Meier and T. Owen. Cometary Deuterium. *Space Sci. Rev.* 90, 33-43 (1999).
110. A. Milani. Trojan Asteroids Belt: Proper Elements, Stability, Chaos and Families. *Cel. Mech. Dyn. Astron.* 57, 59-94 (1993).
111. H. Mizuno, K. Nakazawa and C. Hayashi. Instability of a Gaseous Envelope Surrounding a Planetary Core and Formation of Giant Planets. *Prog. Theor. Physics* 60, 699-710 (1978).
112. A. Morbidelli et al. Source Regions and Time Scales for the Delivery of Water to Earth. *Meteoritics and Planet. Sci* 35, 1309-1320 (2000).
113. A. Morbidelli, and H. Levison. Scenarios for the Origin of the Orbits of the Trans-Neptunian Objects 2000 CR105 and 2003 VB12 (Sedna). *Astron. J.*, **128**, 2564-2576 (2004).
114. A. Morbidelli, H. Levison, K. Tsiganis and R. Gomes. Chaotic Capture of Jupiter's Trojan Asteroids in the Early Solar System. *Nature* 435, 462-465 (2005).
115. S. Mostefaoui, G. Lugmair, P. Hoppe and A. El-Goresy. *New Astronomy Review*, **48**, 155-159 (2004).
116. D. Nesvorny, J. Alvarellos, L. Dones and H. Levison. Orbital and Collisional Evolution of the Irregular Satellites. *Astron. J.*, 126, 398-429 (2003).
117. K. Noll. Solar System Binaries. *Proc. Asteroids, Comets, Meteors 2005* (eds. S. Ferraz-Mello and D. Lazzaro), IAU Symp. 229, Cambridge University Press, Cambridge (2006).

118. J. Oort. The structure of the cloud of comets surrounding the Solar System and a hypothesis concerning its origin. *Bull. Astron. Inst. Neth.*, 11, 91-110 (1950).
119. T. Owen, P. Mahaffy, H. Niemann, S. Atreya, T. donahue, A. Bar-Nun and I. de pater. A Low Temperature Origin for the Planetesimals that Formed Jupiter. *Nature* 402, 269-270 (1999).
120. N. Peixinho, A. Dorresoundiram, A. Delsanti, H. Boehnhardt, M. Barucci and I. Belskaya. Reopening the TNOs Color Controversy: Centaurs Bimodality and TNOs Unimodality. *Astron. Ap.* 410, L29-32 (2003).
121. J. Petit and O. Mousis. KBO Binaries: How Numerous Were They? *Icarus* 168, 409-419 (2004).
122. J. Pollack, O. Hubickyj, P. Bodenheimer, J. Lissauer, M. Podolak and Y. Greenzweig. Formation of the Giant Planets by Concurrent Accretion of Solids and Gas. *Icarus* 124, 62-85 (1996).
123. K. R. Popper. *The Logic of Scientific Discovery*. Hutchinson and Co., London.
124. R. Prinn and B. Fegley. Kinetic Inhibition of CO and N₂ Reduction in Circumplanetary Nebulae. *Ap. J.* 249, 308-317 (1981).
125. D. Prialnik. In *COMETS II* (eds. M. Festou, H. Weaver and H. U. Keller), Univ. Az. Press, Tucson, pp. 359-387 (2005).
126. D. Rabinowitz et al. Photometric Observations of 2003 EL61. *Ap. J.* 639, 1238-1251 (2006).
127. S. Raymond, T. Quinn and J. Lunine. Making Other Earths: Dynamical Simulations of Terrestrial Planet Formation and Water Delivery. *Icarus* 168, 1-17 (2004).
128. H. Rickman, J. Fernandez and B. Gustafson. *Astron. Ap.*, 237, 524 (1990).
129. G. Rieke et al. Decay of Planetary Debris Disks. *Ap. J.*, 620, 1010-1026 (2005).
130. S. Sheppard et al. A Wide Field CCD Survey for Centaurs and Kuiper Belt Objects. *Astron. J.* 120, 2687-2694 (2000).
131. S. Sheppard and D. Jewitt. An Abundant Population of Irregular Satellites Around Jupiter. *Nature* 413m 261-263 (2003).
132. S. Sheppard and D. Jewitt. Extreme Kuiper Belt Objects 2001 QG298 and the Fraction of Contact Binaries. *Astron. J.*, 127, 3023-3033 (2004).
133. S. Sheppard, D. Jewitt and J. Kleyna. Ultradeep Survey for Irregular Satellites of Uranus: Limits to Completeness. *Astron. J.*, 129, 518-525 (2005).
134. G. Shoemaker, C. Shoemaker and R. Wolfe. In *ASTEROIDS II* (eds. T. Gehrels and M. Matthews), Univ. Az. Press, pp. 487-523 (1989).
135. J. Stansberry et al. Albedos, Diameters and Density of Kuiper Belt Object (47171) 1999 TC36. *Ap. J.* 643, 556-566 (2006).
136. D. Stephens and K. Noll. HST Photometry of Trans-Neptunian Objects. *Earth, Moon and Planets* 92, 251-260 (2004).
137. D. Stephens and K. Noll. Detection of Six Trans-Neptunian Binaries with NICMOS. *Astron. J.*, 131, 1142-1148 (2006).
138. J. Sunshine et al. Exposed Water Ice Deposits on the Surface of Comet Tempel 1. *Science* 311, 1453-1455 (2006).
139. S. Takahashi and W. Ip. A Shape and Density Model of the Putative Binary EKBO 2001 QG298. *PASJ* 56, 1099-1103 (2004).
140. N. Takato, S. Bus, H. Terada, T-S. Pyo, and N. Kobayashi. Detection of a Deep 3 μ m Absorption Feature in the Spectrum of Amalthea. *Science* 306, 2224-2227 (2004).

141. N. Takato, H. Terada and T-S. Poo. Crystalline Water Ice on the Satellite of 2003 EL61. Preprint (2006).
142. G. Tancredi, J. Fernandez, H. Rickman and J. Licandro. Nuclear magnitudes and the Size Distribution of Jupiter Family Comets. *Icarus*, 182, 527-549 (2006).
143. S. Tegler and W. Romanishin. Two distinct populations of Kuiper-belt objects. *Icarus*, 392, 49-51 (1998).
144. S. Tegler and W. Romanishin. Extremely red Kuiper-belt objects in near-circular orbits beyond 40 AU. *Nature* 407, 979-981 (2000).
145. S. Tegler and W. Romanishin. Resolution of the kuiper belt object color controversy: two distinct color populations. *Icarus* 161, 181-191 (2003).
146. S. Tegler, W. Romanishin, and S. Consolmagno. Color Patterns in the Kuiper Belt: A Possible Primordial Origin. *Ap. J.* 599, L49-52 (2003).
147. E. Thommes, M. Duncan and H. Levison. The Formation of Uranus and Neptune Between Jupiter and Saturn. *Astron. J.*, 123, 2862-2883 (2002).
148. I. Toth. Impact-Generated Activity Period of the Asteroid 7968 Elst-Pizarro in 1996. *Astron. Ap.* 360, 375-380 (2000).
149. C. Trujillo, D. Jewitt and J. Luu. Properties of the Trans-Neptunian Belt: Statistics from the Canada-France-Hawaii Telescope Survey. *Astron. J.*, 122, 457-473 (2001).
150. C. Trujillo and M. Brown. A Correlation between Inclination and Color in the Classical Kuiper Belt. *Ap. J.* 566, L125-128 (2002).
151. C. Trujillo, M. Brown, D. Rabinowitz and T. Geballe. Near-Infrared Surface Properties of the Two Intrinsically Brightest Minor Planets. *Ap. J.*, 627, 1057-1065 (2005).
152. K. Tryka, R. Brown and V. Anicich. Near-Infrared Absorption Coefficients of Solid Nitrogen as a Function of Temperature. *Icarus* 116, 409-414 (1995).
153. W. Ward and J. Hahn. Dynamics of the Trans-Neptune Region: Apsidal Waves in the Kuiper Belt. *Astron. J.*, 116, 489-498 (1998).
154. S. Weidenschilling. On the Origin of Binary TransNeptunian Objects. *Icarus* 160, 212-215 (2002).
155. P. Weissman and S. Lowry. The Size Distribution of Jupiter-Family Cometary Nuclei. *Lunar Planet. Sci. Conf. Abstract No.* 2003. (2003).
156. P. Wiegert and S. Tremaine. The Evolution of Long-Period Comets. *Icarus*, 137, 84-121 (1999).
157. D. Yeomans et al. Estimating the Mass of Asteroid 253 Mathilde from Tracking Data During the NEAR Flyby. *Science* 287, 2106-2109 (1997).
158. C. Yoder. Notes on the Origin of the Trojan Asteroids. *Icarus* 40, 341-344 (1979).
159. B. Zuckerman, T. Forveille and J. Kastner. Inhibition of Giant-Planet Formation by Rapid Gas Depletion Around Young Stars. *Nature* 373, 494-496 (1995).