Astronomers' Universe

Michael Moltenbrey

Dawn of Small Worlds

Dwarf Planets, Asteroids, Comets

Springer

Astronomers' Universe

More information about this series at http://www.springer.com/series/6960

Michael Moltenbrey

Dawn of Small Worlds

Dwarf Planets, Asteroids, Comets

Springer

Michael Moltenbrey
Munich
Germany

ISSN 1614-659X ISSN 2197-6651 (electronic)
Astronomers' Universe
ISBN 978-3-319-23002-3 ISBN 978-3-319-23003-0 (eBook)
DOI 10.1007/978-3-319-23003-0

Library of Congress Control Number: 2015954813

Springer Cham Heidelberg New York Dordrecht London

Cover illustration: Artist's concept of an asteroid belt around Vega. Credit: NASA/JPL-Caltech

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Contents

1 **Introduction** 1

2 **Asteroids** 17

3 **Comets** 73

4 **Trans-Neptunian Objects** 141

5 **Dwarf Planets** 175

6 **Exploration of Small Solar System Bodies** 215

Further Reading 269

Index 271

Chapter 1
Introduction

Almost since the beginning of time, human beings have been fascinated by the starry sky above their heads. What were these little points of light up there? Not having names for them yet, people wondered about the moon and the Sun. Evidence for this fascination traces well back into the earliest stages of civilization, as e.g. Cave paintings from the Stone Age suggest. The paintings of Lascaux in France are a very famous example for this. They were created between 17,000 and 15,000 BC and epic the Pleiades, the zodiac and probably also the summer sky.

The observation of the starry sky had been limited to the naked eye for many thousands of years. The people identified the Sun, the Moon, and some of the planets (Mercury to Saturn). The outermost gas giants, Uranus and Neptune, remained concealed to them. From time to time, an omen appeared in the sky—a comet! The view on the world out there only dramatically changed with the advent of optical telescopes. When Galileo Galilei (1564–1642) pointed his scope at Jupiter, he was one of the first to realize that there was more. It was probably the first time that a human being was able to see what a planet actually is. Before this event, they were often merely considered to be what their name suggests: Wandering stars. Their visual appearance was not different to any of the other stars except for their movements. Yes! They actually moved! They appeared to move relative to the background stars. What were they?

Galilei was surprised to see a disk with structures on it when looking at the planet. This was something he had not expected. Yet, he saw even more. There were four bright dots near Jupiter. They appeared to change positions over the course of several observing sessions. He correctly interpreted them as moons of Jupiter. To honor their discoverer they are named the four Galilean Moons today.

It was the first time that new bodies in our solar system were found. The planets Uranus and Neptune followed. Nevertheless, a general notion developed that our solar system merely consisted of the Sun, the planets and their moons and a few comets (see Fig. 1.1).

Scientists in the eighteenth and nineteenth century, however, were wondering when looking at the distribution of the planets within the solar system. A planet

M. Moltenbrey, *Dawn of Small Worlds*, Astronomers' Universe,
DOI 10.1007/978-3-319-23003-0_1

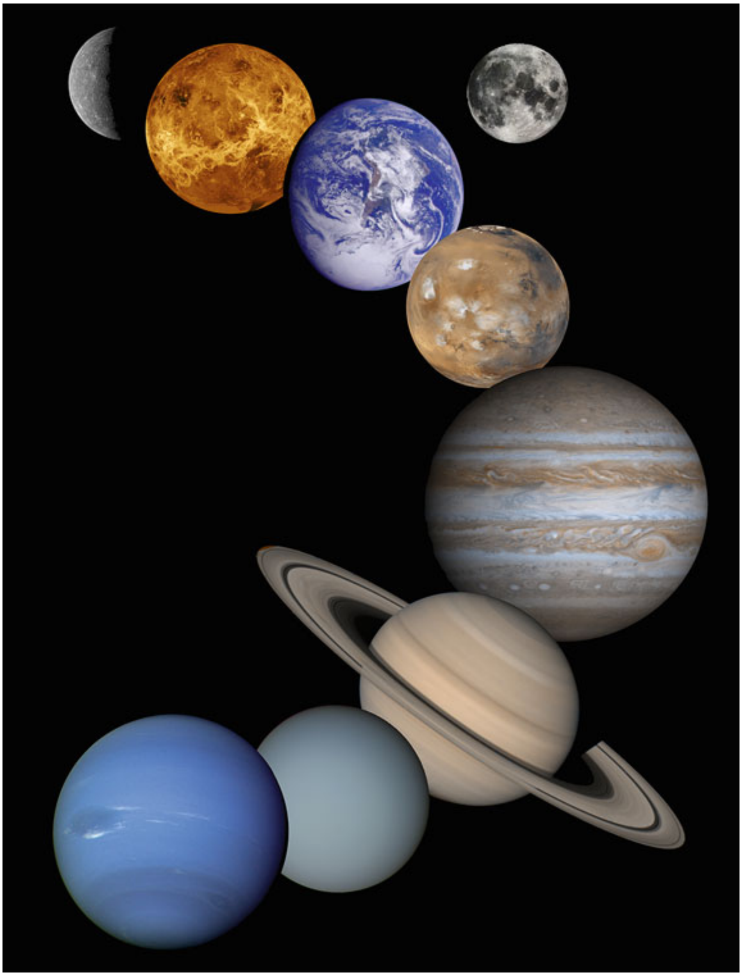


Fig. 1.1 Montage of planetary images taken by spacecraft managed by the Jet Propulsion Laboratory in Pasadena, CA. Included are (from *top* to *bottom*) images of Mercury, Venus, Earth (and Moon), Mars, Jupiter, Saturn, Uranus and Neptune (credit: NASA/JPL)

appeared to be missing. By then, thanks to Johannes Kepler (1571–1630) and Sir Isaac Newton (1642–1726), it had been possible to determine the orbits of objects in our solar system. Yet, there was an unexplainable gap between the orbits of Mars and Jupiter. A, by then, famous law, the so-called Titius-Bode law, postulated that a planet should exist there.

In 1800, Baron Franz Xaver von Zach (1754–1834) and Johann Hieronymus Schroeter (1747–1816) founded the probably first international research project, the "Himmelspolizey" (engl. the "sky police") while working at the observatory in Gotha/Germany. The group divided the starry sky into 24 distinct regions and each member, located in different European countries, was asked to systematically search his region for the missing planet.

The “missing planet” was found by chance by the Italian astronomer and theologian Giuseppe Piazzi (1746–1826), who had not been a member of the “Himmelspolizey”, during the night of New year’s day 1801 (January 1, 1801) while working at the observatory of Palermo/Italy. Soon after, more and more planets were discovered in that region. All of them appeared to be small and faint. Even in the largest telescopes of that time, it was not possible to see more than a faint star-like dot. Were these really planets? They were so different from the other planets that had already been known by that time. All of them showed details of their respective surfaces. Yet, the new ones remained completely indistinct.

The astronomers came to the conclusion that these objects had to be a new group of solar system bodies. The German-British astronomer William Herschel (1738–1822) coined the term “asteroid”, meaning star-like, for them. Soon after, the discovered asteroids clearly outnumbered any of the other known solar system bodies. Today, more than 600,000 asteroids in the main asteroid belt between Mars and Jupiter are known. More exist at other locations in the solar system.

The exploration of the solar system continued. It was found that another larger body somehow perturbed Uranus’ orbit. Mathematical predictions were made of the orbit of the potential new planet. Indeed, a planet was found there. The discovery of Neptune in 1846, however, did not solve all the problems but even triggered a new search for another planet. Astronomers speculated that Uranus’s orbit was being disturbed by another planet besides Neptune. In 1930, a young astronomer, Clyde Tombaugh (1906–1997), working at the Lowell Observatory in Flagstaff (Arizona/USA) discovered Pluto which was then subsequently classified as the ninth planet in our solar system. An intensive debate began about Pluto and its status in our solar system. Observations subsequent to its discovery revealed that Pluto was by far smaller than any of the other planets. It was not possible to resolve any planetary disc of Pluto even with the largest telescopes available. Could this tiny object be really responsible for the perturbations of Uranus’ and Neptune’s orbits?

During the course of the following decades, Pluto’s size and mass was continuously reduced from initially about the mass of Earth down to approximately 1/500 of Earth’s mass. Was this a planet? The debate continued and intensified after the discovery of other small bodies in the region beyond Neptune in the 1990s. The first to be discovered was a object called 1992 QB1 discovered by the American astronomers David Jewitt and Jane Luu in 1992. In the following years more and more of these objects were found in the trans-Neptunian region.

Some of these objects were so large and similar to Pluto as to question its planetary status. Consequently, in 2006, the International Astronomical Union (IAU) introduced the new class of dwarf planets and demoted Pluto into it.

New populations of new types of objects popped up everywhere in the solar system. Up to then, astronomers had only distinguished a few classes of solar system bodies: the Sun, the planets, the moons, asteroids and comets. Yet, some of the newly discovered objects, such as the centaurs, could not be clearly classified into any of these categories as they showed, for example, properties of comets and asteroids together. Furthermore, some bodies that had been considered as asteroids turned out to be extinct comets, i.e., comets that do not exhibit any activity anymore. The number of newly discovered objects increased and with them the chaos in classification.

In 2006 together with the definition of a planet, the IAU therefore decided to put them together in a new group of objects, the so-called Small Solar System Bodies (SSSB). This group comprises all solar system bodies except the planets, dwarf planets and planetary satellites or moons. These currently include most of the Solar System asteroids, most Trans-Neptunian Objects (TNOs), comets, and other small bodies even including interstellar dust.

This book will provide an overview of these small bodies and dwarf planets that are present in our solar system. We will cover asteroids, comets, and trans-Neptunian objects. In addition, we will also deal with dwarf planets and their most prominent representative, Pluto.

1.1 Orbits and Resonances: A Brief Introduction

Before starting our journey to the small bodies in our solar system, we need to introduce some terms and concepts that will be necessary for the further understanding. Those readers, who are already familiar with these concepts can skip this section and directly go to the following dealing with the formation of our solar system.

1.1.1 What Are Orbits?

All the objects in our solar system move along orbits around the Sun. Figure 1.2 shows the basic structure of our solar system, including the orbits of the planets Mercury, Venus, Earth, Mars, Jupiter, Saturn, Neptune. Further shown are the asteroid belt between the orbits of Mars and Jupiter, the Kuiper belt beyond Neptune and the dwarf planet Pluto.

The orbits of the planets are not exact circles but so-called ellipses. You can think of ellipses as some kind of circle that has been stretched to the one or the other direction. The extent to which they are stretched is called the eccentricity e. If we do not stretch the circle at all we have no eccentricity ($e = 0$), the more we stretch, the more elongated the circle becomes. Eccentricity e is increasing in the latter case.

Ellipses are an important concept we have to be familiar with. Most of the planets have near circular orbits. However, this is not true for many of the small bodies. They often have highly elliptical orbits, i.e. extremely elongated circles.

Figure 1.3 depicts the basic structure of an ellipse. Each ellipse has to focuses. In the case of our solar system, the Sun is at one of the focuses. The distance between the focuses and the center gives an indication on how much the ellipse is stretched.

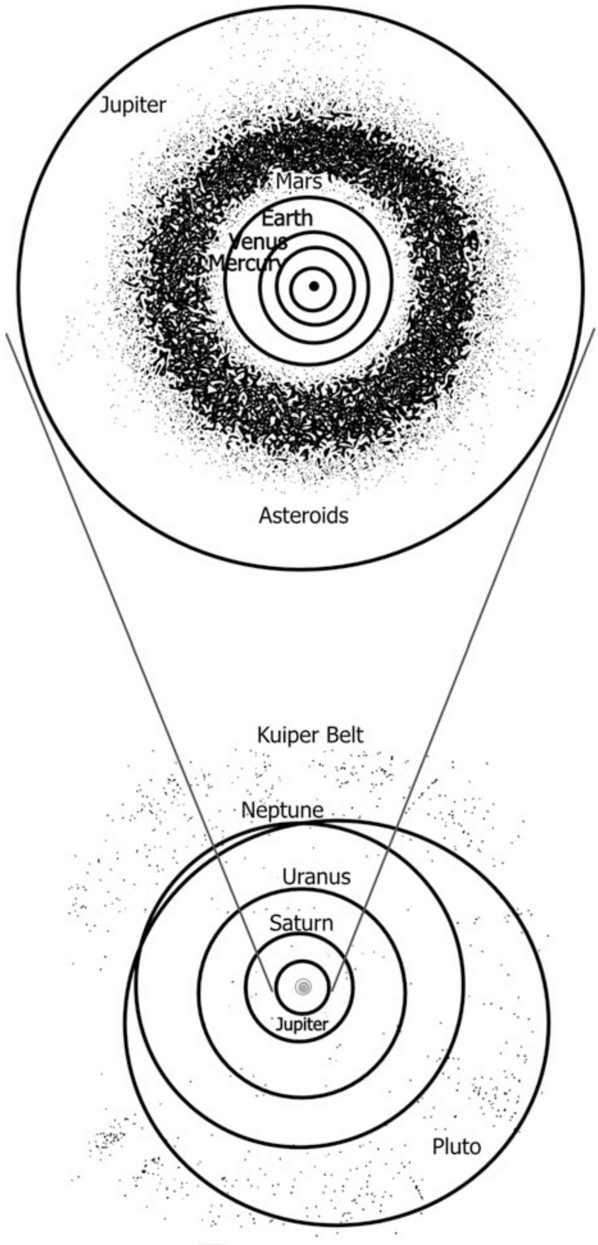


Fig 1.2 Basic structure of our solar system including the planets, the asteroid belt and the Kuiper belt

If we have a circle, i.e. an eccentricity of 0, the two focuses will fall together at the center. In particular two parameters will be important for the following chapters. On the one hand, this is the so-called semi-major axis a which is size of the orbit and the eccentricity e which defines how much the ellipse deviates from a perfect circle.

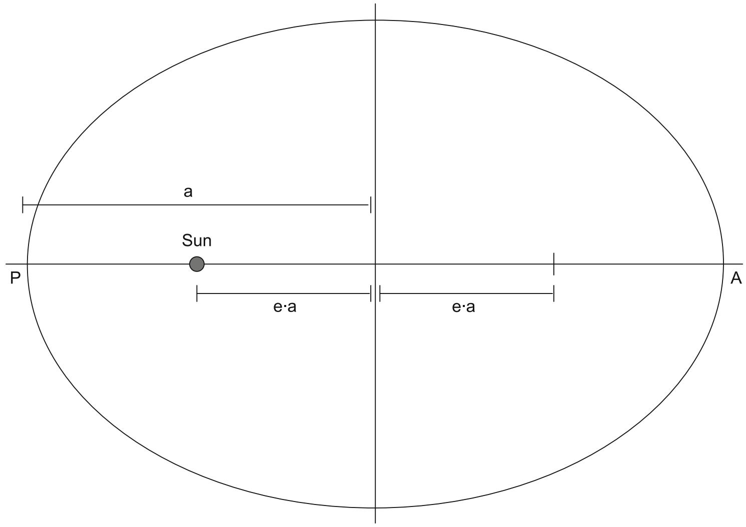


Fig. 1.3 Basic structure of an ellipse with a semi-major axis a. The Sun is in one of the focuses. The letters P and A stand for perihelion and aphelion

The average distance from Earth to Sun is roughly equivalent to the semi-major axis of Earth's orbit. A further important characteristic of an orbit is its inclination i towards the ecliptic plane. So what is the ecliptic plane? The orbit of Earth around the Sun defines a plane, which is termed the ecliptic. Most of the planets have their orbits more or less also in this plane. The value by which they deviate is called the orbit's inclination.

1.1.2 Orbital Resonances

Another important concept for the following chapters are the so-called orbital resonances. An orbital resonance occurs when two orbiting bodies exert a regular, periodic gravitational influence on each other. This is the case when their orbital periods are related by a ratio of two integers. Orbital resonances greatly enhance the mutual gravitational influence of the bodies, i.e., their ability to alter or constrain each other's orbits. In most cases, the results are unstable orbits due to orbital perturbation caused by the gravitational interaction of the two bodies. Under some circumstances, a resonant system can be stable and self-correcting, so that the bodies (permanently) remain in resonance. A prominent example of such stabilizing resonances is the 2:3 resonance between Pluto and Neptune. In this resonance Pluto will make two orbits for every three orbits Neptune makes. We will cover this resonance later on in more detail.

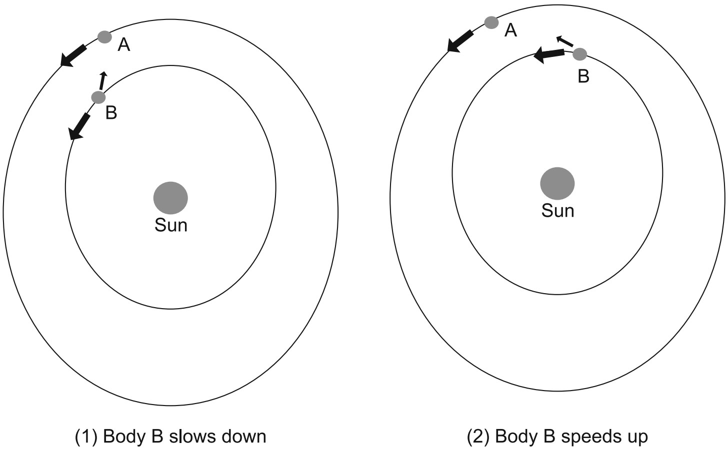


Fig. 1.4 An example on the effects of orbital resonances of two objects

Many destabilizing resonances can be found in the main asteroid belt between Mars and Jupiter. The asteroids there are temporarily locked in resonance with Jupiter. The gas giant with its enormous mass causes strong gravitational interaction with the asteroids. Their orbits are perturbed to a smaller or larger extent.

Figure 1.4 depicts the possible effects of an orbital resonance. In situation (1), object B is slowed down as it is "pulled back" by the gravitational interaction with A. In situation (2), B is accelerated as it is pushed forward by the interaction with A.

1.2 Formation of Our Solar System

Before we continue our journey to the small solar system bodies, we first need to have a look at how our solar system formed. This is crucial for the understanding of how the small bodies formed and why they are the way we can observe them today. Furthermore, only with this background it is possible to understand their development, composition and orbits. We will go through the basic principles of the formation of our solar system and give an overview on the processes that took place. Further details will be provided in the following chapters, e.g., on the formation of the asteroid belt or the Kuiper belt.

Today, the most widely accepted model describing the formation of our solar system is the so-called nebular hypothesis. It was originally proposed by Pierre-Simon Laplace (1740–1827), Emanuel Swedenborg (1688–1772) and Immanuel Kant (1724–1804) in the eighteenth century. Since then the basic model has undergone several more or less severe refinements but is in its basic assumptions still considered to be valid.

So, how did our solar system form? All began about 4.6 billion years ago with a giant molecular cloud spanning across about 65 lightyears. Nowadays, we know

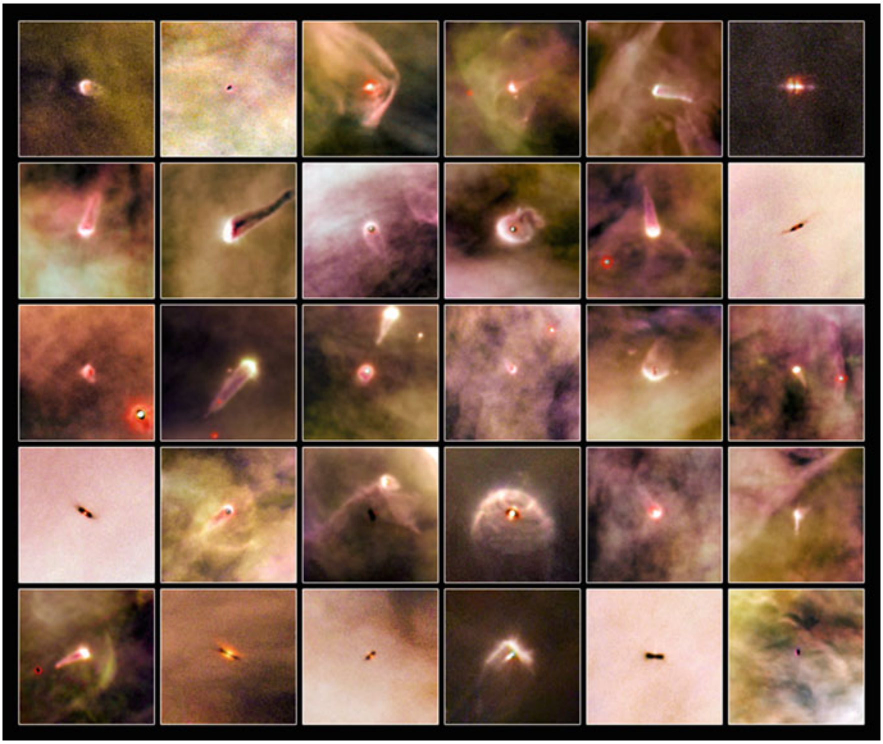


Fig. 1.5 An overview of protoplanetary discs, that were discovered in the Orion Nebula (credit: NASA/ESA and L. Ricci (ESO))

that many such clouds exist in the Universe and we are able to observe the star formation there as well as the formation of planetary systems (see Fig. 1.5). Prominent examples of such star forming regions are the Orion nebula, the Trifid nebula and the Eagle nebula.

These giant molecular clouds mainly consist of hydrogen and helium. Over a period of several millions of years, these clouds tend to collapse and fragment.

In the case of the cloud from which our solar system formed, each of the fragments had a size of about 3.25 lightyears. Of course, you may ask what causes the fragmentation? The cloud is floating in the galaxy why should it change? Various processes are conceivable. Just imagine a nearby star is exploding in form of a supernova. The shock waves originating from the supernova will propagate through the cloud and may cause local instabilities likes waves in the ocean. Some parts may become denser than others and thereby the seed for fragmentation is sowed. One of these collapsing fragments formed what became the Solar System. The masses of such protostellar nebulae can range from merely fractions of the mass of our Sun to several times the mass of our central star. The process of collapsing will last about 100,000 years for a nebula with having the approximate mass of our Sun.

Fig. 1.6 One of protoplanetary discs discovered in the Orion Nebula (credit: NASA/ESA and L. Ricci (ESO))

The nebula further collapsed. It had a certain amount of angular momentum and because of the conservation of exactly this momentum, the nebula spun faster as it collapsed. As the material within the nebula condensed, the atoms within it began to collide with increasing frequency, converting their kinetic energy into heat. At the center of the nebula the concentration of molecules and atoms was the highest within the whole nebula. The frequency of collisions was also higher which eventually caused the center to become hotter than the surrounding disc.

Over a period of about 100,000 years the competing forces of gravity, gas pressure, magnetic fields and rotation caused the protostellar nebula to flatten into a protoplanetary disc. This disk extended up to about 200 AU. Its center formed a hot, young protostar. A protostar is pre-stage to a normal star. However, the fusion of hydrogen has not yet begun in a protostar. Such protoplanetary discs have already been observed in distant regions, e.g. in the Orion nebula (see Fig. 1.6).

The rotating protoplanetary disc further provided material to the protostar. Hence, the contraction further continued. The protostar gained more mass and the temperature and pressure in it continuously increased. After about 50 million years, a turning point was reached. Temperature and press at the protostar had become so high as to trigger nuclear fusion. A star is born, our Sun, was born.

The nuclear fusion process created an internal heat source. The heat increased the pressure within the young star that countered gravitational contraction until hydrostatic equilibrium was achieved, i.e., until the outwards directed pressure and the inwards directed gravitation had the same strength. Our Sun became stable.

1.2.1 Formation of the Planets

The protoplanetary disc around our Sun yet remained and comprised the remnants of the solar nebula in the form of gas (mainly hydrogen and helium) and dust (various other heavier molecules such as silicates). It is important to see that the disc contained both types of material. Otherwise a formation of a solar system as ours is not possible.

While the protoplanetary disc continued to rotate around the Sun and delivered further material to it in an accretion process, the disc also started to differentiate. The dust began to settle through the gas to the central plane of the disc and formed a thin disc of dust there.

Then, the disc further differentiated into distinct regions owing the different temperatures present in various parts of the disc. The inner region close to the Sun ranging up to about 4 AU was too warm for volatile molecules like water and methane to condense and remain solid. They simply evaporated. However, heavier molecules with a higher melting point such as metals (iron, nickel, aluminum) and rocky silicates remained there and condensed.

A bit farther out, also carbonaceous compounds, which have a slightly lower melting point compared to metals, remained present in solid form. There was, however, a problem. All these particles that remained in the inner region are not very common in the Universe. Their abundances are much lower than for volatiles such as methane and foremost hydrogen and helium. Hence, the terrestrial planets, which should be born from them, were limited in their growth.

When we leave the inner region behind us, we cross the so-called frost line which lies in the region of the current asteroid belt between the orbits of Mars and Jupiter. Beyond this line, temperatures are low enough to allow the presence of water ice in solid form. The farther we leave the Sun behind us, the lower the ambient temperatures become and further even more volatile molecules such as methane can remain frozen.

From Dust Grains to Terrestrial Planets

The terrestrial planets essentially formed in the inner region of the protoplanetary disc. The dust grains in orbit around the Sun in this inner region agglomerated trough direct contact to millimeter-sized objects. The particles were so small that the gravitational effects among them were negligible. Through further direct contact these tiny seeds clumped together up to several hundred meters in diameter. This process continued until these objects had accreted enough material to form objects of about 10 km in size, the so-called planetesimals.

This was the point of time, when gravity prevailed. The planetesimals influenced each other gravitationally and collisions occurred. These collisions, of course, caused disruption but also could lead to further accretion. In this phase, in total, the accretion dominated the disruption by collisions. The planetesimals grew until only a few of them had survived and had formed planetary embryos of about 0.05 Earth masses. Subsequent mergers and collisions led to the formation of the terrestrial planets as we know them today. Some of these collisions must have

Fig. 1.7 Artist's conception showing a celestial body about the size of our moon slamming at great speed into a body the size of Mercury (credit: NASA/JPL-Caltech)

been very dramatic. One such collision is believed to have formed our Moon and another to have blown away Mercury's outer mantle leaving behind merely its naked core (see Fig. 1.7).

Yet, the terrestrial planets were still immersed in the protoplanetary disc. The gas within the disc, of which the planets were surrounded, did not move as rapidly around the Sun as the planets did. This resulted in a drag caused by the transfer of angular momentum. The planets slowed down and migrated to orbits closer to the Sun. When the protoplanetary disc finally dissipated, we will come to that in a moment, the migration stopped and the terrestrial planets had arrived at their current orbits.

The Evolution of Giants

Beyond the frost line, the formation of planets was the result of different processes. In that region, water ice and other form of ices were present. Yet, water ice dominated them all being the most abundant icy material in the protoplanetary disc and today's solar system. In general, the ices that formed the gas giants were more abundant than the metals and silicates that formed the terrestrial planets, allowing the giant planets to grow massive enough to capture hydrogen and helium, the lightest and most abundant elements. The planetesimals accumulated about up to four times the mass of Earth within a period of about 3 million years. A runway process began during which the young gas giants accreted further material rapidly and thereby increased their sizes drastically.

Jupiter is believed to have formed first. It accreted much of the ices in the protoplanetary disc. Hence, less material was left for the formation of the other gas planets. The next one that formed was Saturn and for the just mentioned reason is smaller and less massive than Jupiter.

Uranus and Neptune are believed to have formed after Jupiter and Saturn. The latter ones had already used up most of the icy materials. In addition, the young Sun, a so-called T Tauri star, was more active and had a much stronger solar wind than we can experience today. This strong wind had already blown away much of the disc material when the two ice giants Neptune and Uranus formed. As a result, the planets accumulated little hydrogen and helium. Their cores grew to masses equivalent to Earth.

However, there exists a timing problem regarding the formation of the gas giants. At their present positions, the protoplanetary disc was less denser populated and it would have taken about a hundred million years for their cores to have formed. By that time, however, the protoplanetary disc would have already been dissipated a long time ago leaving no material behind for the accretion process. Current models suggest that after between 3 and 10 million years, the young Sun's solar wind had cleared away all the gas and dust in the protoplanetary disc by blowing it into interstellar space. The growth of the planets ended.

How can this problem be solved? Basically, two possibilities are conceivable. First, a yet unknown process was involved in forming the ice giants that somehow led to a faster accretion of material or provided the necessary material at the region where they formed. Secondly, the two planets did not form at their current locations.

The latter possibility is considered to be the most plausible one. The known processes are at least partially well understood and can explain most of the formation features. No further unknown process would be necessary except for the problematic case of the ice giants. And to cite the Greek philosopher Ptolemy (90–168): "We consider it a good principle to explain the phenomena by the simplest hypothesis possible."

Hence, if Uranus and Neptune did not form at their current locations, they must have formed somewhere else and then migrated to their current orbits. Computer simulations can help us understand the various scenarios. In a currently well-accepted scenario, it is assumed that all of the giant planets (Jupiter, Saturn, Uranus and Neptune) have formed closer to the Sun and that they were also closer together. They have reached their current orbits after a phase called planetary migration. As turns out, the migration is also necessary to explain the remaining features of the outer solar system.

The Asteroid Belt and the Outer Solar System

Before we describe the process of planetary migration, we first need to know more about the so far neglected elements of our young solar system. These played a key role in the further evolution of the system.

The early solar system after the dissipation of the protoplanetary disc did not only comprise the Sun and the planets. A huge amount of smaller bodies ranging from dust particles, to small rubbles and to planetesimals was still left. On the outer edge of the inner solar system between 2 and 4 AU, a large gap exists in which no planet had formed. It is filled with thousands and thousands of smaller rocky bodies, which we call asteroids today. In the early phase, this zone was much denser populated and there was enough material in form of smaller bodies and planetesimals in there to form about two or three Earth sized planets.

The conditions sounded promising for the existence of a planet. Why did no planet form there? The simple answer is: Jupiter! Jupiter is the by far largest and heaviest body in the solar system besides the Sun. He is approximately 2.5 times more massive than all the remaining planets combined. This tremendous mass of about $1{,}899 \times 10^{27}$ kg implies strong gravitational effects on the other planets and in particular on the small bodies of the solar system. The proximity of the primordial asteroid belt to Jupiter meant that after the gas giant had formed, about 3 million years after the Sun, the region's history changed dramatically.

Jupiter's gravity destabilized the orbits of some of the bodies in this area and increased their velocities. The gas giant kicked some of the bodies out of the solar system or injected them into the inner or outer solar system. The increased velocities further resulted in heavier collisions in which the bodies were shattered. The shattering clearly dominated any present accretion processes.

A similar area evolved beyond the gas giants. However, while in the primordial asteroid belt bodies of rocky or metallic nature dominated, the primordial outer solar system was the realm of small icy bodies. A large disc of these small icy bodies is believed to have existed there, the primordial Kuiper belt. At this distance from the Sun, the density of the protoplanetary disc was low and hence the accretion was too slow to allow planets to form before the dissipation of the disc, and thus the initial disc lacked enough mass density to consolidate into a planet. Today's he Kuiper belt lies between 30 and 55 AU from the Sun. The primordial Kuiper belt, also called the proto-Kuiper belt was much denser and closer to the Sun, with an outer edge at approximately 30 AU. Its inner edge would have been just beyond the orbits of Uranus and Neptune, which were in turn far closer to the Sun when they formed (most likely in the range of 15–20 AU), and in opposite locations, with Uranus farther from the Sun than Neptune.

Planetary Migration or Nothing Stays the Same

We now have all the basic ingredients necessary to understand what was going on during the phase of planetary migration. The early solar system before the planetary migration consisted of the eight planets whereas Uranus and Neptune were in opposite locations. The outer planets were much closely spaced and more compact than in present days. The planets then migrated until they reached their current positions. How can we know that? What did happen during the migration? There are many open questions. Computer simulations help to understand better the processes that were involved. Various models have been proposed and discarded again. A currently promising model is the so-called Nice model named for the location of the Observatoire de la Côte d'Azur in Nice where they were initially developed by Rodney Gomes, Hal Levinson, Alessandro Morbidelli and Kleomenis Tsiganis.

This model assumes that the four giant planets (Jupiter, Saturn, Uranus and Neptune) were originally found on near-circular orbits between about 5.5 and approximately 17 AU. Today, their orbits lie between 5.2 and roughly 30 AU. In the early system, Neptune and Uranus had swapped positions making Uranus the outermost of the gas giants. A large, dense disk of small, rock and ice planetesimals, the proto-Kuiper belt, extended from the orbit of the outermost giant planet to some 35 AU. Today, we have roughly three populations of small bodies in the outer solar system: the Kuiper belt, the Scattered Disk and the Oort cloud.

Small bodies at the proto-Kuiper belt's inner edge occasionally passed through gravitational encounters with the outermost giant planet, which change the small bodies' orbits. Most likely these small bodies were scattered inwards towards the inner solar system. By scattering the small bodies inwards, an exchange of angular momentum takes place which in turn pushes the planet slightly outwards. The inwards moving objects then come closer to the next gas giants influence zone, its Hills sphere. The process repeats and the planetesimal is further casted inwards. So, step by step, not only the small bodies move inwards, but also the gas giants change their orbits to farther distances from the Sun.

Of course the push, the gas giants received while interacting with the small bodies is very small, almost negligible and hardly influences the planet's orbit. Yet, the cumulative effect of many small bodies encounters, shifted the orbits of the gas giants significantly over time.

This process comes to an end when the small bodies encountered Jupiter, the by far most massive planet in our solar system combining in it more than two times the mass of the sum of all other planets together. Jupiter's enormous mass leads to a different kind of interaction. In most cases, the gas giant forced the small bodies on highly eccentric orbits or even kicked them out of the solar system. This is also considered to be the hour of birth of the Oort cloud. By scattering the small bodies outwards, Jupiter very slowly moves inwards due to the preservation of angular momentum.

About 500–600 million years of slow but gradual migration passed when Jupiter and Saturn reached their 1:2 mean motion resonance which increased their respective orbital eccentricities. This strong resonance caused a destabilization of the entire solar system in which essentially Jupiter pushes Saturn to its current position also due to mutual interactions with the two ice giants, Uranus and Neptune. Also these two end up with by far more eccentric orbits and finally Uranus and Neptune swap positions making Neptune the outermost planet.

In particular, the migration of Neptune had severe consequences for the proto-Kuiper belt in which it now fully immersed. The ice giant thereby approached the small bodies therein. By its gravitational influence, Neptune succeeded to capture some of them into resonances while it pushed others into more or less chaotic orbits with higher eccentricities. This caused temporary chaos within the belt.

Many of the planetesimals, however, were casted inwards closer to the Sun. Their fates were then finally decided by their encounter with Jupiter. This process thinned out the proto-Kuiper belt drastically. Some scientists believe that more than 90 % of its small icy bodies were lost thereby.

The surviving bodies, either captured into resonances or thrown into more chaotic orbits, were in consequence pushed outwards. Some of them formed what we call the Kuiper belt nowadays; others, especially those with more inclined and eccentric and potentially unstable orbits established the scattered disk.

Yet, not only the planetesimals were affected but also the two ice giants. Friction within the belt made their orbits more circular again leading to the situation of the solar system, as we know it today.

This 1:2 mean motion resonance of Jupiter and Saturn also had a big influence on the primordial asteroid belt. A large number of planetesimals was captured in the outer asteroid belt at distances larger than 2.6 AU from the Sun. In that part, a collisional erosion occurred in which much smaller fragments of the original planetesimals were created by collisions. These smaller bodies were so small that they could be influenced by the solar wind and blown away by it. This removed about 90 % of the original material from the primordial belt.

Furthermore, while Jupiter migrated further inward, the gas giant perturbed the orbits of many bodies in the primordial belt. Some of them were either casted either inwards into the inner solar system or outwards towards the Kuiper belt or towards a region at the fringe of our solar system, the Oort cloud. Some of them were even kicked out of the system. This led to a further loss of material. Astronomers believe that the primordial belt had a total mass of about one time the mass of Earth. At the end of its formation it had been reduced to about 0.1 % of that original mass.

Our preparatory work is done and we can start our journey to the small bodies of our solar system.

Chapter 2
Asteroids

We begin our journey to the small bodies in our solar system with asteroids. These are one of the two most famous types of small solar system bodies. If asked, the man on the street would most probably confirm that he had heard of asteroids. Yet, the knowledge he has is often limited to Hollywood movies or news in the media.

When we look into the starry sky to observe objects of our solar system, we are usually fascinated by the Moon, the planets and the impressive comets with their sometimes stunning tails. Asteroids are often neglected as they often are considered to be boring. Dead and cold bodies made of rock, hardly visible. Indeed, only Ceres, now being classified as a dwarf planet, can be seen by naked eye and only under perfectly dark skies which has become rare nowadays. But even with the help of telescopes, it is almost impossible to resolve asteroids and identify any surface details.

Asteroids only become a matter of public interest either when media attention is drawn to them, which happens quite regularly, or when a new blockbuster movie on that topic is airing. In both cases a collision with our home planet is usually the main subject. In particular, the media reports of potential collision candidates are often strongly exaggerated. When kept in perspective, it turns out that the alleged near misses happen at such great distances that already from the beginning it had been clear to scientists that a collision was impossible.

Yet, if we let this straining after effect aside, we will find interesting aspects of these small worlds that are by far more complex than we might expect.

The term asteroid is derived from Greek, meaning star-like object. As of today we know of about 680,000 asteroids in the main belt. The actual number is thought to be in the millions. Only a few of them have diameters larger than 100 km. Asteroids are not massive enough to reach hydrostatic equilibrium, i.e. their mass is not sufficient to gain a spherical shape. This, however, makes them also interesting as they are often irregularly shaped (see Fig. 2.1).

M. Moltenbrey, *Dawn of Small Worlds*, Astronomers' Universe,
DOI 10.1007/978-3-319-23003-0_2

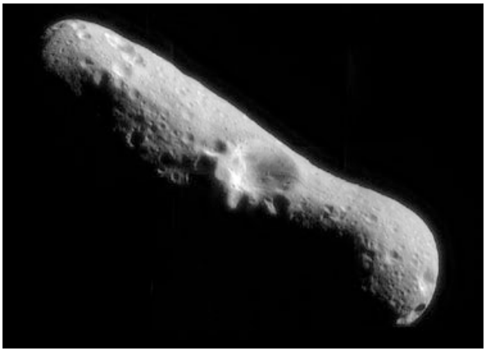


Fig. 2.1 Mosaic of the asteroid Eros taken by the NEAR-Shoemaker spacecraft on February 14, 2000 (credit: NASA/ NEAR Project (JHU/APL))

Although the vast majority of asteroids orbit the Sun in the main asteroid belt between Mars and Jupiter. Yet, there are other groups and families, like the Trojans, that we will learn about in the latter course of this chapter.

These small bodies are remnants of the protoplanetary disc, so-called planetesimals that existed during the formation of our solar system. In spite of their common origin, asteroids are composed of a variety of different chemical compounds. Accordingly, asteroids can be classified as carbon-rich (C-type), stony (S-type) or having metallic compositions (M-type).

We will discuss all this in the following. Beforehand, it is interesting to have a closer look at their initial discoveries.

2.1 History and Early Observations

Other than with comets which have been known for thousands of years, observations of asteroids only dates back about 200 years. Before the advent of telescopes, mankind was completely unaware of them, as they are in general invisible to the naked eye.

This changed at the end of the eighteenth century. In 1766, the German scientist Johan Daniel Titius (1729–1796) formulated a simple mathematical law that was later adapted by Johan Elert Bode (1747–1826) and came to be known as the Titius-Bode-Law. This law relates the semi-major axis a of each planet outward from the Sun to a simple mathematic progression of numbers

$$a = 4 + n; \quad n = 0, 3, 6, 12, 24, 48, 96, 192, 384$$

If we apply the Titius-Bode law to our solar system, we obtain the following numbers.

4, 7, 10, 16, 28, 52, 100, 196, 388

Dividing these numbers by 10 results in

0.4, 0.7, 1.0, 1.6, 2.8, 5.2, 10.0, 19.6, 38.8

Do these numbers remind you of something?

Let us compare these numbers to the semi-major axes of the planets in our solar system

Planet	Titius-Bode Law	Distance (AU)
Mercury	0.4	0.39
Venus	0.7	0.72
Earth	1.0	1.0
Mars	1.6	1.52
X	2.8	X
Jupiter	5.2	5.20
Saturn	10.0	9.54

Isn't it impressive that this simple law works? The people of that time believed in it. The law was widely accepted in the scientific community. This was further solidified when the German-born British astronomer Sir William Herschel (1738–1822) discovered Uranus at roughly the distance proposed by the law. Today we know that it is probably nothing more than a strange coincidence.

One thing that drew the attention of the scientific community was the gap between Mars and Jupiter where, according to the law, a planet should have existed at a distance of 2.8 AU.

The hunt for the unknown planet started. In 1800, Baron Franz Xaver von Zach (1754–1834) and Johann Hieronymus Schroeter (1747–1816) founded probably the first international research project, the "Himmelspolizey" (engl. the "sky police") while working at the observatory in Gotha, Germany. The group divided the starry sky into 24 distinct regions and each member, located in a different European country, was asked to systematically search his region for the missing planet.

Though the search efforts were considerably high, the Himmelspolizey lost the race. The "missing planet" was found by chance by the Italian astronomer and theologian Giuseppe Piazzi (1746–1826) during the night of New Year's Day 1801 (January 1, 1801) while working at the observatory of Palermo, Italy. Piazzi had been observing the night sky when he discovered a faint star in the constellation Taurus. The "star" was not registered in any star map or catalogue that he knew. Additionally, he was able to follow the object moving relative to the background stars over the course of several nights.

Piazzi was aware of the search going on by the Himmelspolizey, so he decided to send his observation records and notes to von Zach. Unfortunately, the lucky discoverer got sick and was no longer able to track the object for some time. It

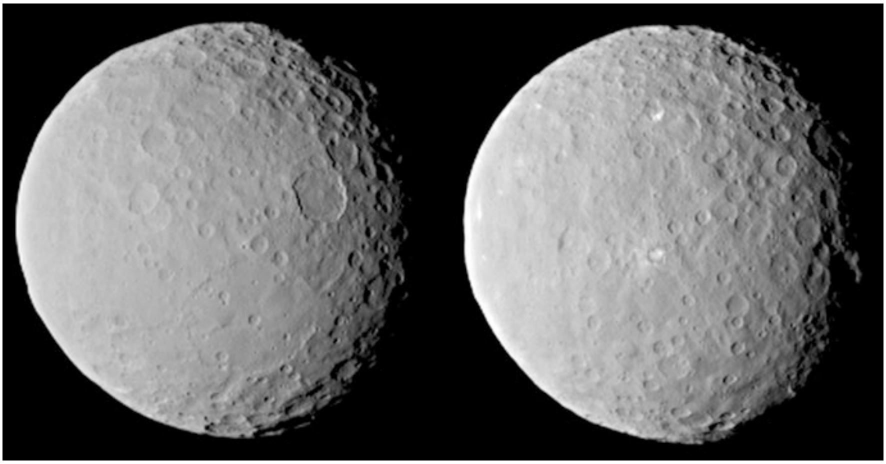


Fig. 2.2 Images of the dwarf planet Ceres taken on Feb. 19, 2015, from a distance of about 46,000 km by NASA's Dawn spacecraft (credit: NASA/JPL-Caltech/UCLA/MPS/DLR/IDA)

took a long time for Piazzi's observations to be published. During this time, the "missing planet" was lost and initially all efforts to rediscover it failed.

The German mathematician Carl Friedrich Gauss (1777–1853) had developed a method, the method of least squares, which allowed him to determine an object's orbit by using only a few limited numbers of recorded positions. The derived orbit suggested the object to be between Mars and Jupiter at about 2.8 AU distance from the Sun, exactly where the missing planet was supposed to be. Using Gauss' calculations, Heinrich Wilhelm Olbers (1758–1840) was finally able to rediscover the missing planet on December 31, 1801.

Its discoverer, Piazzi, named the planet Ceres (see Fig. 2.2) after the Roman Goddess of Agriculture.

Olbers managed to find three other objects similar to Ceres in the same region of space over the next few years: Pallas (1802), Juno (1803), and Vesta (1807).

It took about 38 years until the fifth object, Astraea, was discovered by the German amateur astronomer Karl Ludwig Hencke (1793–1866) in December 1845. In the meantime, it had been a common belief that more new "planets" existed in this region between Mars and Jupiter. These objects were still considered to be planets. This, however, meant that at its time of discovery in 1846, Neptune was not considered to be the 8th planet but the thirteenth. In the following years the discovery rate of new objects increased drastically.

Severe doubts about their planet status arose. Could there be so many planets at similar distances to the Sun although they appeared to be much smaller than the big planets? A new group of solar system bodies was created and the four objects were thus classified as asteroids. Hence, Ceres was dethroned as a planet and demoted to an asteroid. This would not be its last status change. When, in 2006, the International Astronomical Union created the new class of dwarf planets, Ceres was promoted into this group. It is, thus, the only dwarf planet residing in the inner solar system. All others are located in the trans-Neptunian region.

Until 1891 about 300 asteroids had been identified. That year a marked a turning point in the discovery of small objects. The German astronomer Maximilian Franz Joseph Cornelius Wolf (1863–1932) pioneered the use of astrophotography to automate the discovery of asteroid. Until then, the discovery of new asteroids (and also comets) had been a very cumbersome and ineffective task. The visual findings at the telescope's eyepiece had to be compared to star charts and catalogues. Now, with the advent of astrophotography it became much easier as the objects were directly banned on photographic plates. Due to their motion asteroids appeared as small streaks on long exposure photos and could thus easily be identified. Using this method, Wolf alone was able to discover more than 200 asteroids.

The number of known asteroids has grown since then to about 680,000. This sounds like a tremendous number, but the total mass of these objects is, as we will see later on, almost negligible compared to that of the planets. If we look at the mass of Ceres which is about 9.35×10^{20} kg, we can see that this is even considerably lighter than the mass of our Moon (7.349×10^{22} kg) which is far from being the largest moon in our solar system. With this mass Ceres is by far the largest and heaviest asteroid.

2.2 Naming of Asteroids

Before we turn to more details about asteroids, it may be helpful to see the naming scheme of asteroids.

Newly discovered asteroids are first given a provisional designation in order to be able to properly identify them. This designation consists of the year of discovery, and an alphanumeric two-letter code indicating the half-month of discovery and the sequence within this half-month. The alpha-numeric code is relatively easy to create: The letter “A” is assigned to the first half of January, “B” to the second half and so on until the last half of December is designated “Y”. The letter “I” is omitted as it could too easily be mixed up with the number “1”. The second letter defines the order of discovery within the half-month: “A” meaning the first asteroid in this half month, “B” the second and so until we reach number 25 with letter “Z”. “I” is again not used.

This scheme was sufficiently adequate for quite some time. However, with new discovery methods number of new asteroids found drastically increased. Often more than 25 per half month were detected. Thus, an additional subscript number was added to the alpha-numeric code indicating the number of times that the letters have cycled through.

Let us have a look at a few examples. The designation “2015 AA” refers to the first asteroid discovered in the first half of January 2015. Asteroid “1950 FC1” is the 28th asteroid discovered in the second half of March 1950. A last example: what does “1992 QB1” mean? Yes, it is the 27th asteroid discovered in the second half of August 1992. We will see later on, that this object is of special pertinence.

Okay, so much about the provisional designations. As soon as the orbit of the asteroid is confirmed, meaning it really exists, can be found again and is indeed an asteroid, it is given a number enclosed in brackets (subsequently counted) and sometimes it is also given a name. However, naming has gone out of fashion for quite some time. The sheer number of new discoveries makes it difficult to assign a name to each one.

In addition, originally the name was taken from Roman or Greek mythology but, the reservoir of available names was soon exhausted. Nowadays, names from various sources can be used, such as a famous person, the spouse’s name or a TV character. Just to mention a few: (2309) Mr. Spock, (9007) James Bond and (26858) Misterrogers.

However, the choice of name, contrary to what it might appear from the above, is not completely free. For some objects the name space is limited. The centaurs, objects between Saturn and Neptune which show characteristics of both comets and asteroids, are restricted to the names of the centaurs known from Greek mythology. The names of the Jupiter Trojans, an important class of asteroids co-orbiting the gas giant, have to be chosen among the heroes of the Trojan War.

2.3 Origin and Evolution

Heinrich Olbers postulated a first hypothesis on the formation of the asteroid belt in 1802 shortly after he had discovered the second asteroid, Pallas. He suggested that both Ceres and Pallas were remnants of a much larger planet that originally orbited the Sun in the same region between Mars and Jupiter approximately 2.8 AU from the Sun.

He thought the planet had been destroyed either by a massive internal explosion or a comet impact. The fragments then formed the asteroids or planets, as they were then called.

Yet, this hypothesis soon fell in disgrace after the advent of spectral analysis which allowed the determination of the chemical compositions if these bodies. It turned out that this varied significantly from asteroid to asteroid. How could these diverse bodies have originated from a common body?

In addition, an event having the huge amount of energy that would be needed to tear a planet apart was considered to be unlikely to arise. No process was known by that time that could unleash such a force.

The last counter-argument brought forward was the relatively low mass of all the asteroids put together. Was that sufficient to form a planet? And so the scientific community discarded this hypothesis.

2.3.1 From the Protoplanetary Disk to a Primordial Belt

Other hypotheses were formulated, but nowadays astronomers believe that the asteroid belt formed in the same process as the planets in the protoplanetary disk. We have already learnt about this in a previous chapter when we discussed the formation of our solar system.

Here, we will just focus on the details necessary to understand the origin of the asteroid belt. An accretion process took place in the protoplanetary disk in which small particles within the disk collided and stuck together. By this, clumps of small particles developed that gradually grew in size. When a certain threshold of the mass of these clumps was exceeded, gravitational effects replaced the sticky collisions. The clumps were massive enough in order to attract others and collide with them. The clumps grew into planetesimals whose interactions led to the formation of the terrestrial planets and the gas giants.

Why then, following the same process, did no planet develop in the region between Mars and Jupiter? This is owed to the influence of the gas giants and in particular Jupiter's gravitational influence. Astronomers assume today that after the dissipation of the gas and dust of the protoplanetary disk at the end of the formation of our solar system, the four gas giants Jupiter, Saturn, Uranus and Neptune were originally much closer together in a region stretching from a distance of about 5.5 to 17 AU from the Sun. At present, their orbits lie in the region of about 5–30 AU.

The orbits of the planetesimals in the early, primordial asteroid belt were strongly perturbed by Jupiter's gravity, which essentially hindered the formation of another planet. Due to the perturbations, the number of collisions of the planetesimals significantly increased and became more successful than the accretion of the planetesimals.

Orbital resonances existed with Jupiter within this early belt, some being disruptive, and some stabilizing. Thus, in some resonance bands the planetesimals were not perturbed, in others, however, massively. This led to a thinning out of bodies in the latter ones. The scattered objects moved along more or less arbitrary orbits in the early solar systems and were often on a collision course with one of the planets or their moons. Impact craters on terrestrial planets and rocky, planetary moons are the best evidence for these events.

2.3.2 Planetary Migration Reshuffles the Pack

Then, by interacting with another disc of planetesimals beyond the orbits of the outermost gas giants, the proto-Kuiper belt (see Chap. 4), a process called planetary migration started in which the outer gas giants Saturn, Neptune and Uranus moved gradually outwards while Jupiter moved towards the Sun. Several hundred million years (500–600 million years) of slow but gradual migration passed when Jupiter and Saturn reached their 2:1 mean motion resonance. This increased their respective orbital eccentricities. This particular strong resonance caused a destabilization of the entire solar system in which essentially Jupiter pushed Saturn to its current position with the help of mutual interactions with the two ice giants, Uranus and Neptune.

This 2:1 mean motion resonance also had a big influence on the early asteroid belt. Following the so-called Nice-model describing the planetary migration a large number of planetesimals were captured in the outer asteroid belt at distances greater than 2.6 AU from the Sun. In that region, collisional erosion occurred in which much smaller fragments of the original planetesimals were created by collisions. These smaller bodies were so small that they could be influenced by the solar wind and effectively be blown away. This removed about 90 % of the original material from the primordial belt.

Furthermore, while Jupiter migrated further inward, the aforementioned resonance bands swept across the belt and dynamically excited the region's population. Just think about a region that had been stable and now became subjected to Jupiter's devastating gravitational influence. This led to a further loss of material. Astronomers believe that the primordial belt had at one time a total mass equivalent to that of Earth. At the end of its formation it had been reduced to about 0.1 % of that original mass.

2.3.3 Further Evolution, or, Nothing Stays the Same

Even after this eventful formation history, the asteroid belt did not come to a halt. If we were to assume that, asteroids are composed of primordial material of the protoplanetary disk, we would be wrong. Asteroids developed further and underwent a considerable evolution, significantly changing their chemical compositions.

The larger asteroids were subjected to a brief period of radioactive decay of the aluminium isotope ^{86}Al and iron isotope ^{60}Fe during which the asteroids partially melted. The resulting heavier elements such as nickel (Ni) and iron (Fe) sank to the inner core due to gravitational effects, just as a stine sinks in water. The lighter compounds such as silicates remained in the upper, outer layers of the asteroid. Thus, the once uniformly composed bodies were transformed to have differentiated metallic cores and silicate mantles.

In addition to radioactive decay, surfaces were also likely to melt and thus subject to alteration due to impacts and bombardments by micrometeorites. Basically, all asteroids were subjected to these events.

There is one more aspect dating back to the formation of the belt which acts to destroy the assumed homogeneity of asteroids in the asteroid belt. At a distance of 2.7 AU from the Sun, a so-called "snow line" developed. Beyond this line, planetesimals were able to accumulate ice as their temperature dipped below the freezing points of the volatiles. These objects may exhibit a different surface structure as those within the line. The bodies become more comet-like. Indeed, in

2006, it was announced that comets had been detected within the asteroid belt beyond the snow line, the main belt comets. We will deal with these in Chap. 3.

2.4 Distribution Within the Solar System

Before we have a closer look at the asteroids' orbital and physical characteristics, it is advisable to provide a broad overview on the distribution of these bodies within the solar system and to bring some order into the apparent chaos. Thereafter, we will address each group of asteroids individually in detail.

In general, we should note that most asteroids have more chaotic and eccentric orbits compared to the planets, which travel around the Sun in near-circular orbits.

Let us start our considerations by looking at the asteroids within the orbit of Mars, these being inner solar system asteroids. Several groups exist. Most of them are composed of quite small bodies with diameters of 5 km or less. Some cross the orbits of planets, others move completely within or outside of these.

The first ones we encounter when traveling from the Sun to the outer regions of our Solar System, is a postulated group named Vulcanoids named after the hypothetical planet Vulcan, whose existence was disproved in 1915. Some scientists proposed the existence of the Vulcanoids in order to explain certain aspects of Mercury's orbit. This group is supposed to be very close to the Sun, i.e., well within the orbit of Mercury. So far, no Vulcanoid has been discovered and it is not clear whether any exist. Another large group comprises the Near Earth Asteroids (NEA) whose orbits pass close to the orbit of the Earth. Their perihelia are usually considered to be at distances of less than 1.3 AU from the Sun. This group draws particular attention as its members may be potential collision candidates. Therefore, a large number of automatic survey programs have been set up in order to identify as many NEAs as possible. Among them are the Lincoln Near Earth Asteroid Research (LINEAR), Catalina Sky Survey, Pan-STARRS, NEAT, and LONEOS. So far these programs have proven to be very successful.

Figure 2.3 illustrates the further sub-classification of NEAs. There is an Amor-type and Apohele-type. Then there is the category of Earth-crossers meaning Near Earth Asteroids that actually cross the orbit of Earth. Among them are the Apollo-, Aten- and Arjuna-asteroids.

If we move a bit farther out in the solar system, we encounter the main asteroid belt between Mars and Jupiter, which comprises about 90 % of all known asteroids in our solar system. The main asteroid belt is further subdivided into the inner main belt ranging from about 2.06 to 2.5 AU distance from the Sun, the middle belt (2.5 to 2.8 AU), and the outer belt from 2.8 to approximately 3.3 AU distance from the Sun.

Additionally, some smaller groups of asteroids can be found outside the main belt whose orbits are often in stabilizing resonances with Jupiter. What makes them particularly interesting and distinct from the main belt asteroids are their often highly inclined orbits compared to the ecliptic plane. The Hilda-group, Hungaria-group and the Cybele-family are just a few to name.

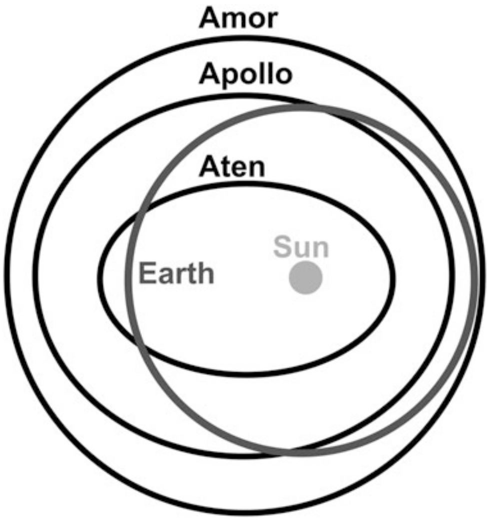


Fig. 2.3 Types of near earth asteroid orbits (based on work by Berklas; Public Domain)

Then, when we leave the asteroid belt and pass the orbit of Jupiter, we discover more strange fellows that cannot be easily classified anymore. Among them are two big groups, the centaurs, which we briefly discussed before, and the Damocloides named after (5335) Damocles. The aphelia of the latter ones lie often beyond the orbit of Uranus and have very eccentric orbits thus putting their perihelia often into the inner solar system.

The last important group is the Trojan asteroids. There are special in some respects. They are co-orbiting a planet at its Lagrange points. The most prominent and largest group is Jupiter Trojans. Yet, Trojans of Venus, Mars, Uranus, Neptune and even Earth are known.

2.5 Vulcanoids: The Phantoms

In our detailed discussion of the various asteroids, we first came across the Vulcanoids. As previously stated, these are a special group of objects as their population merely exists hypothetically. No evidence of these bodies has been found so far. Yet, quite a few astronomers believe they may exist for several reasons. One reason is that in all other similar areas around a more massive object in the solar system, such bodies do actually exist. Why should there be none between Mercury's orbit and the Sun?

The Vulcanoids are supposed to orbit in a dynamically stable zone inside Mercury's orbit. To recap, they are named after the hypothetical planet Vulcan that was proposed to exist there but whose existence was disproven in 1915.

Okay, we were able to show that Vulcan does not exist. Why is it so difficult with the Vulcanoids? The answer is quite simple. If they are out there, they are very close to the Sun. This only allows Earth-based observations during twilight or solar eclipses. Not the best conditions as you can imagine. In addition, it is hardly possible to use large telescopes for their detection. The closeness to the Sun and the potential damages caused by high intensity light falling into them accidentally would severely damage their detectors. Most probably no one will be able to risk this.

Yet, other techniques have been used so far. Planes equipped with astronomical devices, for example, were sent up into the sky in order to allow observations above the interference of Earth's atmosphere. This was tried several times without success.

What if we go farther above into space? It is true that specialized space probes may help. However, they are too expensive to be sent up there to merely search for these hypothetical Vulcanoids.

Could we not use already operating space probes for this? This is exactly what has been done a few times. SOHO carried out a search for them in 1998 but did not find any objects larger than 60 km in diameter in the predicted zone. Consequently, we have to look for objects smaller than that size.

The Solar Terrestrial Relations Observatory (STEREO), launched in 2006 by NASA, subsequently made another attempt. Its results rule out any Vulcanoids larger than 6 km in diameter. Furthermore, the "Mercury Surface Space Environment, Geochemistry and Ranging space craft" (MESSENGER), launched in 2004 by NASA, also failed to detect any Vulcanoids.

However, these unsuccessful attempts do not rule out their existence per se. In any case, we can derive some of their physical and orbital characteristics.

2.5.1 Physical Characteristics

The Vulcanoids have to be very small. If they were larger than 6 km in diameter they would have already been discovered. Most probably, their size is larger than 100 m. Otherwise the radiation pressure and Yarkovsky effect would be too devastating.

They further should be rich of elements with a very high melting point such as iron and nickel. Most likely, they should be hot enough to glow reddish.

2.5.2 The Stable Band: Why Is It Apparently Empty?

The Vulcanoids are supposed to have semi-major axes of less than 0.387 AU, the orbit of Mercury. It is expected that a gravitationally stable band exists inside the orbit of the tiny planet ranging from about 0.06 to 0.21 AU. In all other similar stable regions in the solar system, objects have been found. There seems to be, at first glance, no reason why the one so close to the Sun should be empty. Yet, this environment is even more hostile than other places in the solar system. Non-gravitational forces such as a very strong radiation pressure from the Sun and the Yarkovsky effect play a significant role there and may have cleared the region quite some time ago.

Are there no indications for their existence? It is known that a large object hit Mercury early in its development. This collision stripped away much of the planet's crust and mantle. The resulting debris might still be orbiting the Sun in the Vulcanoid zone.

However, all these are only hypothetical discussions and it still needs to be shown whether the Vulcanoids remain phantoms, like the from time to time proposed companions of our Sun or additional planets in our solar system.

2.6 Near Earth Asteroids

As we leave the realm of hypothetical objects, we first encounter the Near Earth Asteroids when departing from the Sun. As their name indicates, these objects can come pretty close to Earth on their orbits. Depending on their actual orbits and inclination they can become potentially dangerous for our home planet. For this reason, several survey programs have been set up.

The NEAs are not located in the main asteroid belt but travel in the region of the inner solar system. Some of them, though, have orbits stretching into the main belt or even beyond.

The first NEA that had been detected is (433) Eros (see Fig. 2.4) and which had subsequently been visited by the NEAR spacecraft.

Yet, not all NEAs are similar. They can differ quite a lot in their orbits and hence are further classified based on their semi-major axes, perihelia and aphelia. Some can cross Earth's orbit, some come close to the Earth from the outside, others merely orbit inside Earth's orbit and so on. You can see the different types of NEAs in Fig. 2.3. We will discuss their characteristics in the following. Finally, the classification is not cast in stone. NEAs can change from one type too another due to orbital perturbations caused by the planets.

2.6.1 Amor Type

The Amor type asteroids are a large group of NEAs which do not cross Earth's orbit but can come very close to it. Their perihelia, i.e. the point on their orbit closest to the Sun, lie between 1.017 and 1.3 AU). The lower limit corresponds to Earth's aphelion, its farthest point on its orbit from the Sun, whereas the 1.3 AU limit at the outer fringe is more or less arbitrarily defined as being "close".

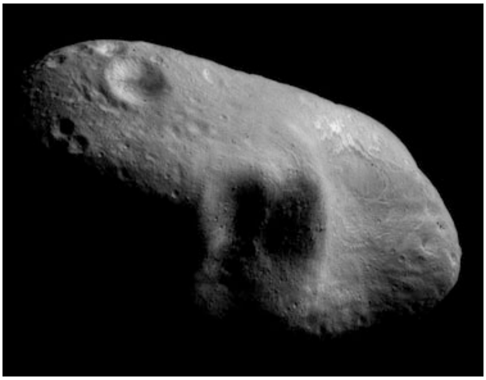


Fig. 2.4 Mosaic of the asteroid Eros taken by the NEAR-Shoemaker spacecraft. Eros' saddle and a shadowed feature to its left, was taken from a distance of 204 km (credit: NASA/NEAR Project (JHU/APL))

The prototype of this group is the asteroid (1221) Amor which was discovered in 1923 and orbits between its perihelion of 1.08 AU and aphelion of 2.76 AU. Today, about 3800 Amor asteroids are known. By the way, as an interesting side note, Phobos and Deimos, the two moons of Mars, are thought to be captured Amor asteroids.

The Amor type asteroids are further sub-divided into four smaller groups, Amor I, II, III and IV.

The Amor I objects have semi-major axes between Earth and Mars (between 1.017 and 1.523 AU) without actually crossing Earth's orbit. About 20 % of all Amors belong to this sub-group. Some of them have such moderate eccentricities that their orbits lie completely within the region between Earth and Mars, thus forming something like an "Earth-Mars belt". One representative of this latter group is the 700 m sized (15817) Lucianotesi, discovered in 1994, which has a semi-major axis of 1.325 AU, a perihelion of 1.168 AU, an aphelion of 1.481 AU and an eccentricity of 0.118.

Other Amor I asteroids cross the orbit of Mars such as the aforementioned Eros. As you can see, NEAs are merely defined by their closeness to Earth and the inner solar system. Whether their orbits stretch much farther towards the outer solar system is not relevant in that respect.

Amor II type asteroids are characterized by longer semi-major axes which reach into the region between Mars and the main asteroid belt (1.52–2.12 AU). Their orbits, however, are only moderately eccentric ($0.17 < e < 0.52$). Yet, their aphelia most often lie within the main belt. As all of them cross the orbit of Mars, they are sometimes titled Mars-crossers. About one-third of all Amors belong to this group. The most famous representative of Amor II asteroids is the name-giver itself, (1221) Amor with a semi-major axis of 1.920 AU (very close to the belt), a perihelion at 1.086 AU (coming pretty close to Earth) and an aphelion of 2754 AU somewhere in the middle of the asteroid belt. With its 1.5 km in diameter, Amor is a relatively large Amor asteroid.

The third type, Amor III, comprises about half of all known Amors, thus being the most prominent group. Their semi-major axes bring them even farther out into the asteroid belt ranging between 2.12 and 3.57. For some of them, the semi-major axes already lie beyond the asteroid belt. Yet, the higher eccentricities of the Amor III compared to the I's and II's still brings them at their perihelion close enough to Earth in order to fall under the classification of an NEA. Furthermore, as their orbital eccentricities are relatively high, the orbits of about one-third of all Amor III asteroids stretch beyond the asteroid belt and come about 1 AU close to Jupiter, like (719) Albert and the largest known NEA (1036) Ganymed. You should, however, not confuse the latter one with the Galilean Moon Ganymede. The asteroid has a diameter of about 32 km and was discovered in 1924. Its semi-major axis is at 2.662 AU. Its eccentricity of $e = 0.537$ brings it as close as 1.233 AU to the Sun and at its aphelion to 4.091 AU.

Many of the Amor III asteroids even go further and cross the orbit of Jupiter, such as (5370) Taranis which reaches its aphelion at 5.454 AU and its perihelion at 1.221. Its semi-major axis is at the fringe of the asteroid belt with a distance of 3.338 AU.

There is one problem that becomes apparent when looking at the Amor IIIs: the classification of asteroids into groups is not clear cut at all. Asteroids can belong to different groups at the same time making the classification a bit fuzzy. As the Amor IIIs potential have parts of their orbits in the asteroid belt, several of them have also been assigned to subgroups of the belt, e.g. The Alinda group which comprises asteroids being in 3:1 resonance with Jupiter and 1:4 resonance with Earth. Their semi-major axes are at about 2.5 AU and Jupiter perturbs their orbits and their eccentricities increase. One day, if they come close enough to one of the inner, terrestrial planets they will be freed from the resonance with Jupiter.

The last group is Amor IV, a very small group of about 14 asteroids. Their semi-major axes are well beyond the asteroid belt with more than 3.57 AU. All of them are Jupiter-crossers and have very high eccentric orbits ($0.65 < e < 0.75$). The only named Amor IV asteroid is (3552) Don Quixote, and is in some respects an awkward asteroid. Its orbit is with about 30° highly inclined towards the ecliptic plane and is more similar to that of a comet. Astronomers assumed that it might be an extinct comet. We will see later that when comets run out of volatile compounds on their surfaces after many passes of the Sun, they become burnt out and exhibit no further characteristics of a comet such as a tail or coma when they travel again into the inner solar system. Scientists believe that quite a few objects we consider as asteroids today are in fact extinct comets that have reached their end of their cometary existence. However, in 2013, the Spitzer Space Telescope (see Fig. 2.5) detected a faint coma and tail in infrared light. Most likely, these are created by CO_2 sublimation, i.e., frozen carbon dioxide ice on the surface of the body is warmed up by solar radiation and ejected into space in the form of CO_2 gas, thereby forming the coma and subsequently the tail.

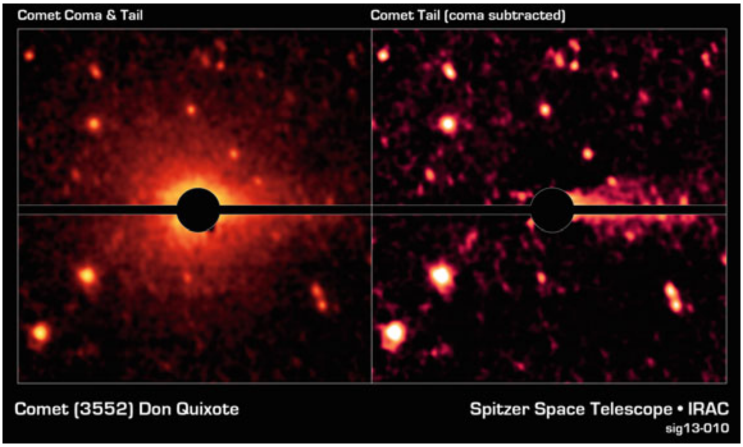


Fig. 2.5 Asteroid Don Quixote turns out to be a comet. The *left image* shows Don Quixote's coma and tail as revealed in infrared light by Spitzer. The *right image* represents a more elaborate image processing step, in which the glow of the coma has been removed based on a model comet coma. *Bright speckles* around Don Quixote are background stars; the *horizontal bar* covers image artifacts caused by the image processing (credit: NASA/JPL-Caltech/DLR/NAU)

2.6.2 Apollo Type

The Apollos are named after (1862) Apollo discovered in 1932 and have semi-major axes larger than 1 AU, so beyond the orbit of Earth. Yet, their perihelia are located at distances smaller than 1.017 AU, Earth's aphelion. Hence, many of the Apollo's are Earth-crossers and depending on their actual orbits (including the inclination), they may potentially become dangerous to Earth. The asteroid (1862) Apollo itself has a semi-major axis of 1.471 AU, an aphelion of 2.295 AU and a perihelion of 0.647 making it also an Earth-crosser.

On February 15, 2013, a meteor exploded over the Russian city of Chelyabinsk and caused severe damages to the infrastructure in this Siberian city (see Fig. 2.6). The provenance of this meteor was later found to be an Apollo asteroid; it exploded approximately 40 days after its perihelion. The asteroid 2011 EO_{40} is one of the candidates proposed for the role of the parent body of the Chelyabinsk meteor. Other published orbits are similar to the 2-km-diameter asteroid (86039) 1999 NC_{43}. This shows impressively what Earth-crossers may be possible to end up.

The largest known Apollo and also the largest known Earth-crosser is (1866) Sisyphus with an approximate diameter of 8.5 km. To put this into perspective, the object irrespective of whether it was a comet or asteroid that collided with Earth 65 million years ago and triggered the extinction of the dinosaurs, was comparable in size.

The origin of the Apollos is believed to at least partially lie in the main asteroid belt. The asteroids there may be perturbed by the influence of Jupiter's gravitation. Some of them may finally end up as Apollos, such as e.g. (4179) Toutatis. In

Fig. 2.6 Trails of the Chelyabinsk meteor that exploded on February 15, 2013 (credit: Alex Alishevskikh; CC BY-SA 2.0)

particular, the orbits of asteroids being in 3:1 mean motion resonance with Jupiter are heavily disturbed and become more and more eccentric. One day, when they come close to one of the inner planets this resonance with the gas giant will be unlocked and they may turn into an Apollo.

2.6.3 Aten Type

The Aten asteroids are named after (2062) Aten discovered in 1976 and are characterized by semi-major axes smaller than 1 AU, and aphelia larger than 0.9833 AU, a distance which corresponds to Earth's perihelion distance. They are thus by definition Earth crossers and can also, like the Apollos, become potentially dangerous to the blue planet. Some of them have their perihelia well inside the orbits of Mercury and Venus and thus are also Mercury- and/or Venus-crossers. Hence, they may be collision partners to most of the inner planets.

Quite a few had recent close encounters with Earth. On 15.2.2013, asteroid (367943) Duende passed Earth at a distance of 27,599 km. This is only about one-tenth of the distance between the Earth and Moon and inside the orbit of geostationary satellites. It literally scratched our home planet. Yet, the consequence of a collision would have been moderate, as the asteroid had a mere diameter of about 40 m (see Fig. 2.7). When Duende approached Earth it had been classified as an Apollo asteroid. However, due to the very close encounter with Earth, its orbit was heavily perturbed now rendering it an Aten type asteroid.

A similar, potentially dangerous Aten asteroid is (99942) Apophis, which was discovered in 2004 and would be a far greater threat with its diameter of 325 ± 15 m. Soon after its detection first estimations of its orbit suggested an extremely close flyby of Earth on April 13, 2029. Looking at the statistical variances in these early

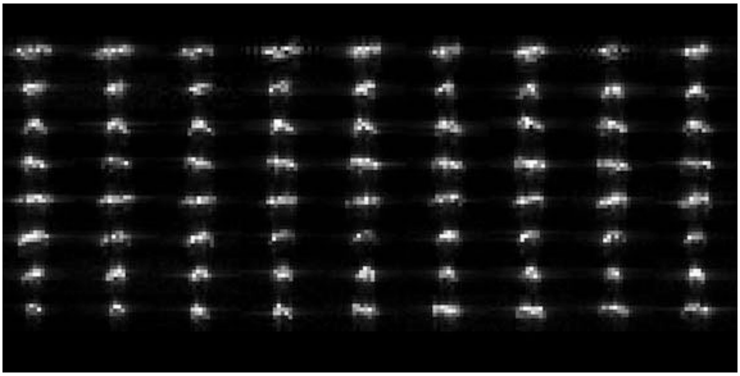


Fig. 2.7 Initial radar images of Asteroid 2012 DA_{14} Duende. The data were collected on the night of Feb. 15 to 16, 2013, after the asteroid had made its closest approach to Earth and was exiting the Earth-moon system. During the observations, the space rock's distance to Earth increased from 120,000 to 314,000 km (credit: NASA/JPL-Caltech)

estimations, a collision could not be excluded. The consequences thereof would have been devastating.

The impact energy of a collision would be close to 900 megatons of TNT. The impact energy of the Tunguska event is estimated to be in the 3–10 megaton range. The biggest hydrogen bomb ever exploded, the Tsar Bomba, was around 57 megatons. The energy of 900 megatons roughly corresponds to an earthquake of strength 9.0 on the Richter magnitude scale. The earthquake that triggered the devastating tsunami on 26.12.2004 had about the same magnitude.

Later, more reliable computations revealed that Apophis will indeed be passing Earth closely on that date yet well beyond the orbits of geostationary satellites at a distance of about 31,000 km. Still pretty close but an impact can be ruled out. However, the asteroid will give a stunning performance on the sky that day and will be with approximately 3.3 mag of brightness an easy naked-eye object. We will be able, clear skies provided, follow the fast moving object on the night sky during its passage. Though Apophis will not harm Earth, its encounter with our home planet will have dramatic consequences for itself. Its orbit will be heavily perturbed and eventually kick the asteroid out of the Aten group and make it an Apollo type asteroid. Its semi-major axis of 0.92 AU will change to a semi-major axis of 1.1 AU.

2.6.4 Atira Group

The Atira group is (still) a relatively small group of asteroids that orbit within the orbit of Earth. They are not Earth-crossers and thus can be classified as Inner Earth Objects or asteroids (IEA). They are named after (163693) Atira, the first asteroid ever discovered being an IEA. All members, except one are larger than 300 m. Currently, the group is constituted of Atira, 1998 DK_{36}, (164294) 2004 XZ_{130}, 2004

JG_6, 2005 TG_{45}, 2006 KZ_{39}, 2006 WE_4, 2007 EB_{26}, 2008 EA_{32}, 2008 UL_{90}, and 2010 XB_{11}.

2.7 The Main Asteroid Belt

The main asteroid belt, roughly lying between the orbits of Mars and Jupiter (2.0–3.4 AU), is probably the most famous, also in public notion, asteroid concentration in our solar system. It comprises the vast majority of all known asteroids with well above 600,000.

This sounds like a huge number and therefore, we may easily be deceived to conclude this would also mean a high mass concentration in the belt. The contrary, however, is true. Though there are so many bodies, about half of the total mass is concentrated in the four largest objects of the belt: the asteroids Vesta, Pallas and Hygiea and the belt's only dwarf planet Ceres. The latter one is also the largest object in the belt with a diameter of approximately 950 km. Vesta, Pallas and Hygiea are about half that size with respective diameters of about 400 km. Most of the remaining asteroids are by far smaller in size ranging down to the size of dust particles. The total mass of the whole asteroid belt corresponds to roughly 5 % the mass of our Moon or roughly one-third that of Pluto.

Another common misassumption is that since there are so many objects, collisions should be very likely and an unhindered crossing very difficult. Yet, in fact, it is mostly empty space. The asteroids are highly scattered on various orbits and cover a larger zone and hence very thinly distributed. Many spacecraft have passed this area so far without having any significant problems or encountering impacts.

We have already discussed the origin and evolution of the asteroid belt previously in this chapter and thus will not repeat it. We will, however, see in the following that though there are a huge number of objects in the zone of roughly 2.0–3.4 AU, these are not evenly distributed.

2.7.1 Dimensions and Distribution Within the Belt

Jupiter has a tremendous influence on the asteroid belt. The gravitational effects induced by its huge mass can easily perturb orbits of the asteroids. In particular, destabilizing mean motion resonances, i.e. when the mean orbital period of an asteroid is an integer fraction of the orbital period of Jupiter, play a very important role.

In some of these regions, no stable asteroid orbits are possible. Jupiter's gravity perturbs the orbits of these bodies forcing them into unstable, random orbits with larger semi-major axes. Hence, the area gradually is thinned out and gaps develop. These gaps are named after the American astronomer Daniel Kirkwood (1814–1895) who discovered them while studying the distribution of semi-major axes of

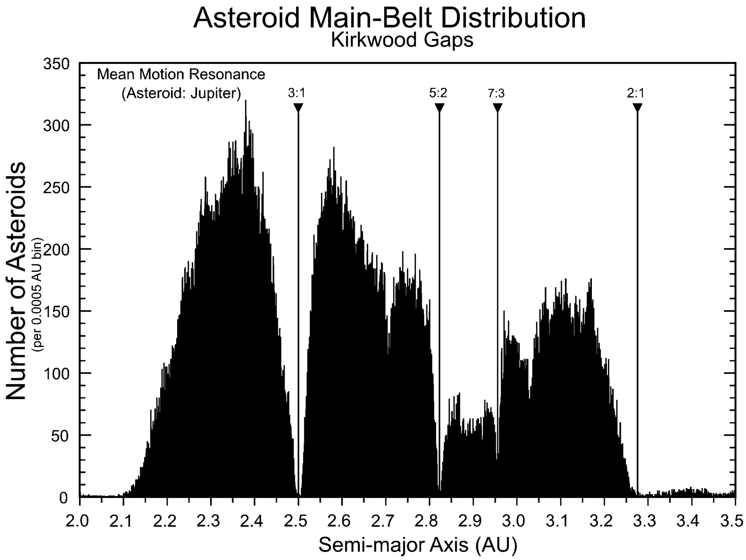


Fig. 2.8 This histogram shows the primary Kirkwood gaps in the asteroid main-belt. These gaps (labeled “3:1”, “5:2”, “7:3”, “2:1”) are caused by mean-motion resonances between an asteroid and Jupiter (credit: Alan Chamberlain, JPL/Caltech)

asteroids in the belt. He noted “gaps”, i.e. missing occurrences of objects at certain distances from the Sun (see Fig. 2.8). He further investigated this phenomenon and found out that they were located at positions where the period of revolution about the Sun was an integer fraction of Jupiter’s orbital period. Kirkwood proposed that the gravitational perturbations of the gas giant led to the removal of asteroids from these orbits. This should later on turn out to be true.

There are many resonances and mostly smaller gaps within the asteroid belt. Yet, a few stand out. You can also easily find them in Fig. 2.8. There is a strong resonance at 4:1, i.e. while objects there orbit the Sun four times, Jupiter does ones. The gap is located at 2.06 AU and forms the inner border of the asteroid belt towards the inner solar system. Another gap is at 3:1 resonance, the so-called Hestia gap. Then, gaps at 5:2 and 7:3 follow. The Hecuba gap at the 2:1 mean motion resonance (about 3.28 AU distance from the Sun) forms the outer border of the asteroid belt.

Also taking these gaps into account, the asteroid belt is often divided into three distinct zones. Zone I, the inner asteroid belt, reaches from the 4:1 resonance at 2.06 AU to the 3:1 resonance at 2.5 AU. It is followed by zone II, the middle asteroid belt, ranging from the edge of zone I to the 5:2 resonance (2.82 AU). Zone III, the outer asteroid belt subsequently stretches from the end of zone II to the 2:1 resonance (3.28 AU).

2.7.2 Asteroid Families or Asteroids Are Seldom Alone

Even within these zones, or more refined between the gaps, the asteroids are not evenly distributed but certain concentrations of these occur. In 1918, the Japanese astronomer Kiyotsugu Hirayama (1874–1943) noticed that some asteroids had similar characteristics compared to others, such as the same semi-major axes, similar inclination and eccentricities as well as a comparable chemical composition. Hirayama concluded that these bodies might have a common origin and, subsequently, decided to group such asteroids into asteroid families. Today, we count about 20–30 families in the asteroid belt, depending on the definition. Some families still need to be verified. Among them are a few large families whose numbers of members dominates the population of the belt. We will have a brief look at some of the most relevant families in the following.

In the inner belt, we find the Flora family, a relatively large group of asteroids orbiting in a distance between 2.15 and 2.35 AU from the Sun. Their orbits are only slightly inclined between 1.5 and 8° towards the ecliptic plane. About 4–5 % of all asteroids in the belt belong to this family.

Another dominant group is the Vesta group whose members have semi-major axes ranging from 2.25 to 2.5 AU and have orbital inclinations between 5 and 8°. This group is even larger than the Flora family, uniting about 6 % of all asteroids in it. The composition of this group differs from the Floras in that the latter ones are a relatively homogenous group of asteroids having comparable size. The Vestas, however, are dominated by a large member, which is also its namesake, Vesta (see Fig. 2.9), which is clearly larger than all the other members. Scientists thus assume that both groups have a different origin. In the case of the Vesta group it is proposed that it is the result of a cratering event. Yet, we will come back to that in a minute.

The Nysa group is a smaller group within the inner belt whose members occupy orbits at similar distances as the Vestas but have lower inclinations ($1° < i < 5°$).

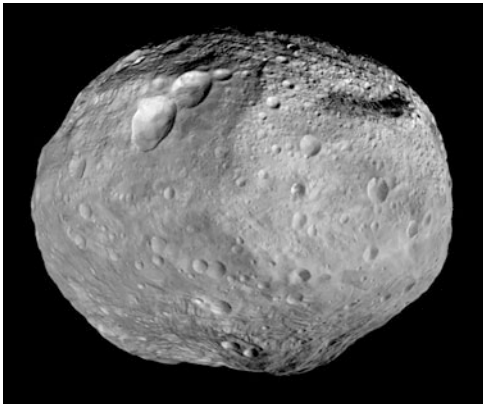


Fig. 2.9 Mosaic of Vesta taken by NASA's Dawn spacecraft (credit: NASA/JPL-Caltech/UCAL/MPS/DLR/IDA)

The two dominating families of the middle belt are the Eunomia and the Gefion groups. Both are located behind the Hestia gap. The semi-major axes of the former one's members vary between 2.5 and 2.8 AU, whereas the latter ones lie in a region stretching from 2.7 to 2.8 AU. The Eunomias orbits are inclined 11–16° towards the ecliptic plane while the Gefions are slightly less inclined ($7.5° < i < 16$). About 5 % of all main belt asteroids are considered to belong to this group.

The Gefion group is a nice example that it is not sufficient to merely consider the orbital elements in order to determine the membership to a family. Within this family, the orbit of a dwarf planet is located. Yet, the chemical composition of Ceres is completely different from those of the Gefions. A common origin can thus be for sure denied.

Moving farther outwards into the belt, we encounter the Koronis, EOS, Themis and Hygiea families. The orbits of Koronis family are inclined between 0 and about 3.5°. The family comprises most of the asteroids having their semi-major axes in a distance of 2.8–2.95 AU, i.e. between the 5:2 and 7:3 resonances. Its most prominent member is (243) Ida (see Fig. 2.10).

The members of the EOS family orbit in a distance of 2.99–3.03 AU from the Sun and are inclined by 8–12° towards the ecliptic plane. The semi-major axes of the Themis family members stretch from 3.08 to 3.24 AU which are therefore at the outer fringe of the asteroid belt. Their orbits are moderately inclined ($i < 3°$). The Hygiea family, named after their most prominent member Hygiea being one of the largest asteroids in the belt, lies in similar region as the Themis family. Yet, the orbits within this family are more inclined.

2.7.3 Two Families at the Fringe of the Belt

Two larger families exist at the respective borders of the asteroid belt: the Hungaria family at its inner border and the Cybele family at its outer border.

The semi-major axes of the Hungaria family vary between 1.78 and 2.0 AU. The gap at the strong 4:1 mean motion resonance with Jupiter forms their outer border.

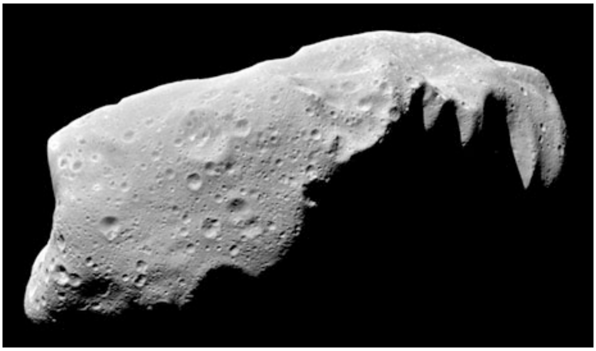


Fig. 2.10 Mosaic of the asteroid (243) Ida of five image frames acquired by the Galileo spacecraft at ranges of 3057–3821 km on August 28, 1993, about 3.5 min before the spacecraft made its closest approach to the asteroid (credit: NASA, Jet Propulsion Laboratory)

Hence, the family officially lies outside the main asteroid belt. Their origin and future is nevertheless closely tied to the belt. However, due to the proximity of Mars whose semi-major axis is at 1.524 AU, it is the red planet that determines their fate at the inner border of the family. In general, the Hungaria family is the innermost and densest group of asteroids.

Its origin is well known while for many other families this still remains to be in the dark. On the one hand, if asteroids are close to the 4:1 resonance with Jupiter, their orbits are strongly perturbed by the gas giant, which then forces them into extremely eccentric and unstable orbits. We have seen this already when we learnt about the Kirkwood gaps. The 4:1 gap thus forms a kind of natural border not only to the asteroid belt but also to the Hungaria family.

On the other hand, asteroids within the 4:1 resonance zone are strongly affected by Mars if their orbits have low inclinations. At high inclinations the influence of the red planet is not strong enough. Mars perturbs their orbits and essentially will kick those asteroids with lowly inclined orbits out of the region leaving only highly inclined objects behind.

This leads to the orbital characteristics of the Hungaria family. As mentioned before, their semi-major axes stretch from 1.78 to 2.0 AU and their orbits are inclined by 16–34° towards the ecliptic plane.

The Cybele family is a group of asteroids slightly outside the main belt with a semi-major axis between 3.27 and 3.7 AU, an eccentricity of less than 0.3, and an inclination of less than 25°. The group is named after the asteroid 65 Cybele. As of 2010, the group is thought to have formed by the breakup of larger object in the distant past. Some of the largest Cybele asteroids are 87 Sylvia, 65 Cybele (the group's namesake), 107 Camilla, 121 Hermione, 76 Freia, 790 Pretoria, and 566 Stereoskopia. Some members of the group also have asteroid moons, such as 121 Hermione.

2.7.4 Formation of Asteroid Families

We have already seen that members of the same family share similar orbital characteristics and are of comparable, if not identical, chemical composition. About 33–35 % of all main belt asteroids belong to families. This shows that the existence of such groups is more than just coincidence. But how did these families develop?

In general collisions between asteroids form the basis for families. We can distinguish two different scenarios. On the one hand, in scenario I, the collision is strong enough in order to completely shatter the parent bodies. This can happen, when the colliding bodies have similar size or the impactor is at least large enough to fully break the other one up. The result of scenario I is a group of relatively small bodies.

In scenario II, on the other hand, the impactor is not massive enough to cause a complete distruction of the parent body. It is more a cratering event in which the

Fig. 2.11 Artist's concept of building planets through collisions (credit: NASA/JPL-Caltech)

impactor collides with the larger parent body, causes a probably large crater and triggers the ejection of material from the surface of the parent body into space (see Fig. 2.11). Thus scenario II results in a family comprising one single large body and a swarm of much smaller ones.

Several scenario II families exist in the main asteroid belt, such as the Vesta family, the Pallas family or the Hygiea family.

Yet, asteroid families do not exist for eternity. On the contrary, they have a limited lifetime of up to several billion years. Looking at the age of our solar system, this may at first glance almost seem to be "eternity". As for all asteroids, the orbits of the members are subject to perturbations by the planets, in particular Jupiter, and other asteroids. Eventually, a family member could be forced to leave the group.

Also collisions play an important role. Colliding bodies, family members with each other or "external" asteroids, may cause a further fragmentation of the objects. At some point in time, they may become so small as to be affected by the solar wind or by the Yarkovsky effect which can then move them out into one of the mean motion resonances with Jupiter.

Once there, as we have seen before, the gravitational influence of Jupiter can ultimately kick them out of the belt.

We should, however, be careful when determining the membership to a family. So-called "interlopers" may exist. These are bodies moving on very similar orbits but indeed to not belong to the family. How can we exclude such "false" members? The answer is by looking at their chemical compositions. As the interlopers and the family asteroids do not have a common origin, it is very unlikely that they share the same chemical characteristics.

We have already learnt about one such interloper: Ceres, which "smuggled" itself into the Gefion family. However, the chemical compositions differ considerably.

2.8 Trojan Asteroids

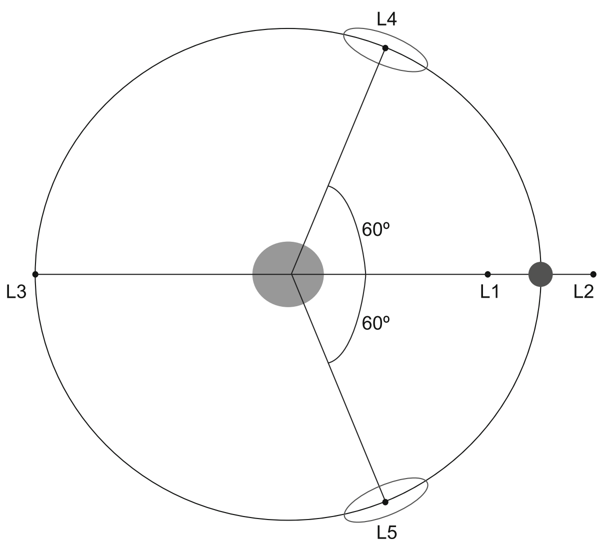
Trojans are a special type of asteroid, not with respect to their composition but for their locations. Contrary to all other asteroids or asteroid-like bodies in our solar system, they share the orbit with a planet or sometimes with a larger moon. They co-orbit the planet around the Sun and are thus in 1:1 mean motion resonance with it. This is atypical since normally objects on a planet's orbit are removed from that over time due to gravitational interactions with the planet. This process is also called "clearing the neighborhood", a definition which will become important when we talk about dwarf planets and their differences compared to a full-blown planet.

So, how can Trojans exist when planets are supposed to throw them out of their orbits? Is this not contradictory?

In fact, there are some exceptions, called the Lagrange points, to the rule of clearing the neighborhood. In 1772, the Italian-French mathematician Joseph-Louis Lagrange (1736–1813) studied a restricted three-body problem. These problems consider three bodies and their respective gravitational interactions. In general, they are difficult to solve. In the restricted case, the mass of one body is negligible because it is too small compared to the others. Hence, we only have to consider the masses of two objects M_1 and M_2. Lagrange determined five special locations (see Fig. 2.12) where the combined gravitational pull of the two large masses provides precisely the centripetal force required to orbit with them. We have already come across these Lagrange points in Chap. 1.

In the present case of Trojan asteroids, the two Lagrangian points L_4 and L_5, sometimes also called the Trojan points, are of special importance. They lie approximately 60° ahead and 60° behind the planet on its orbit. When a planet orbits the Sun, both bodies actually orbit around their common barycenter. Due to

Fig. 2.12 Illustration of the location of the five Lagrange points L_1 to L_5



the much higher mass of the Sun compared to the planet's mass, the barycenter is in this case very close to the Sun's center. Now consider an object with a mass much smaller than that of the Sun or the planet located at either L_4 or L_5. This object is then subject to the combined gravitational force that acts through the barycenter. Hence, the smaller body can co-orbit around the barycenter with the same orbital period as the planet and thus being in 1:1 mean motion resonance with it. This situation can be stable over a very long period of time.

2.8.1 What Are Trojans?: A Bit of Definition

Trojans are a specific type of asteroids, as we have seen, that co-orbit a planet. The first Trojans discovered were those at the L_4 and L_5 points of Jupiter. Thus, for a long period of time, the term "Trojan" merely referred to Jupiter Trojans and was often used synonymously. However, when asteroids co-orbiting other planets in our solar system were discovered, the term was generalized. Today, we understand a Trojan to be an asteroid or asteroid-like body co-orbiting a planet near its L_4 and L_5 Lagrange points, thus either moving ahead the planet or behind it.

Nowadays, Trojans of most planets are known: Venus, Earth, Mars, Jupiter, Uranus and Neptune. Yet, no Trojans have been found so far with Mercury and Saturn. The former is a special case due to its closeness to the Sun. The lack of any Trojans so far for the lord of the rings, Saturn, is not yet understood and remains an unsolved mystery. It is assumed that it is somehow related to the formation of the solar system and the later capture of Trojans.

In the following we will start our journey through these fascinating bodies with the historically first ones that had been discovered—the Jupiter Trojans—before we turn to the other ones. Far less is known about the latter ones than is for Jupiter's companions. Most probably, however, all of them show similar characteristics and may have a comparable origin.

2.8.2 Jupiter Trojans or How It All Began

The Jupiter Trojans are the largest group of objects known at the L_4 and L_5 points of any of the planets. It is assumed that their number may equal that of the number of asteroids in the main belt. These asteroids are distributed in two elongated, curved regions around these points (see Fig. 2.13). The semi-major axes are at about 5.2 AU, which does not come as a big surprise since they co-orbit the giant gas planet. As of January 2015, we know about 6178 Trojans of Jupiter. Yet, how were they discovered? What do we know about them? All this will be subject to the following discussion.

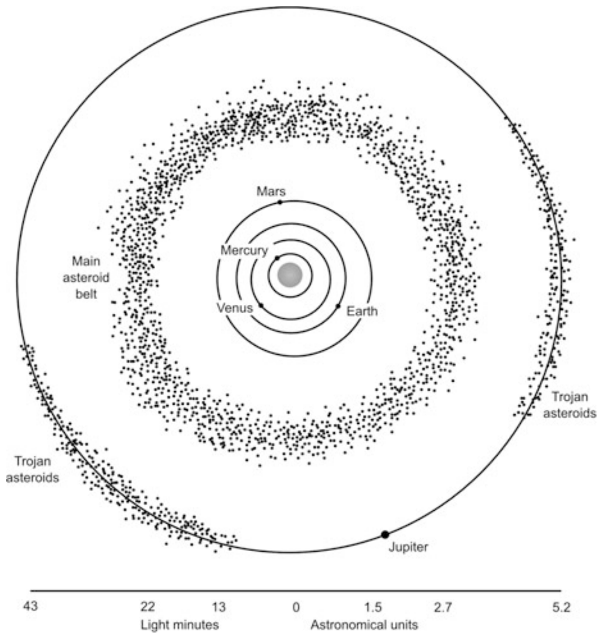


Fig. 2.13 The inner Solar System, from the Sun to Jupiter also showing the asteroid belt and the Jupiter Trojans' Greek and Trojan camps (credit: Mdf; Public Domain)

2.8.3 History of the Discovery of Jupiter Trojans

In 1772, Lagrange studied, as we already know, a restricted three-body problem. During his research, he predicted the existence of small bodies sharing an orbit with a planet being trapped at the L_4 and L_5 points and slowly librating around these positions. Yet, it took about 100 years to prove Lagrange's theory that had raised many doubts in the scientific community at that time.

It was the American astronomer Edward Emerson Barnard (1857–1923) who made the first recorded sighting of a Trojan, (12126) 1999 RM_{11} in 1904. Unfortunately, he did not recognize this object as a Trojan. Rather he thought it to be the newly discovered moon of Saturn, Phoebe, which was very close to the actually observed location that night. The true identity of Barnard's object was only revealed in 1999 when astronomers established its orbit, which clearly showed its Trojan nature.

Thus, the honor and merits of discovering the first recognized Jupiter Trojan passed over to the German astronomer Max Wolf (1863–1932). In 1906, while he was working at the Heidelberg-Königstuhl State Observatory, he identified (588) Achilles close to Jupiter's L_4 point.

Further Trojans soon after followed discovered by the German astronomer August Kopff (1882–1960) in 1906–1907: (624) Hektor at L_4 and (617) Patroclus being the first L_5 Trojan.

2.8.4 Naming of Jupiter Trojans or Greece vs. Troy

In general, the naming of asteroids with conventional names is no longer limited to any specific names. Arbitrary names can be and have been chosen. Yet, there is an important exception. The nomenclature of Trojan asteroids is restricted to heroes of the Trojan War as told by Homer in his epos “Ilias”.

In the case of the Jupiter Trojans, the actual name selection depends on the location. All asteroids located near the L_4 point belong to the so-called Greek camp. Hence, names of Greek heroes are to be chosen. Bodies at L_5 are named after Trojan heroes as this group is considered to be the Trojan camp.

Yet, there are two exceptions to this naming convention. (617) Patroclus, a Greek hero, is located at L_5 and thus in the Trojan camp. On the other side, (624) Hektor, the famous Trojan hero that had been defeated by Achilles, belongs to the Greek camp at L_4. How did that come about? The explanation is quite simple. They were named before the nomenclature was introduced.

2.8.5 Origin of Jupiter Trojans

The Jupiter Trojans are in its majority dark or dark reddish bodies with a quite irregular shape. They show undoubtedly some similarities to trans-Neptunian objects but also share characteristics with asteroids from the outer main asteroid belt. Thus, what do we know about their origin? How did they get into their current orbits? Did they form in situ?

Two main theories of their origin exist while the second one involving the Nice model seems to be the more promising one though it still lacks to show all details, the first one is given for the sake of completeness.

The first model assumes that the Trojans formed in the same part of the solar system as Jupiter did and entered their current orbits while the gas giant was forming. According to modern theories, in the last stage of Jupiter’s formation process a runaway growth of its mass took place through the rapid accretion of large amounts of hydrogen and helium that were still present in its neighborhood. Jupiter soaked them up like a sponge. During this phase, which only lasted about 10,000 years, Jupiter’s mass increased by a factor of ten.

This huge mass increase led to capturing still existing nearby planetesimals. The capture process, according to the model, appears to have been quite successful since about 50 % of the captured bodies still should be at their positions.

This model, however, introduces some drawbacks, which cannot be easily overcome. First, the number of captured bodies the model predicts exceeds the actually observed population of Jupiter Trojans by about four orders of magnitudes. Where did all of the excess bodies go? Orbital perturbations which can still happen at the L_4 and L_5 Lagrange points cannot be hold accountable for all of the lost asteroids when considering the limited time span since their capture. Secondly, the orbits of the observed Trojans have a larger inclination than they should have according to the model's predictions. Thirdly, the same mechanism should also have caused the formation of Saturn Trojans. Yet, so far none have been found and it is still unclear whether Saturn seems to be an exception.

The problems just mentioned led to new models and proposals that have tried to overcome them. The above-mentioned second model is based on the Nice model which we have already touched a few times and which will be in particular important in relation to trans-Neptunian objects.

According to this model, the Jupiter Trojans were captured during the phase of planetary migration, which is supposed to have taken place 500–600 million years after the formation of the solar system. We have seen that when Jupiter and Saturn reached their 2:1 mean motion resonance a process was triggered that drastically changed the outer solar system. Uranus, Neptune and Saturn moved farther outward while Jupiter slightly inwards. This caused further strong disruptions in this region for the small bodies that still existed there, e.g. in the proto-Kuiper belt. Many of them were thrown towards the inner solar system. Inevitably, they would have come close to the outer planets and potentially close to the L_4 and L_5 points. Their combined gravitational force in addition to the influences inhibited by the migrating gas giants would have perturbed and destabilized the orbits of any Trojans that had existed there during that period. During this time, the Trojan co-orbital region is termed "dynamically open". Planetesimals from the proto-Kuiper belt wandering towards the inner solar system may have crossed this region and temporarily inhabited it. After the migration of the planets came to an end and thus the period of orbital instability had ended, the Trojan region is "dynamically closed" again, i.e. any planetesimals or bodies captured then will stay. Other than in the first model, the Nice model suggests that asteroids scattered from the outer regions of the primordial asteroid belt (at distances greater than 2.6 AU from the Sun) were captured as Trojans by Jupiter. This matches quite well with the characteristics we know of the Trojans: the eccentricity, their highly inclined orbits and their libration angle. Furthermore, many observed Jupiter Trojans seem to be of D-class asteroids, which resembles to asteroids of the outer main belt. Today's TNOs in contrast appear to have a different composition.

Yet, final evidence is missing and no one can tell today which model is true, if at all. Further research is required. Still the main problem remains that Trojans are very difficult to observe due to their extreme faintness.

2.8.6 Physical Characteristics of Jupiter Trojans

Unfortunately, only very little is known about the masses, chemical composition, rotation periods and other physical characteristics of the Jupiter Trojans. Most of them are dark, dark reddish or neutral, featureless bodies having an irregular shape. Their albedo, i.e. the ratio of reflected radiation from the surface to incident radiation upon it, is relatively low ranging between 3 and 10 %. No major differences can be seen between the two swarms.

Spectroscopic analysis revealed that they are mostly D-type asteroids, i.e. of the spectral type that also predominates the outer regions of the main asteroid belt. Only a smaller number is classified as C- or P-type. Scientists were not able to detect any direct evidence of water ice or any organic compounds. Yet, indications exist. The Trojan (4709) Ennomos was found to have very high albedo of 0.18 compared to other Trojans. This may indicate the presence of water ice, which has a higher reflectivity rate than many other compounds.

(911) Agamemmnon and (617) Patroclus show very weak absorptions at 1.7 and 2.3 μm, which may hint to the existence of some organic compounds.

Furthermore, it is interesting to observe that the spectra of Jupiter Trojans are very similar to those of the irregular shaped moons of the gas giants and of cometary nuclei. This suggests that all of them may have a common origin.

This assumption was further hardened when scientists were able to measure the density of Patroclus in 2006. It seems that the Trojan's density is lower than that of water ice (0.8 g/cm^3 compared to 0.9167 g/cm^3 of ice). Such a result more resembles to trans-Neptunian objects or comets than to main belt asteroids. However, one of the first Trojans discovered, Hektor, appears to have a density of 2.480 g/cm^3 which speaks more for asteroid origin.

To make a final determination more results are needed. Looking at the assumed origin of the Trojans according to the Nice model, it may be true that they are of both asteroid- and cometary origin and that the population is simply diverse. Those scattered objects from the outer main belt that were captured as Trojans by Jupiter may have had originally originated either from the primordial asteroid belt, hence having a higher density, or be some of the proto-Kuiper belt objects that had been injected to the inner solar system during the period of planetary migration.

2.8.7 Orbits of Jupiter Trojans

Not surprisingly the orbits of the Jupiter Trojans have semi-major axes of 5.2 ± 0.15 AU, as they co-orbit Jupiter and are in 1:1 resonance with the planet. They are distributed throughout elongated, curved regions around the L_4 and L_5 Lagrange points (see Figs. 2.12 and 2.13). Each of the two swarms stretches for about 26° along the orbit of Jupiter from these points.

The most striking aspect is that their orbits are highly inclined towards Jupiter's orbital plane with inclinations ranging up to 40°. The Trojans, however, are not fixed at their positions but slowly librate around their respective equilibrium points, i.e. either L_4 or L_5. During this process they periodically move closer to and farther from Jupiter.

2.8.8 Trojan Families

We know that families exist in the main asteroid belt. Do Jupiter Trojans also form families? Astronomers have shown that at least a dozen families exist. However, families and their respective members are much more difficult to identify than in the asteroid belt. Why is this?

Jupiter Trojans are locked within a by far narrower region than the relatively widely scattered asteroids in the main belt. The swarms are much more compressed. Consequently, families tend to overlap and be in close neighborhood with other non-family Trojans. Furthermore, due to their faintness it is not easy to individually observe them and measure individual spectra. Thus, it remains a challenging task to define family members.

From the groups known so far, we know that they are much smaller in size than respective families in the main belt. David Jewitt showed in 2004 that the currently largest group, the Menelaus family, consists of only eight members.

In addition, the Trojan families are more dynamic than their main belt counterparts. Since the density of Trojans is high in a small region, collisions take place more frequently thereby ejecting members from a group or degrading them.

2.8.9 How Many Jupiter Trojans Exist?

A definitive answer to this question cannot be given. Today we know of about 6,800 Jupiter Trojans but the actual total number is estimated to be much higher. Some models suggest that the L_4 swarm consists of 160–240,000 asteroids with a diameter larger than 2 km and about 600,000 with a diameter larger than 1 km. It is very likely that L_5 hosts a comparable number of asteroids. This would mean that the number of Jupiter Trojans is similar to the number of asteroids in the main asteroid belt.

This number is based on the assumption that all Trojans have a similarly low albedo between 0.03 and 0.1.

However, it has recently turned out that this may not be true for all of them, as can be seen by the high albedo of Ennomos with a value of 0.18. This suggests that the number of Trojans may be overestimated. Further albedo measurements of many more Trojans are required in order to arrive at a statistically reliable estimation.

2.8.10 Venus Trojans

Many of the Trojan populations of other planets in our solar system are not as all as impressive, also in terms of numbers, as the Jupiter Trojans are. Nevertheless, they are worth a visit. If we look into the inner solar system, we can identify Venus to be the closest planet to the Sun that has Trojan asteroids. Although for a very long time

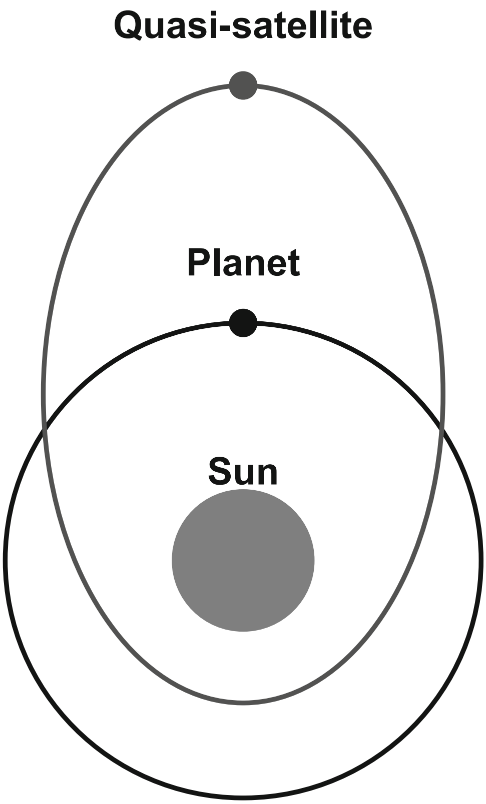


Fig. 2.14 Illustration of the orbit of a Quasi-Satellite (credit: Bryan Derksen; Public Domain)

it had not been clear whether any exist. Some quasi-satellites of Venus have been known since 2001: (322756) 2001 CK_{32}, 2002 VE_{66} and 2011 XE_{133}.

A quasi-satellite is different from a true satellite (e.g. a moon) in that it is an object in a specific type of co-orbital configuration (1:1 orbital resonance) with a planet where the object stays close to that planet over many orbital periods. The quasi-satellite's orbit around the Sun takes exactly the same time as the planet's, but has a different eccentricity (see Fig. 2.14). In contrast to true satellites, quasi-satellite orbits lie outside the planet's Hill sphere, i.e. the zone of gravitational influence of the planet on smaller bodies, and are thus unstable. Over time they tend to evolve to other types of resonant motion, where they no longer remain in the planet's neighborhood, then possibly later move back to a quasi-satellite orbit, etc.

In July 2013, the first and so far only true Venus Trojan, though only being a temporary one, was found by Pan-STARRS at Venus' L_4 point: 2013 ND_{15}. It is originally an Aten asteroid and its semi-major axis (0.7235 AU) is very similar to that of Venus but it has a high eccentricity (0.6115) and small orbital inclination (4.794°). ND15 currently has a perihelion of 0.2810 AU and an aphelion of 1.1660 AU which also makes it, by definition, a NEA and a Mercury-, Venus- and Earth-crosser. Its diameter is supposed to be in the range of 40–100 m.

Predictions of its orbit show that it will leave its current Trojan orbit in about 500 years and may become a quasi-satellite of Venus. Yet, it cannot be ruled out that 1 day it will be re-established as a temporary Venus Trojan.

2.8.11 Earth Trojans

Currently, only one Earth Trojan is known. 2010 TK_7 is located at Earth's L_4 point and has an approximate diameter of 300 m. The Wide-field Infrared Survey Explorer (WISE) spacecraft discovered it in October 2010. It is only weakly bound to the L_4 point and periodically librates between the L_4 and L_3 points over a period of about 390 years. During its move on that very elongated loop it can reach distances ranging from 20 million kilometers from Earth up to 300,000 million kilometers (when being close to L_3, i.e. almost at the opposite side of the Sun with respect to Earth).

Its semi-major axis is currently at 1.0004 AU. This means, however, that the Trojan is slightly more distant from the Sun than our home planet and thus has a slightly slower orbital velocity than Earth following Kepler's laws of planetary motion. Currently, it drifts towards Earth on its libration orbit. The interaction with our planet continuously reduces TK_7's semi-major axis which will increase its orbital velocity. This increase, in turn, will eventually make it move faster away from Earth. In 2209, TK_7 will be close to Earth's L_3 point.

Figure 2.15 shows the position of the loop in selected future years, changing from its current position just ahead of the Earth to its farthest position from our planet in the year 2209. After that, the loop reverses its drift and slowly moves back towards the Earth. It gets back to its 2011 position around the year 2400, completing a full libration period. Numerical studies show that the loop will librate this way and remain on the leading side of the Earth for at least the next several thousand years.

2.8.12 Mars Trojans

Not many Mars Trojans are known. Their characteristics and origin still remain in the dark. As of today, seven possible Trojans of the red planet are known. Yet, the Minor Planet Center only has officially acknowledged four of them.

The first Mars Trojan was discovered by David H. Levy and Henry Holt at Palomar Observatory in 1990: (5261) Eureka. It is located at Mars' L_5 point and its infrared spectrum is typical for an A-type asteroid. Two other officially acknowledged Trojans exist at the L_5 Lagrangian point: (101429) 1998 VF_{31}, (311999) 2007 NS_2. Though not being "official" Trojans, three other asteroids can be found there whose orbits indicate after numerical studies to be Trojan-like: 2001 DH_{47}, 2011 SC_{191}, 2011 UN_{63}. Only one asteroid at L_4 is known, (121514) 1999 UJ_7.

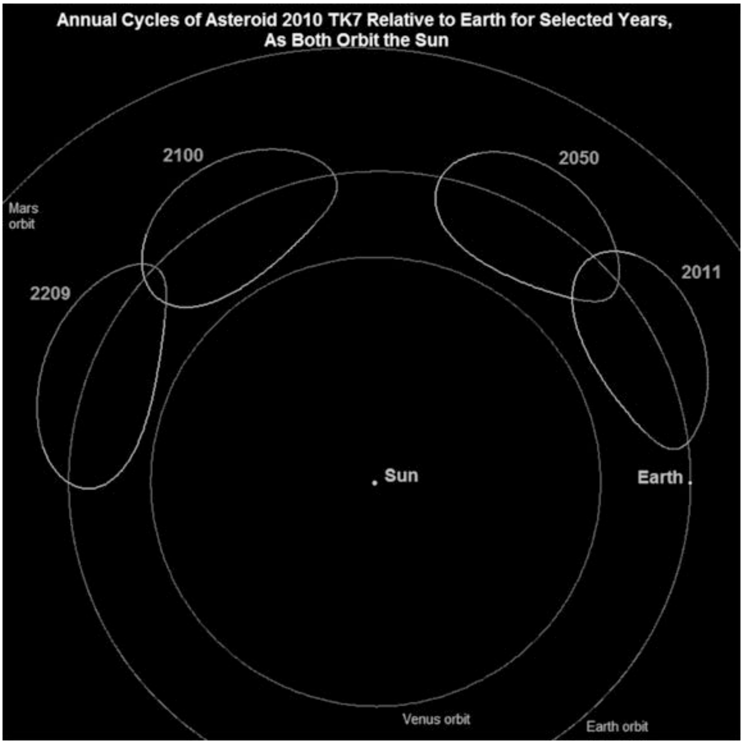


Fig. 2.15 The diagram shows the position of the annual cycle of 2010 TK_7 in selected future years, changing from its current position just ahead of the Earth to its farthest position from our planet in the year 2209. After that, the cycle reverses its drift and slowly moves back towards the Earth. It gets back to its current 2011 position around the year 2400, completing a full period of what is called "libration" (credit: NASA, Paul Chodas & Don Yeomans, NASA/JPL Near-Earth Object Program Office)

Not much is known about them, as they are difficult to observe. It is also not clear how they came into their current positions. How did Mars capture them? The mechanisms which led to the Jupiter Trojans and probably also the Neptunian Trojans do not seem applicable in the case of Mars as its mass is by far too low. Other processes are therefore discussed. One model, mentioned by Alessandro Morbidelli, assumes that in the final stage of terrestrial planet formation, the orbit of proto-Mars was possibly perturbed by lunar-sized bodies. These larger bodies were remnants of the planetary accretion process. These bodies collide finally with either of the terrestrial planets or they will be ejected out of the Solar System at the end of the heavy bombardment phase.

A major consequence of close encounters, or even impacts, between proto-Mars and these lunar-sized bodies is a chaotic wandering of the proto-planet resulting in a shifting of its semi-major axis. During each change of semi-major axis, the regions surrounding both L_4 and L_5 move with the planet to a new location possibly populated by small planetesimals. These were probably leftovers of the runaway growth of planetary embryos. A fraction of the local planetesimals may be trapped and become Mars Trojans. Such planetesimals might have inclinations between 15 and 30°, which is necessary to become a stable Mars Trojan, otherwise stable orbits are not possible. We have already encountered this problem of high inclinations when we discussed the Hungaria family, a few sections above.

2.8.13 Saturn Trojans: Where Are They?

If we go farther out into the solar system and skip Jupiter, since we already dealt with the gas giant and its Trojan populations, we can first make a stop at Saturn. To the surprise of many astronomers, no Trojans seem to exist with the ringed planet. The reasons for that are not yet understood.

The two dominant gas giants in our solar system, Jupiter and Saturn, are believed to have formed in the same way, by accretion of an icy rocky core followed by the massive accretion of nebular gas. During the growth of the two planets, Trojans might have been captured from a reservoir of planetesimals near the planet's orbit by different mechanisms or during the phase of planetary migration as we have seen and discussed with respect to the Jupiter Trojans according to the Nice model.

In the case of Jupiter this has led to the formation of a large population of Trojans comparable in number to the main belt asteroids. Hence, if we assume similar formation processes Saturn also could have captured Trojans like Jupiter. However, no Saturn Trojans have been observed by now.

Although their discovery and observation is more challenging than with the Jupiter counterparts due to the greater distance, it should be still achievable if we assume similar sizes as for the Jupiter Trojans. That also means that if the Saturn Trojan population were as numerous as the Jupiter Trojan population, at least one Saturn Trojan should have been discovered until now.

As this is not the case, scientists have discussed various reasons for their absence. Dynamical instability could simply be the reason for the apparent absence of Saturn Trojans. Another reason might be that there were only very few planetesimals in the reservoir region around Saturn. The migration of Saturn itself could be another reason. The planet would have continuously gained and lost asteroids while migrating.

In addition, studies on the stability of Saturn Trojans have shown that the stable regions near the Lagrangian points L_4 and L_5 of Saturn are very small compared to the Jupiter Trojan region. Orbital perturbations due to the near 2:5 mean motion resonance between Jupiter and Saturn may be one reason for this.

The orbit of a Saturn Trojan would usually become unstable because its eccentricity is increased due to these perturbations until the Trojan has a close encounter with Saturn. The L_4 and L_5 regions of Saturn are far more chaotic as their Jovian counterparts, which is at least partially owed to the overlap of 2:5 resonance with

Jupiter and the 1:1 mean motion resonance with Saturn. This result indicates that the Saturn Trojan region is intrinsically unstable because of chaotic effects.

Hence, there seems to be only a very small region restricted to small eccentricities and inclinations where orbits may survive over timescales comparable to the age of the solar system. The statistical probability that Trojans wit exactly these characteristics exist is quite low and may therefore explain their absence.

2.8.14 Uranus Trojans

As of 2015, only one Trojan companion of Uranus is known. 2011 QF_{99} was discovered in 2011 during a deep survey of trans-Neptunian objects conducted with the Canada–France–Hawaii Telescope which is located near the summit of Mauna Kea in Hawaii not far away from the Keck telescopes.

2011 QF_{99} is believed to be about 60 km in diameter and temporarily orbits near Uranus's L_4 Lagrangian point, i.e., ahead of Uranus. It will continue to librate around L_4 for at least 70,000 years and will remain a Uranus co-orbital for up to 3 million years before becoming a centaur, a special class of objects we will deal with soon which share characteristics of comets and asteroids together.

Uranus trojans are generally expected to be unstable and none of them are thought to be of primordial origin for similar reasons as given for the Saturn Trojans.

2.8.15 Neptune's Trojan Asteroids

Compared to the Trojans of other planets we have learnt about before, the Neptunian population is believed to be much richer, though only a few have been discovered so far. The main reason for the low discovery rate is supposed to be due to the large distance from Earth of these small and faint objects.

Currently, 12 Neptune Trojans are officially confirmed, nine at the planet's L_4 Lagrangian point and 3 at L_5. The first one that was discovered is 2001 QR_{322} at L_4 by the Deep Ecliptic Survey. Its diameter is supposed to be in the range of 60–160 km.

In 2005, the discovery of the high-inclination Trojan 2005 TN_{53} has indicated that the Neptune Trojans populate thick clouds, which has constrained their possible origins.

The orbits of Neptune Trojans are highly stable. Numerical studies suggest that Neptune may have retained up to 50 % of the original post-migration Trojan population over the age of the Solar System.

Neptune Trojans are not fixed to either L_4 or L_5 but are able to librate up to 30° from their associated Lagrangian points over a period of about 10,000 years.

The interesting fact that most of the known Neptune Trojans have highly inclined orbits helps in understanding their origin, as we will see.

The existence of high-inclination Neptune Trojans indicates that they were captured instead of being formed in situ or by collision. Thus, capture or formation of the Trojans at the Lagrangian regions likely occurred during or just after the planet formation epoch, when conditions in the solar system were vastly different from those now.

The captured population already had to be dynamically excited for high-inclination Trojans to exist. Although resonant trans-Neptunian objects are thought to have been captured by sweeping resonances during planet migration, this process would have caused the escape of Neptune Trojans. Irregular planetary migration would result in the depletion of the associated Trojan reservoir.

The estimated equal number of large L_5 and L_4 Trojans indicates a common capture mechanism for both L_4 and L_5 Trojans. The original population of Trojans probably contained many objects on dynamically unstable orbits, and the current Trojan population continues to contribute centaurs.

Spectroscopic analysis revealed that the Neptune Trojans seem to be similar in their chemical composition to Jupiter Trojans, irregular satellites of the outer planets, and possibly cometary nuclei. This suggests a common origin. All these different populations may have been subsequently dispersed, transported, and captured in their current locations during or just after the planetary migration phase.

2.9 The Centaurs: Comets or Asteroids?

We now turn to another interesting group of small solar system bodies that has not received much attention outside the astronomical community. Additionally, this group, called the Centaurs, has been a worry to the involved scientists. What are these objects?

Centaurs are a strange type of small solar system bodies, something that had not been anticipated before by astronomers. Before their discovery, the world of small bodies had been clearly differentiated into asteroids and comets. Centaurs, however, show characteristics of both, comets and asteroids. How should they consequently be classified?

The International Astronomical Union (IAU) thus decided to introduce the new group of "centaurs" in order to tackle this problem. These objects are named after the beings from Greek mythology that were a mixture of horse and humans (see Fig. 2.16), simply to show that they are something in between, some kind of hybrid.

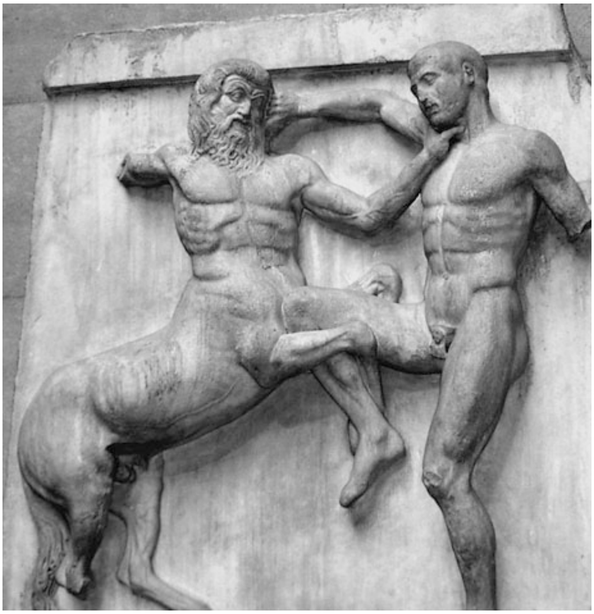


Fig. 2.16 Lapith fighting a centaur. South Metope 31, Parthenon, ca. 447–433 BC (credit: Adam Carr, CC BY-SA 3.0)

2.9.1 What Are Centaurs?

There is still a dispute going on in the astronomical community about a uniform classification scheme. Several proposals exist that differ slightly. In general, a centaur is a small solar system body having its semi-major axis somewhere between the outer planets, i.e. between Jupiter and Neptune. They also have unstable orbits, as they are not protected by any mean-motion resonances with any of the outer planets. On the contrary, the gravitational interactions with these gas giants contribute significantly to their instability.

Currently, three more or less widely accepted definitions exist and are open to debate. The first definition is proposed by the Jet Propulsion Laboratory (JPL) in Pasadena (USA). According to the JPL, a centaur has a semi-major axis between those of Jupiter and Neptune, i.e. between (5.5 and 30.1 AU).

The Minor Planet Center (MPC) is more restrictive and defines a centaur as an object that has its perihelion beyond the orbit of Jupiter and its semi-major axis is inside the orbit of Neptune.

The most complex definition is given by the Deep Ecliptic Survey (DES), a project to find Kuiper belt objects (KBOs), using the facilities of the National Optical Astronomy Observatory (NOAO) at the Kitt Peak National Observatory (KPNO) in Arizona (USA). Its classification is based on simulated changes in behavior of the present orbits when being extended over 10 million years. DES, hence, defines a centaur to be a non-resonant object, i.e. with potentially unstable orbit, whose instantaneous perihelion is less than the osculating semi-major axis of Neptune at any time during the simulation. The osculating semi-major axis, in this case, means the gravitational Kepler orbit at a given time that Neptune would have about the Sun if perturbations were not present.

2.9.2 Physical Characteristics of Centaurs

One aspect centaurs share with most small solar system bodies is, unsurprisingly, that they are very small. In contrast to the asteroids of the main asteroid belt or the NEAs, an additional problem arises with their much larger distance from us. Hence, it is hardly possible to directly observe any surface features on these tiny objects mostly having diameters of only several tens of kilometers.

Therefore, indirect methods are required that may provide indications as to their physical characteristics. Their colors and spectra are good candidates for this purpose.

The colors of centaurs are very diverse but can be classified into two groups: very red, such as (5145) Pholus, and blue like (2060) Chiron. Until now, it is not yet understood why the centaurs exhibit this wide range of colors and particularly these two main groups. Many theories exist. The colors may indicate a different origin, making the centaurs an in themselves inhomogeneous group. It is also proposed that the differences may be due to different levels of space weathering from radiation and/or their temporal cometary activity. Furthermore, some astronomers suggest that multiple competing processes may be the cause, e.g. reddening by radiation and re-surfacing due to collisions.

Spectra can provide a more detailed understanding of the physical characteristics of the Centaurs. Yet, since these objects are so small and faint, their spectra are very difficult to obtain and are often ambiguous. Spectral analysis, nevertheless, has provided some interesting insight into this matter.

Using this method, water ice has been detected on surface of some Centaurs, such as (2060) Chiron, (10199) Chariklo and (5145) Pholus. Further, models based on such analysis suggest that Chariklo's surface consists of a mixture of so-called tholins and amorphous carbon. Tholins are special types of heteropolymer molecules formed by the effects of solar UV radiation of simple organic compounds such as methane or ethane. Tholins have not been detected on Earth so far. Yet, a great abundance of these molecules exists on the surfaces of the icy bodies in the outer solar system. It has been found on Saturn's moon Titan and several trans-Neptunian objects, i.e. objects whose orbits are beyond the orbit of the outermost gas giant. It thus seems likely that tholins also cover the surfaces of Centaurs. In any case, these molecules give some reddish brown color to the objects they cover.

The centaur Pholus exhibits a dark reddish color. This together with its spectra indicates that it surface comprises dark amorphous carbon and a mixture of water ice, methanol ice, olivine grains, and tholins.

Most observed Centaurs so far appear to be of simple or moderate chemical complexity. Chiron, to the contrary, shows a quite complex structure as its spectra vary with the time the observation is made. Water ice, e.g., has not always been detectable. So far, it has been assumed that it was only present at periods of low cometary activity and disappeared during high activity periods. The reasons for that are unknown. Recent studies published in the beginning of 2015 revealed that Chiron is the second centaur after Chariklo having rings. These rings could also be accountable for the variability of detectable water ice if we assume that the water ice is mainly present in the rings. Hence, the amount of water ice would depend on the view we have on the rings: edge-on views, e.g., would hardly show any water ice.

2.9.3 Similarities with Comets

We have talked a lot about the hybrid nature of Centaurs, partially being of cometary nature and partially like an asteroid. However, what exactly makes Centaurs so similar to comets and what are the consequences?

Chiron was detected in 1977 and classified as the first Centaur. By that time it behaved like any other asteroid that had been discovered so far. It was only its orbit, which put it into the group of Centaurs. Astronomers were then highly surprised when Chiron approached its perihelion in 1988 and 1989 and started to show a coma, something that had only been expected with comets. When a comet approaches the Sun its core, the nucleus, heats up and begins to outgas and hence begins to display a visible atmosphere around the nucleus consisting of gas and dust. This atmosphere is called a coma.

This discovery caused some problems. Was Chiron really an asteroid or rather a comet? Chiron showed characteristics of a comet, its size however (about 233 ± 14 km diameter) was much larger than any comet. Typically, cometary nuclei have diameters of a few kilometers. Halley's comet, probably the most famous representative of its group, has a diameter of only about 15 km.

Further, centaurs have been monitored for cometary activity.

At least two of them have confirmed activity: (60558) Echeclus and 166 P/NEAT. You can already see at their designations that their discovery story and initial classification is different. 166 P/NEAT bears the designation of a comet. It had been discovered as a comet but was later found to be on a Centaur orbit and thus was reclassified as a centaur. Echeclus, on the contrary, was detected and classified as an asteroid, hence bearing a minor planet designation (60558). Sometime after its discovery it started to show cometary activity and since its orbit fitted, it was reclassified as a centaur.

A clear distinction is often not possible. For this reason, many centaurs have both a comet and an asteroid designation. Taking Chiron as an example again, it has two designations: (2060) Chiron and 95 P/Chiron.

2.9.4 Orbits and Origin of Centaurs

The orbits of centaurs are not uniform at all. On the contrary, they cover a wide range of eccentricities from highly eccentric, like the centaurs Pholus, Asbolus, to almost circular orbits, such as for Chariklo. Their orbits, however as we have already seen, are independent from their actual eccentricities highly unstable as they are not protected by any resonance with the outer planets but are subject to perturbations caused by gravitational interactions with them.

The centaurs' orbits can give indications as to their origin. Dynamical studies that have been carried out suggest that centaurs merely represent an intermediate state of objects coming from the outer regions of our solar system on their way to potentially join the Jupiter family of short-period comets. However, their origin still remains unclear.

Possible candidates are Kuiper belt objects (KBO) or scattered disk objects (SDO), both of which we have already discussed in more detail earlier in this book. At this stage, a brief description of them shall suffice. KBO and SDO belong to the so-called trans-Neptunian objects, i.e. objects that are located beyond the orbit of Neptune ranging from about 30 AU to approximately 55 AU distance from the Sun. The KBO are gathered in a doughnut-like structure and orbit on more or less stable orbits, which are locked in mean motion resonances with Neptune. The scattered disk overlaps partially with the Kuiper belt but its members have perihelia larger than 30 AU and orbits that are more highly inclined than those of KBOs. Furthermore, SDOs are not protected by any resonances. Hence, their orbits can be perturbed by Neptune. This renders their orbits relatively unstable. Although, KBOs have relatively stable orbits, recent studies have shown that some of them, the plutinos (objects in 2:3 resonance with Neptune and named after their most prominent member Pluto), can potentially be perturbed by close encounters with the dwarf planet Pluto.

SDOs due to their unstable orbits should be the best candidates for centaurs. Their physical characteristics, especially their colors however do not fit with the bicolored distinction do not match. Plutinos and other KBO in contrast show similar bicolored characteristics. Differences could be explained by space weathering.

These potential centaur candidates could be perturbed by Neptune or Pluto or both of them and forced into orbits crossing the one of Neptune. In particular, the gravitational interaction with the ice giant may then reduce their orbits to semi-major axes smaller than the orbit of the planet and thus turn them into the class of centaurs (see the definitions above).

Their orbits remain chaotic and may change rapidly with further encounters with the gas giants. In some cases, especially during encounters of Jupiter, their perihelia can be reduced to within the inner solar system to make them comets of the Jupiter family.

However, even then, their journey continues inexorably towards their "deadly" end. No happy end is foreseen for these small solar system bodies. They will either collide with the Sun, crash into a planet or be expelled into interstellar space.

2.10 Composition of Asteroids

Asteroids have very differing compositions and are thus not easy to classify according to this feature. Their composition is not well understood yet. Only a few detailed samples are known, in particular by exploration through space probes.

If we consider Ceres, though being a dwarf planet, it is supposed to have a rocky core covered by an ice mantle. We will come back to this object in Chap. 5.

Vesta, the second largest object in the main belt, probably has a nickel-iron core, an olivine mantel and a basaltic crust.

Hygiea, discovered in 1849 and the fourth largest asteroid, is of completely undifferentiated primitive composition of carbonaceous chondrite.

Smaller asteroids are thought to be mere rubble piles held together loosely by gravity.

How can we determine the composition of asteroids, in particular, as we are not able to send a space probe to each of them? The classical method in general assumes that we need three characteristics for this: the albedo, the surface spectrum and the asteroid's density. The last characteristic is rather difficult to determine. An accurate determination is usually only possible if the asteroid has at least one moon whose orbit can be observed.

All the asteroids having moons that are known and that have been analyzed more closely turned out to be mere piles of rubble.

2.11 Spectral Classification of Asteroids

With the ever increasing number of discovered asteroids, the wish arose to classify them. In 1975, a taxonomy was introduced by Clark R. Chapman, David Morrison and Ben Zellner which was based on the color of the asteroids, their albedo and their spectral shape. By that time, astronomers believed that these three characteristics would correspond to the composition of the surface material quite well. Today, we know that this is not that easy.

Originally, according to the proposed scheme, asteroids were divided into three distinct types. C-type asteroids are dark carbonaceous objects. About 75 % of all known asteroids belong to this group. The second group is formed by the S-type asteroids, i.e., stony objects comprising about 17 % of all known asteroids. U-type asteroids, the third group, comprise all objects not fitting into the C- or S-group.

This scheme has been further expanded over time due to the increasing number of observations. It has turned out the spectra are by far more differentiated and the original scheme was to coarsely defined.

Nowadays, two classification schemes are widely accepted. We will only briefly discuss them and their major groups, which are based on the original scheme. A more detailed discussion would go far beyond the scope of this book. In the reference list at the end of the book, further information can be found.

The first scheme was proposed by David J. Tholen in 1984. He defined 14 types. Based on Tholen's scheme, Schelte J. Bus and Richard P. Binzel, introduced a modified version in 2002 comprising in total 24 types—the SMASS classification (Small Main-Belt Asteroid Spectroscopic Survey).

Both schemes comprise three broad categories: C, S, and X. C and S are basically based on the original definition. However, further sub-categories were introduced. X-type asteroids are mostly metallic asteroids.

In the following we will only quickly browse through some of the categories. C-type asteroids are, as already mentioned, carbonaceous objects and define the most populated group of asteroids (75 % of the known ones belong to this group). The percentage is even higher in the outer region of the main belt beyond distances of 2.7 AU. The true amount of them, however, is difficult to determine, as they are extremely dark objects with low albedos ranging from 0.03 to 0.10.

Basically only D-type asteroids are darker having an even lower albedo. D-type objects show featureless reddish spectra and can be mainly found in the outer asteroid belt and beyond it. The centaur (944) Hidalgo belongs to this group as well as most of the Jupiter Trojans. Considering the Nice model of the formation of our solar system, these objects most likely originated from the Kuiper belt.

The S-type objects form the second most populated group comprising about 17 % of all known asteroids. Compared to the C-type asteroids they are brighter with albedos from 0.10 to 0.22. They mainly consist of iron- and magnesium-silicates, hence having a stony composition. They are most dominant in the inner asteroid belt within 2.2 AU and are still common in the middle belt with semi-major axes ranging up to 3 AU. Yet, they are almost not represented in the outer belt or regions beyond of it.

Not surprisingly, some of the earliest discovered asteroids belong to this category, such as (15) Eunomia und (3) Juno.

2.12 Prominent Asteroids and Centaurs

As we have learnt a lot about asteroids and the special group of Centaurs, it is a good point in time to show some prominent examples of these small solar system bodies. Of course, the selection cannot be complete and a listing all prominent objects in detail would go well beyond the scope of this book. Nevertheless, the following examples should be interesting. If you wonder why Ceres, the first discovered asteroid, is missing in this selection, it will be dealt with in Chap. 5 as it was officially classified as such by a decision of the IAU in 2006.

2.12.1 Vesta: An Atypical and Very Large Asteroid

Vesta, or to be more precisely (4) Vesta, is one of the largest asteroids in our solar system with a diameter of approximately 525 km. Not only owing its size, as we will see, Vesta is also one of the first asteroids that has been discovered. It is the second most massive object in the main asteroid belt after the dwarf planet Ceres. The latter one, nevertheless, being about 3.5 times more massive and approximately two times larger than Vesta. Both Vesta and Ceres together represent about 30 % of the total mass of the asteroid belt.

Yet, Vesta, though being second with regard to mass, it only takes third place in the category of size. Pallas, discovered, a few years earlier is slightly larger.

Vesta is atypical in some aspects. It is one of the few asteroids with a very differentiated surface and most probably the last remaining "original" planetesimal that has been left over from the formation of the terrestrial planets. Vesta, in addition is most likely "mother" of the Vesta family asteroids, which we have dealt with earlier in this chapter already.

Discovery

Vesta's discovery is closely related to the one of the dwarf planet Ceres, the first object that had been found in the region which we nowadays call the main asteroid belt.

Soon after Ceres' discovery, Heinrich Wilhelm Olbers (1758–1840) found another object in that region, Pallas, in 1802. Thereafter, Olbers was sure that both, Ceres and Pallas, were remnants of a much larger body, potentially, a planet that used to orbit there but had been destroyed by either an internal explosion or a collision. He assumed that if that was the case, further remnants had to exist. Olbers informed his friend and colleague Herschel about his theory and suggested to him to start a search for other fragments near the locations where Ceres' and Pallas' orbits intersected. Olbers himself immediately started his quest for further fragments and succeeded to find another large object close to his proposed positions on 29.03.1807. The object he found should thereafter be named Vesta after the Roman goddess of home and hearth, following a proposal of the German mathematician Carl-Friedrich Gauss.

Today we know, that Olbers simply had been lucky and his discovery had been pure coincidence, as the hypothesized planet had never existed. Hence, Ceres and Pallas were not related fragments and the intersection of their orbits more or less arbitrary.

In between the discovery of Pallas and Vesta, the asteroid Juno had been found, making Vesta the fourth identified asteroid. Therefore, its official designation is (4) Vesta. Vesta was the last asteroid to be discovered for a very long period of about 38 years.

Physical Characteristics

Vesta had been discovered but nothing except its orbital elements was known. We have seen it already a few times that even today it can be quite difficult to obtain further information about a small solar system body, e.g., its physical characteristics. Now, imagine how the situation had been at the beginning of the nineteenth century when technology and detailed knowledge about celestial mechanics were by far more limited than nowadays.

By that time, observations were limited to the visible light. Besides visual observations, photometry was a common. Spectral analysis still was in its infancy and was very difficult to apply. Telescopes of very good optical quality were still lacking. Photometry is concerned with measuring the flux or intensity of an astronomical object's (e.g., an asteroid) electromagnetic radiation. By this the light curve of the asteroid can be determined. Since an asteroid is not spherically shaped but rather quite irregular, the amount of sunlight being reflected from its surface depends on the amount of surface we look at and its albedo. Just assume a peanut-shaped asteroid. Sometimes we may look at only one "head", sometimes at the full "body". It is quite clear, if we assume equal albedo that in the latter case much more sunlight will be reflected. Now, let us draw a diagram in which we put the amount of reflected light as a function of time. Then, while our peanut asteroid is rotating, we can see some ups and downs in the diagram. When we only see the "head", there will be a down and in the case of the full body an up. The diagram we just obtained is called a light curve. The pattern will repeat with each rotation of the asteroid.

As you may see, it is hence possible to derive the rotational period of an asteroid from the light curve. We basically only need to determine the time between two occurrences of the same peak in the curve. Though this may sound quite straightforward, the truth is that the devil is in the details. Often the surface of a body does not show the same albedo. Differences may exist such as darker regions, which reflect light much worse than brighter parts. This makes the determination more complicated.

The Problems of Size. . .

For similar reasons the results determined by photometric analysis of Vesta soon after its discovery were lengthily debated within the scientific community. Today, we know, also confirmed by close observations by the spacecraft Dawn, that Vesta's rotational period is at 5.342 h, which is considerably fast for asteroids.

Similar problem that had arisen with its rotational period came up while trying to determine the asteroid's diameter. This is in particular difficult when angular resolution is not sufficient to resolve the "planetary" disc of the object. Just imagine the case of a planet, e.g. Jupiter. When we observe it through a telescope, we can easily resolve its planetary disc revealing the storms etc n it. We can then measure the size of the disc. If we now know the distance to Jupiter, it is a matter of straightforward trigonometry to determine its actual diameter. However, as you can see, we need the object's disc. What can we do if we cannot resolve a disc at all, e.g. as is the case with asteroids?

In this case, other, indirect measures have to be applied. This could be photometric analysis that, as we have just seen, can be used to determine an object's light curve and subsequently derive its rotational period. An at first glance easy method is to measure the asteroid's albedo, i.e., the amount of light reflected from its surface. If we assume an evenly distributed albedo over the asteroid's whole surface we can deduce its mean diameter, as we know that an object of a specific size should reflect a certain amount of light when having a specific albedo. This approach is, however, much depending on the assumed albedo. The diameter can, therefore, vary considerably.

This was also the case with Vesta during its early observations. In 1825, e.g., its mean diameter was determined to be roughly 383 km. Then, in 1879, the American astronomer and physicist, Edward Charles Pickering (1846–1919) found Vesta's diameter to be at 513 ± 17 km which is astonishingly close to its current value. However, further measurements in the following decades lead to more wide-spread results again ranging from about 390 up to 602 km.

If the albedo approach can lead to such wide-spread results, do other, more reliable methods exist? Indeed, they do. One is also related to photometry and considers special situations when the asteroid passes in front of another object, preferably a star. We speak of star occultations. During such occasions, the light intensity or brightness of the star, we can observe, will drop when the asteroid passes in front of it. We can then measure the time of the drop and use this to compute the asteroid's diameter.

Such an occasion turned up in 1991, when Vesta occulted a star. From observations of this event, astronomers could derive an elliptical profile of Vesta with dimensions of 550×462 km which was later on also confirmed by measurements of modern spacecraft, such as Dawn.

Another possibility is to use speckle interferometry. In 1989, Vesta's diameter was determined to be 498–548 km varying during its rotational period.

. . . and Mass

Vesta was also the first asteroid for which its mass was determined. Again, this is a difficult task and usually requires the existence of another orbiting object, such as a moon. Once we know how far away the asteroid is, we can use the orbital periods of the moon or other orbiting body and the distance of this object to the asteroid to measure the asteroid's mass. For this, we measure the angular separation between the moon and the asteroid and use basic trigonometry to convert the angular separation into distance between the asteroid and moon. That conversion, though, first requires that the distance to the asteroid and moon be known. Unfortunately, no moons of Vesta exist. However, a trick helped. Every 18 years the asteroid (197) Arete approaches within 0.04 AU of Vesta. Such a close approach causes gravitational perturbations. Hans G. Hertz estimated the mass of Vesta as $(1.20 \pm 0.08) \times 10^{-10}$ solar masses. More refined estimates followed, and in 2001 the perturbations of 17 Thetis were used to estimate the mass of Vesta as $(1.31 \pm 0.02) \times 10^{-10}$ solar masses. Spacecraft Dawn measured a mass of 2.6×10^{20} kg.

Further Characteristics of Vesta

In general, Vesta is the second-most massive body in the asteroid belt, as we have seen. However, it only has about one-third of the mass of the dwarf planet Ceres, which is the clearly most massive object in the main belt.

Vesta's density is lower than that of the four inner, terrestrial planets but significantly higher than the density of most asteroids. Scientists see this aspect as one part forming its exceptional part within the asteroid population.

It orbits the Sun with a semi-major axis of 2.36179 AU and an orbital period of the 3.63 years. Vesta thus lies interior to the Kirkwood gap located at 2.5 AU and hence in the inner main asteroid belt.

Depending on its orientation, temperatures range from −20 to −190 °C.

Differentiated Surface and Interior

Other aspects that makes Vesta quite unique among other asteroids are its extremely differentiated surface and in particular its interior. We have seen before in this chapter, that most asteroids are simple rubble piles of rock. Vesta, though, is different. Its interior more resembles that of a terrestrial planet. Vesta has a basaltic crust of cooled down lava. This crust covers a rocky mantle and an iron-nickel core (see Fig. 2.17). Some astronomers, therefore, characterize Vesta as one of the last remaining protoplanets in our solar system.

Also Vesta's surface is more differentiated than is the case with other asteroids. It is scarred by impact craters. In particular, the area around the southern pole is dominated by two tremendous impact craters called Rheasilvia with an approximate diameter of 500 km and Veneneia with roughly 400 km in size (see Fig. 2.18). These two craters are partially overlapping whereas Rheasilvia is overlaying Veneneia. From this, we can conclude that the former one is younger than the latter one.

The impacts that resulted in these two craters must have been enormously. Simulations suggest that the impacts excavated about 1 % of Vesta's volume. The debris is the thought to have formed the Vesta family of asteroids, which we encountered previously.

If this is the case and the Vesta family asteroids formed like this, then we can indirectly determine an age range for the craters. The fact that 10 km sized

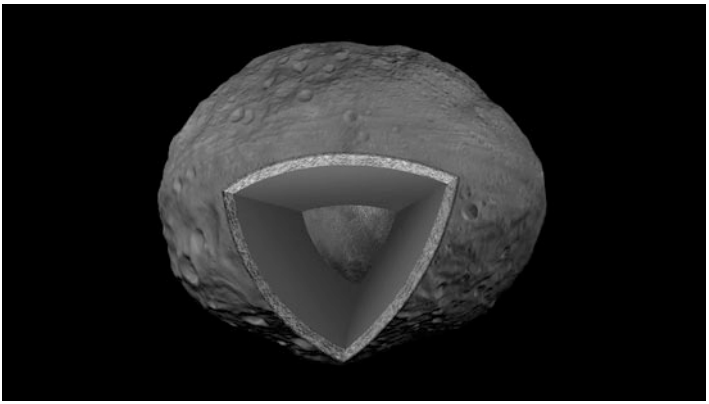


Fig. 2.17 Visualization of Vesta's internal structure (credit: NASA/JPL-Caltech)

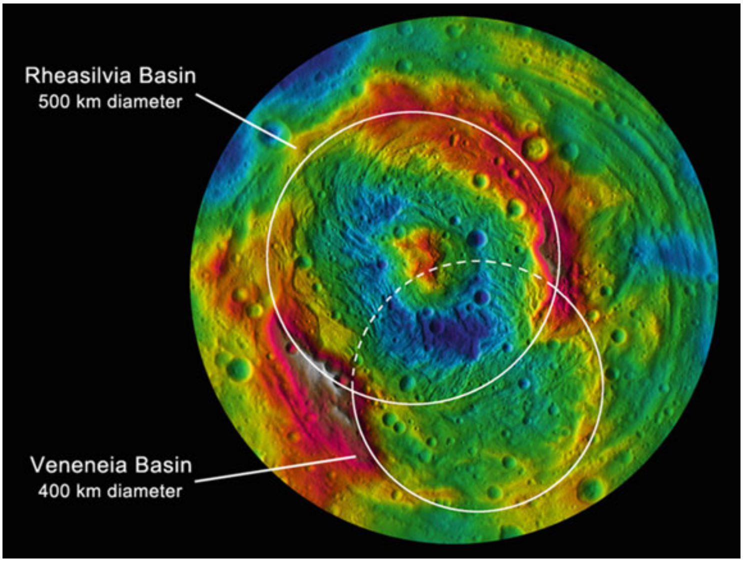


Fig. 2.18 This topographic map from NASA's Dawn mission shows the two large impact basins in the southern hemisphere of the giant asteroid Vesta. The map is color-coded by elevation, with *red* showing the higher areas and *blue* showing the lower areas. Rheasilvia, the largest impact basin on Vesta, is 500 km in diameter. The other basin, Veneneia, is 400 km across and lies partially beneath Rheasilvia (credit: NASA/JPL-Caltech/UCLA/MPS/DLR/IDA/PSI)

fragments have survived collisions etc until the present day indicates that the craters are at most only about 1 billion years old.

Many other craters exist on Vesta. In fact, its whole surface is covered by craters as you can see from Fig. 2.19 which was taken by the Dawn spacecraft. Most of these craters, however, are older and have sometimes significantly degraded, such as e.g. Feralia Planitia which is 270 km across.

A famous and also nice looking group of three craters is the so-called "snowman craters" located in Vesta's northern hemisphere (see Fig. 2.20). If you look at the picture, you can probably easily guess where the name for this group comes from. Their official names from largest to smallest (west to east) are Marcia, Calpurnia and Minucia. Marcia is the youngest and cross-cuts Calpurnia. Minucia is the oldest.

Besides craters, Vesta's surface is also coined by a series of concentric troughs. Most of them are located in the asteroid's equatorial region. Two famous and large ones are the Divalia Fossa (10–20 km wide, 465 km long) and Saturnalia Fossa which is located more to the North (see Fig. 2.21). The latter one is about 40 km wide and more than 370 km long. These troughs are thought to be grabens resulting from the impacts that formed Rheasilvia and Veneneia.

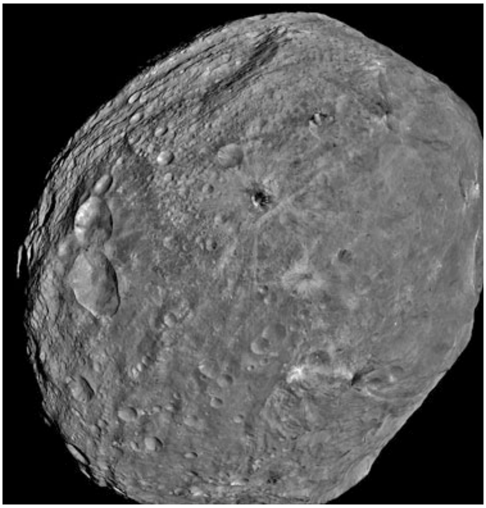


Fig. 2.19 Vesta as seen by Dawn on July 24, 2011 taken from a distance of about 5200 km (credit: NASA/JPL-Caltech/UCLA/MPS/DLR/IDA)

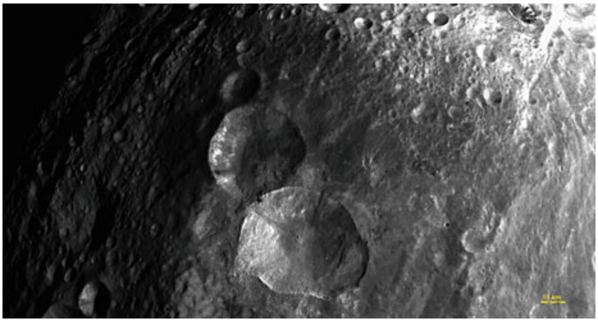


Fig. 2.20 The Snowman craters on Vesta, set of three craters, are located in the northern hemisphere of Vesta. The image was taken on July 24, 2011, from a distance of about 5200 km by the Dawn spacecraft (credit: NASA/JPL-Caltech/UCLA/MPS/DLR/IDA)

2.12.2 (2060) Chiron: The First Centaur

Chiron was discovered by the American astronomer Charles T. Kowal (1940–2011) at the Mount Palomar Observatory on October 18, 1977. He discovered it on images taken 2 weeks before. It is the first object that was classified as a Centaur. The Centaur (944) Hidalgo had been discovered much earlier in 1920 but had not been properly identified as such. Thus, Chiron officially remains the first.

By the time of its discovery, Chiron was near its aphelion (currently 18.891 AU). At this point of its orbit it was too far away from the Sun to show any cometary activity. It was therefore designated a minor planet number and supposed to be an asteroid being on a Centaur orbit. Later on, it was found on several precovery images going back to 1895. Chiron had reached its last perihelion in 1945. However, by then, only a few searches had been made and the focuses of these were not on Centaurs. Chiron hence remained undiscovered.

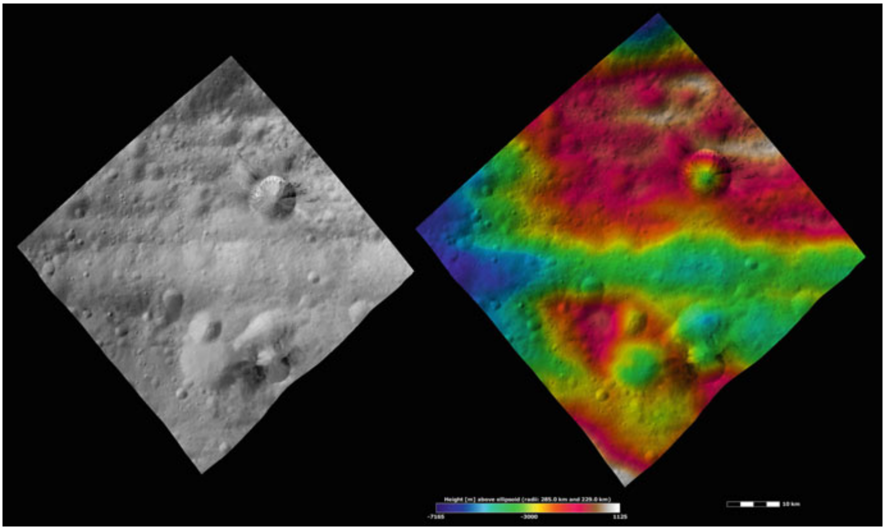


Fig. 2.21 These images show a part of the large trough, Divalia Fossa, which encircles most of Vesta's equator. Divalia Fossa is visible in both the apparent brightness image (*left*) and the topography image (*right*): it is the approximately 10 km wide depression that runs from the *left corner* to the *right corner* of the images (credit: NASA/JPL-Caltech/UCLA/MPS/DLR/IDA)

Chiron's orbit is quite eccentric ($e = 0.37911$) and takes it as close as 8.5114 AU to the Sun at its perihelion, i.e. just inside the orbit of Saturn, and as far away as 18.891 AU at its aphelion meaning just outside Uranus' perihelion (see Fig. 2.22). Its semi-major axis is at 13.708 AU.

Numerical simulations show that the Centaur had its closest approach to the Sun in "recent" history in May 720. During this event, Saturn perturbed its orbit by decreasing Chiron's semi-major axis from 14.4 AU to its current value. This shows again, how vulnerable the orbits of Centaurs are. In Chiron's case, Saturn is the driving force for orbital perturbations, as it never comes as close to Uranus as to the ring planet. When Chiron crosses Uranus' orbit, the gas giant is farther than average from the Sun (see Fig. 2.22).

Further numerical studies suggest that Chiron may leave its current state as Centaur and become a Jupiter family comet in about 1 million years.

Chiron's spectrum has been measure in visible light and near-Infrared. The spectra are neutral and very similar to C-type asteroids and the spectrum of the nucleus of Halley's comet. Campins et al. were able to measure the light curve of Chiron. The light curve is the graph of light intensity of the centaur as a function of time. This curve was then used to estimate the rotation period of the centaur which was then determined to be 5.917 h. Measurements by Spitzer Space Telescope made in 2007 suggest that Chiron has a diameter of about 233 ± 14 km.

In February 1988 a surprising discovery was made. Chiron, by that time, was about 12 AU from the Sun and suddenly brightened by 75 %. This is a phenomenon not uncommon to comets when they approach the Sun and start to develop a coma.

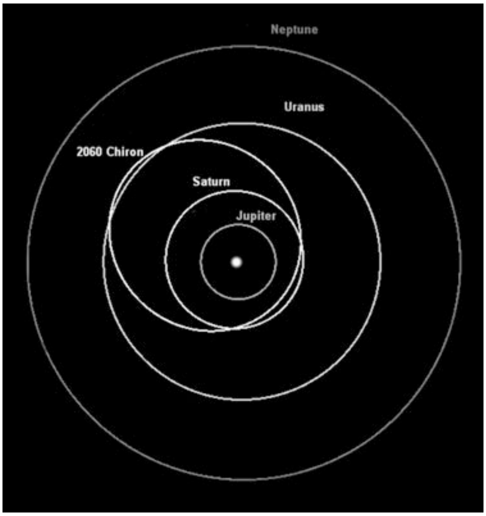


Fig. 2.22 Orbit of the centaur Chiron (credit: based on original work by Reyk; Public Domain)

Yet, Chiron was still considered to be an asteroid. Such an event had never been detected before with an asteroid. What had happened? Had the asteroid Chiron developed cometary activity at such a huge distance from the Sun where most comets remain inactive?

Further evidence for cometary activity was detected by Meech and Belton in April 1989 in form of a coma. Campins et al were then in 1993 able to see a tail originating from the centaur, clear evidence for a comet. One thing that clear right from the beginning was that it was some kind of atypical cometary behavior. At such distances from the Sun, the critical temperature for triggering the sublimation of water ice cannot be reached as the incident solar energy is simply too low. With comets, the sublimation of water ice, however, is in general the driving force. Hence, other processes must have been responsible in the case of Chiron and other centaurs, most likely ices that require a lower sublimation temperature.

Other unusual aspects were found. Chiron's brightness varied and also the amount of water ice that could be detected varied. In times of high cometary activity it was almost not measurable whereas during low activity phases it was. The centaur's cometary behavior has been solely held accountable for these aspects.

In early 2015, new insights were gained that may add to this. In January 2015, it was announced that two rings had been found around Chiron. This would make it the second centaur after Chariklo that has a ring system.

The rings were detected during stellar occultation during which the centaur passes in front of the line of sight of a star and thus blocks the light coming from the star. Such events occurred on 7.11.1993, 9.3.1994, and 19.11.2011. The star's light emissions can then be analyzed. Assuming Chiron had no rings, a simple body without any surrounding material, the centaur would block the star's light in its entirety when passing in front of it. There would be a sudden drop in the light emissions of the star when Chiron was in front of it. Yet, this was not what had happened. Two sharp features were detected at the beginning and the end of the occultation blocking only a fraction of the starlight. These features had been misinterpreted as jets coming from the surface of the Centaur due to its cometary activity.

New analysis, however, revealed that these two features more likely represent rings around Chiron at a distance of about 324 ± 10 km from its center. Each of the rings is supposed to be 3–7 km wide.

The rings could contribute to the explanation of the problem of varying water ice detectability and Chiron's varying brightness. If we assume the water ice to be located in the rings, these values depend on the viewing angles we have on the rings. In 2001, when the indications of water ice disappeared from Chiron's spectra we looked at the rings edge-on.

Where did the rings come from? The origin is unknown. It may be material from the surface of the Centaur which had been frozen there while it had been in the Kuiper belt. When coming closer to the Sun during its transition from a KBO to a Centaur, a sublimation process might have begun and caused the ejection of gas and dust from the surface. This debris then may have formed the rings.

2.12.3 *Chariklo: The One Who Dances with the Rings*

The small solar system body (10199) Chariklo is currently the largest known Centaur and moves on a moderately eccentric orbit ($e = 0.171$) between Saturn and Uranus. Its perihelion is at 13.05 AU and its aphelion at 18.48 AU. Its orbit is highly inclined with 23.41° towards the ecliptic plane. Chariklo's orbital period is 62.5 years.

With a diameter of approximately 250 km it is a bit larger than Chiron (about 230 km). Chariklo has a long orbital half-life as it lies very close to the 4:3 mean motion resonance with Uranus (only 0.09 AU or 13.5 million kilometers from that resonance).

On June 3, 2013, a stellar occultation occurred during which Chariklo passed in front of a 12.4 magnitude star. The analysis of this occultation revealed something astonishing: the Centaur has two rings (see Fig. 2.23). This was a surprising and completely unexpected discovery because it had been thought that rings could only be stable around much more massive bodies, such as planets.

Further analysis of the data showed that one of the rings, the inner one, is about 7 km wide while the second one is less distinct with a width of only 3 km. The two rings, the inner one being named Oiapoque and the outer one Chuí, are about 9 km apart and have respective distances of 396 and 405 km from Chariklo.

At the time of discovery of the rings, Chariklo was the smallest object in the solar system with which rings were found. This changed in 2015, when also two rings were found around the slightly smaller Centaur Chiron.

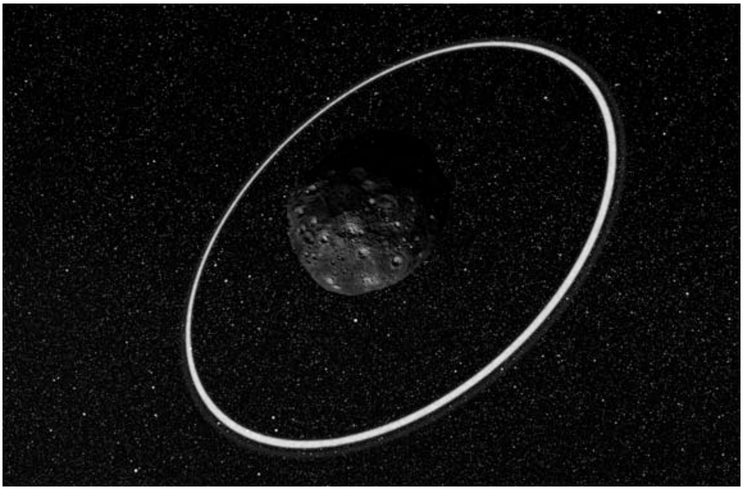


Fig. 2.23 Artist's impression shows a close-up of what the rings of Chariklo might look like (credit: ESO/L. Calçada/M. Kornmesser/Nick Risinger)

Both the rings of Chariklo and Chiron are not stable features but will disperse over a period of a few million years. Chariklo's mass is not high enough to hold the rings forever.

Where did the rings come from? The discovery was only recently and a origin has not yet been found. Many models were proposed. The most promising seems to be that they are probably remnants of a debris disc, which formed after a heavy impact some time ago.

2.12.4 Ida and Dactyl: The First Asteroid with a Moon

The asteroid (243) Ida was discovered on September 29, 1884 by the Austrian astronomer Johann Palisa (1848–1925). It is a S-type asteroid and is most famous for being the first asteroid for which a moon, Dactyl, was detected. On August 28, 1993, Ida was visited by the spacecraft Galileo while it was on its way to Jupiter. Ida was the second asteroid to be visited by a spacecraft and the visit of Galileo provided new and detailed insight into the geology of S-type asteroids. So, what do we know about Ida and Dactyl?

Physical Characteristics Orbit and Family

Most information about Ida and its moon Dactyl were obtained during the flyby of Galileo in 1993. Ida is a relatively small asteroid with dimensions of $59.8 \times 25.4 \times 18.6$ km and a mass of $3.65–4.99 \times 10^{16}$ kg. It has a very elongated shape with a very irregular surface (see Fig. 2.24). One thing that becomes apparent is Ida's strange shape. It seems to consist of two distinct larger bodies. Let us call

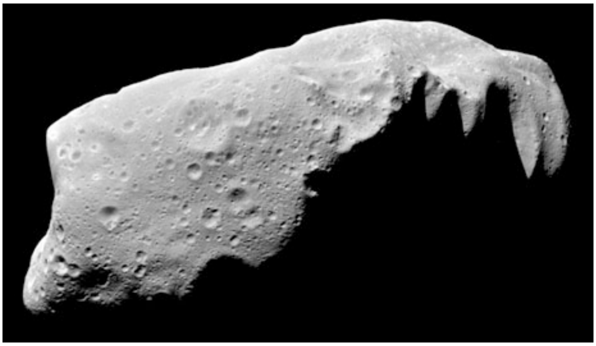


Fig. 2.24 Mosaic of the asteroid (243) Ida of five image frames acquired by the Galileo spacecraft at ranges of 3057–3821 km on August 28, 1993, about 3.5 min before the spacecraft made its closest approach to the asteroid (credit: NASA, Jet Propulsion Laboratory)

these bodies A and B. Scientists assumed that the gap in between these two bodies would be filled with loose debris. However, the high-resolution images taken by Galileo during its flyby did not reveal such debris.

The shape is also responsible for a very uneven gravitational field exhibited by the asteroid. It is lowest at the extremities due to the high rotational speed and is also low at the connection between the two bodies, the "waist". The latter is quite obvious. Less material is concentrated there and therefore a lower mass can be found at the waist. The mass is concentrated in the two halves leaving only bit for the waist.

Ida is a main belt asteroid having a semi-major axis of 2.862 AU, its perihelion at 2.732 AU and its aphelion at 2.991 AU. It is thus member of the middle asteroid belt. It takes Ida about 4.84 years to orbit our Sun once. With a rotation period of 4.63 h it is one of the fastest rotating asteroids in our solar system. The dwarf planet Haumea, located in the outer solar system, has a comparable fast rotation of only 3.9 h.

Ida moves along a near-circular orbit ($e = 0.0452$), which lies almost completely within the ecliptic plane due to its low inclination of 1.138°.

Ida is a member of the Koronis family which is thought to have been formed at least 2 billion years ago in a catastrophic collision between two larger bodies. The Koronis family travels in a cluster along the same orbit. Over 300 have been found but only about 20 are larger than 20 km in diameter. Ida is one of the largest members.

A Scarred Surface

The images taken by Galileo during its flyby showed that Ida is one of the most densely cratered bodies in our solar system. There are so many craters on its surface that scientists speak of a reached saturation level. What does this mean? There are so many craters on the surface in such small distances from each other that any new impact will inevitably destroy or damage an existing crater. There is simply not enough free space left.

The craters vary in size. Practically all possible sizes can be found ranging from very small craters to very extended ones. The visible craters also show different levels of degradation and are of different ages. There are young ones but also some that may be as old as Ida itself.

Besides the visually dominant craters there are also various other surface structures apparent such as grooves, ridges and protrusions.

It is interesting to see that the two bodies A and B differ in their appearance. While B gathers almost all of the larger craters, A lacks any larger crater. The reasons for this are unknown.

Body A is dominated by a large ridge of 40 km named Townsend Dorsum and that stretches 150° around Ida's surface. The other characterizing structure is a large indentation named Vienna Regio.

Body B combines several sets of grooves on its surface. Most of them are about 100 m wide and up to 4 km long.

Composition of Ida

Ida's surface is covered by a thick layer of regolith, i.e. a layer of very small debris. This layer is about 50–100 m thick. It is thought that the regolith on Ida developed during impact events, which must have been very frequent when looking at its surface. Ida's regolith comprises olivine and pyroxene, a quite common composition of regolith in the solar system.

Although the whole surface of Ida is covered with this material and in general appears to be quite homogenous, some smaller color differences have been detected. How does this come?

Regolith does not stay unchanged for eternity. It is subject to space weathering, i.e., it is exposed to electromagnetic radiation coming from the Sun. In particular ultraviolet radiation can be a driving force in the space weathering process during which regolith over time slightly becomes redder. This means, however, the redder the material the older it is. The regolith on Ida's surface is not of the same age. It was created during impact events distributed over the whole lifetime of the asteroid. Those parts with old regolith hence appear redder than the younger parts.

We have already pointed out that Ida is an S-type asteroid. Its internal structure has not been analyzed so far directly. It is, however, assumed that Ida's interior comprises at least some impact fractured rock, so-called megaregolith, which extends from a few hundred meters below the surface to several kilometers deep.

Measurements by the Galileo spacecraft analyzed Ida's spin, which indicated that the asteroid has a consistent density of 2.27–3.10 g/cm^3 and an assumed porosity of 11–42 %.

Dactyl: A Companion for Ida

During the flyby of Galileo many images were taken. On February 17, 1994 Galileo mission member Ann Harch discovered Dactyl while examining delayed image downloads from the spacecraft (see Fig. 2.25). Dactyl, or (243) Ida I Dactyl as is his official designation, is the first moon that has been found orbiting an asteroid. During the discovery, Ida and Dactyl were only separated by about 90 km. Galileo itself was about 10,760 km away from Ida and 10,870 km from Dactyl.

Ida's moon is a heavily cratered, egg-shaped body of $1.6 \times 1.4 \times 1.2$ km. It shows more than a dozen craters with diameters larger than 80 m (see Fig. 2.26).

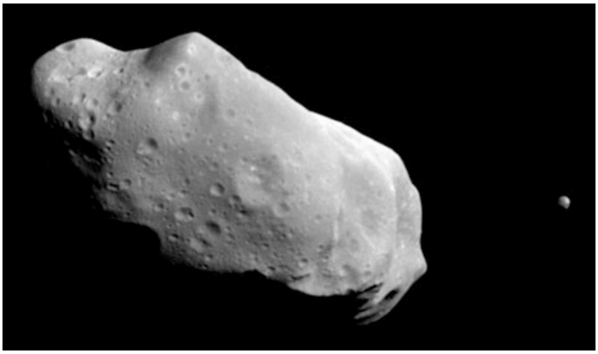


Fig. 2.25 Asteroid Ida with its moon Dactyl as seen by the Galileo spacecraft on August 28, 1993 (credit: NASA/JPL)

Fig. 2.26 Most detailed image of Ida's moon Dactyl as seen by the Galileo spacecraft on August 28, 1993 (credit: NASA/JPL)

Hence, it must also have suffered substantial bombardment during its existence similar to Ida.

Dactyl and Ida share many characteristics be it similar albedo or spectra. It only appears that space weathering is hardly active on Dactyl. Furthermore, the moon is simply too small for the formation of regolith. Therefore, while a thick layer of regolith covers Ida, Dactyl exhibits naked rock.

Origin of the Ida-Dactyl System

Astronomers believe that Ida originated from a breakup of the parent body of the Koronis family. Taking into account the known and observed Koronis members, this parent body should have had a diameter of approximately 120 km. A heavy impact with another large body then triggered the catastrophe and thereby the formation of the Koronis family. Yet, it is not known when this event actually happened.

Dactyl may have originated at the same time as Ida from the disruption of the aforementioned Koronis parent body. However, it may have formed more recently. A large impact on Ida might have thrown material from Ida into space, which was then captured by Ida again and then formed Dactyl. It is extremely unlikely that Ida captured it due to its low mass and size.

Chapter 3
Comets

Probably the most prominent representatives of small solar system bodies (SSSB) are "comets". Their sometimes impressive appearances have often influenced people's thinking, both in cultural and historical aspects, as we will see in the following.

3.1 Early Observations and Considerations

People in the early times were often shocked or at least fascinated when comets suddenly appeared in the sky. One should also keep in mind for a moment, how impressive great, bright comets must have been in a time when the night sky was really dark in the absence of any (modern) light pollution. A view we can hardly enjoy nowadays.

These odd objects simply behaved and looked so differently from what people were used to when looking up into the sky. They suddenly seemed to appear from nowhere and had a strange tail attached looking similar to a cloud of smoke. Was something burning up there? Then as suddenly as it appeared, the phantom vanished. It took a very long time before people really began to understand what the characteristics were behind these appearances. Only then, they were able to derive some regularity associated with the occurrences.

Until then, as was the case with similar celestial phenomena, these fuzzy objects were considered to be signs or omens of the gods often foretelling the deaths of kings or coming catastrophes.

First written evidence of observations of comets traces back to the old advanced civilizations of Mesopotamia. Also, the early Chinese were quite active in their observation and interpretation of the omens. The first written record of a comet dates back to 1095 BC. Though not known by then that it was the same and nowadays famous Halley's Comet, its appearance was recorded several times by scholars at the Chinese imperial court, e.g. in 467 BC, 240 BC, 164 BC and 87 BC.

M. Moltenbrey, *Dawn of Small Worlds*, Astronomers' Universe,
DOI 10.1007/978-3-319-23003-0_3

However, not only in China, Halley's Comet played an important role. There were several occasions in Europe and the Middle East where its appearances were interpreted as important omens, of course only in retrospect. Its occurrence in 12 BC is often thought of being an explanation for the Star of Bethlehem related to the birth of Jesus Christ. Then, in 66 AD, for example, when the comet was again visible, observers in the Middle East retroactively interpreted it as being a bad omen that foretold the destruction of the temple of Jerusalem in 70 AD. The Romans on the other hand interpreted this event to forecast the eruption of the volcano Vesuvius and the resulting extinction of the Roman cities of Pompeii and Herculaneum in 79 AD.

Yet not only in ancient times, comets were considered to be harbingers but this notion also continued in the Middle Ages and is sometimes still valid today.

In 1066, Halley's Comet was again seen up in the sky and thought to be an omen again. It was the year of the famous battle of Hastings in which the Norman-French army of William II of Normandy, also thereafter called William the conqueror, inflicted a devastating defeat to the English army under the Anglo-Saxon King Harold II in which the latter one died. The apparition of the comet and its derived significance was eternalized in the famous Bayeux Tapestry as a fiery star (see Fig. 3.1).

Fig. 3.1 The image shows scene 32 of the Bayeux Tapestry created in the 1070s. Men are staring at Halley's Comet (credit: Myrabella, CC0 1.0)

Then, in 1635, the outbreak of the pest in the course of which more than 90,000 people died in London alone, was seen as a punishment of God as in the beginning of the same year a bright comet was visible in the sky.

In 1910, Earth passed Halley's tail. Some newspapers falsely reported that the tail would contain high concentrations of cyanogen, which could poison millions of people. This media coverage created a hysteria resulting in a tremendous increase of sold gas masks.

In the more recent past, the appearance of comet Hale-Bopp, a very bright and visible comet, triggered a mass suicide of 38 followers of the Heaven's Gate cult in March 1997.

Also popular culture is inspired by comets. You only have to look at the vast number of Hollywood movies dealing with this topic. Furthermore, when the discovery of a "new" bright comet is announced, such as ISON in 2014, mass media report about that though they are silent on most other astronomical topics and number of people heading for local observatories drastically increases.

3.2 First Scientific Considerations

Besides all the religious and astrological interpretations of cometary appearances, scholars in all times have tried to find reasoned, scientifically flavored explanations for the nature of these objects. Aristotle and Ptolemy were some of the first providing a "scientifically" based theory beyond seeing almighty gods to be the causes. In his book "Meteorology", Aristotle laid the fundament for what should be the commonly accepted "nature" of comets for almost 2000 years. He thought comets to be a phenomenon of the upper atmosphere where hot, dry exhalations gathered and from time to time burst into flames creating the smoke tails of comets.

This theory sounds strange with our modern knowledge. Yet, it was not until the sixteenth century that comets were considered to be true celestial objects and being not part of our Earth's atmosphere at all. It was the time when the first truly scientific explanations emerged. The Danish astronomer Tycho Brahe (1546–1601) measured the parallax of the great comet of 1577 and was thereby able to show that comets were indeed celestial objects wandering in the sky far beyond Earth's atmosphere.

Though, Brahe's results seemed to be clear, it still took several decades before this finding was widely accepted. Galileo Galilei (1564–1642), one of the great men of astronomy, denied it. He stuck to the Aristotle's notion of comets moving on straight lines through the upper atmosphere. Even Johannes Kepler (1571–1630), who had determined the elliptical shape of planetary orbits in 1609, believed that comets travelled among the planets along straight lines.

In 1680, German astronomer Gottfried Kirch (1639–1710) discovered a very bright comet. Astronomers from all over Europe tracked its position for several months and were thereby able to determine that it moved in a parabola around the Sun. This was in line with Isaac Newton's theory written down in his "Principia

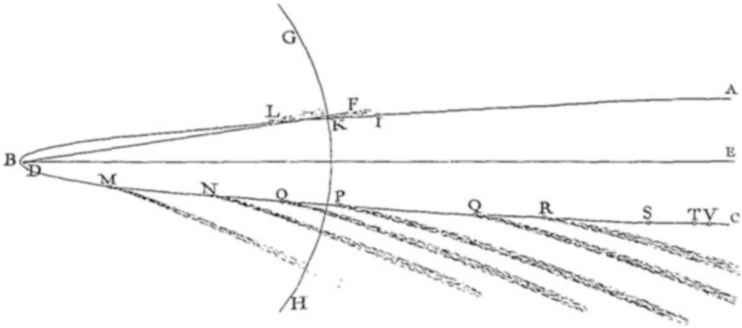


Fig. 3.2 The parabolic orbit of the comet of 1680, fitted to a parabola, as shown in Isaac Newton's Principia Mathematica

Mathematica" published in 1687. Newton proved that an object moving under influence of his inverse square law of universal gravitation must trace out an orbit shaped like one of the conic sections. He used the parabolic orbit of the 1680 comet as an example to demonstrate his theory (see Fig. 3.2).

It only prevailed after 1705 when Edmond Halley (1656–1742) could show that the comet that was visible in 1682 was a periodically returning celestial object. He managed to determine that it orbited on an elongated ellipse around the Sun in 76 years. Looking back into the archives, Halley found out that the comets of 1682, 1607, 1531 and 1456 had very similar orbital elements. He, thus, concluded that these apparitions had been the appearances of the same comet. He then predicted that the comet should re-appear in the sky in 1758–1759. Indeed, the comet returned in 1759. Henceforth, in order to honor Halley, it was decided that the comet should bear his name: Halley's comet. Its official designation, nowadays, is 1P/Halley. We will learn more about the naming conventions later on in this chapter after we discussed the orbital characteristics of comets. Halley's comet has become one of the most famous and well-known comets. We will go into more details in the later part of this chapter.

The second comet found to have a periodic orbit was Encke's comet (2P/Encke). German mathematician and physicist Johann Franz Encke (1791–1865) derived the orbits for a series of comets that had been observed in 1786, 1795, 1805, and 1818. These orbits were very similar to each other, so he concluded that it should be the same comet. He then successfully predicted its return in 1822.

Although we now know that comets are celestial bodies traveling on elliptic or parabolic orbits around the Sun, there still remain many open questions: what are comets? Where do they come from? How are they explored? We will deal with that in the remaining part of this chapter.

3.3 Physical Characteristics of Comets

Besides their orbits, comets' physical characteristics are the cause for their unique appearance. We will thus first discuss these characteristics before dealing with their spatial origins.

A comet is an icy and dirty small Solar System body sometimes referred to as "dirty snowballs". This terminology was created by astronomer Fred Whipple (1906–2004) and is still commonly used today though recent research suggests that this "simplification" might not be correct. Some scientists rather refer to "snowy dirtballs" to characterize a comet.

In any case, when a comet on its orbit approaches the Sun, it heats up and begins to outgas. The solar wind blows the gases and dust ejected from the nucleus away which then displays a visible atmosphere or coma and at least one tail.

Thus, the three main components of a comet, which we will see, are its nucleus, coma and tail.

3.3.1 Nucleus of a Comet

The nucleus or core of a comet is the actual comet itself. It is an amalgamation of rocky material, dust, water ice and frozen gases such as carbon dioxide (CO_2), carbon monoxide (CO), methane (CH_4) and ammonia (NH_3). It forms the central part of the phenomenon we usually call a "comet". All the other visible components, which we usually consider as a comet, are results of processes initially taking place on the surface of the nucleus.

For a long period of time, it was impossible to directly observe a comet's nucleus although it was long falsely believed that one did observe it. What we can usually see is a thin, bright cloud of dust particles surrounding and hiding the real nucleus from our views. This cloud should not be confused with the coma which enwraps the core. It rather appears as a bright condensation within the coma. Only with the help of space probes, such as Giotto and Vega 2, it became possible to directly observe a comet's nucleus (in both cases that of Halley's comet in 1986).

Today, we know that cometary nuclei usually have an elongated and irregular shape. Its gravity is not sufficient to achieve a so-called hydrostatic equilibrium which would force the nuclei into a spherical shape (as is e.g. the case with planets and most of their larger moons). Some of them, such as 19P/Borrelly and 103P/Hartely, look like a peanut others better resemble a potato. Recent observations of comet 67P/Churyumov-Gerasimenko by the Rosetta spacecraft show a "duck-like" structure (see Fig. 3.3).

Before direct observations of a nucleus were possible by visiting spacecraft, many scientists assumed that a cometary nucleus had to be quite large in size. How else could their brightness be explained? However, this assumption turned out to be wrong. On the contrary, a cometary nucleus is small object having usually only a few kilometers in diameter.

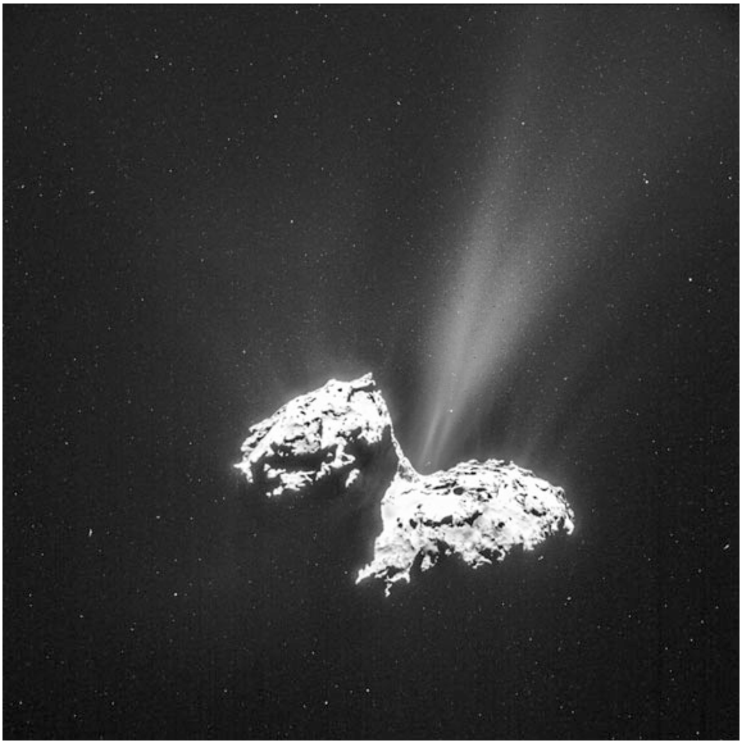


Fig. 3.3 Comet 67P/Churyumov-Gerasimenko as seen by the Rosetta spacecraft from a distance of 124 km on February 6, 2015. The comet's nucleus resembles the look of a "rubber duck" (credit: ESA/Rosetta/NAVCAM, CC BY-SA IGO 3.0)

The following table provides the diameters of a few selected comets.

Comet	Diameter (km)
1P/Halley	4.5–5.6
2P/Encke	1.3–2.9
9P/Tempel	2.3–3.1
C/1995 O1 Hale-Bopp	13.5–56.0

Formation of Cometary Nuclei

Detailed surface images taken by spacecraft show very complex surface structures including mountains, valleys, craters and so on (see Fig. 3.4). Yet, how does a nucleus form?

Several theories have been developed. One of the first models was the so-called sandbank model which assumed the cometary "nucleus" to be a gravitationally bound swarm of dust particles with adsorbed gases, orbiting the Sun. No monolithic structure was foreseen in this model. Over time, this model got less and less popular and finally had to be discarded when the first images of the nucleus of a comet were

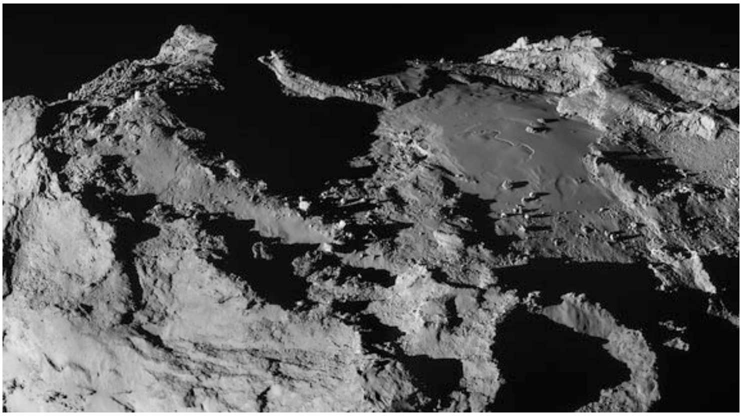


Fig. 3.4 The image shows the Imhotep region on comet 67P/Churyumov-Gerasimenko as seen by the Rosetta spacecraft from a distance of 19.9 km on 28 March 2015. The image has a resolution of 1.7 m/pixel and measures 3.1 × 1.7 km (credit: ESA/Rosetta/NAVCAM, CC BY-SA IGO 3.0)

taken by Vega 2 and Giotto which left no doubt that a cometary "nucleus" could not be a swarm of particles.

Already in the early 1950s, the astronomer Fred Whipple, one of the founding fathers of modern cometary science, introduced the "icy-conglomerate" model for a cometary nucleus which envisioned a single monolithic body composed of a mixture of volatile ices and "meteoritic material". The term "dirty snowball" probably describes the essence of this model best. The number of its supporters grew rapidly over the following decades though it underwent several modifications and refinements in order to better match the observations. One problem that was not properly addressed by the original model was that contrary to observational results, volatile compounds, such as methane (CH_4) and ammonia (NH_3), should not be present in a cometary nucleus for a very long time. Thus, there had to be a mechanism which ensured that these compounds were hold in the nucleus for longer period of times, e.g. several orbits.

The "clathrate-hydrate model" deals with this. Therein, it is assumed that these compounds are embedded in a crystalline structure of water ice, in so-called clathrate hydrates. Just imagine the volatile molecules are locked in a cage of hydrogen molecules which makes them less volatile since more energy is needed to break up the cage and subsequently free them.

Yet, despite all the modifications made, one big problem remained with this model: Cometary nuclei frequently tend to break up or dissolve which seems incompatible or at least raises doubts with the postulated existence of a solid, monolithic core. Nonetheless, it sustained for many decades.

In the 1980s, the interest in cometary science increased owing the return of Halley's comet in 1986. Two similar models emerged during that time: the "fluffy aggregate" and the "primordial rubble pile" model.

The basic idea behind these two models is that a comet's nucleus is an aggregate of smaller icy planetesimals, sometimes also referred to as "cometesimals". These icy components were brought together at low velocity in a random fashion. Since hardly any modifying processes exist on a comet, such as erosion or an atmosphere, the nucleus would, according to these models, preserve its highly irregular initial shape and very porous, easily fragmented structure during their whole lifetime. Their shape and structure would only be altered due to external influences, e.g. when approaching the Sun.

The primordial rubble pile model resembles the idea of rubble pile structures of asteroids. However, it was suggested that the nuclei were original material from the protoplanetary disk. Probably collisions among the cometesimals played an important role in the formation of the cometary nuclei.

Though these two models suggest being a composition of smaller elements, we should not confuse it with the sandbank model. The fragments do not form a swarm of particles as in the sandbank model but are in actual contact in a single nucleus structure and are weakly bonded or gravitationally bound.

Another model, the "icy-glue" model was developed in 1986. It suggests that comets are composed of porous refractory boulders with compositions similar to outer-main belt asteroids, cemented together by icy-conglomerate glue forming a more or less solid crust. The boulders would be responsible for the formation of the observed highly irregular topography of cometary nuclei. Yet, it failed to explain many of the features of cometary breakups, such as the breakup of comet Shoemaker-Levy 9 (D/1993 F2) in 1994.

Nowadays, the fluffy-aggregate and the primordial rubble pile models are thought to provide the best explanations of the underlying structure of the cometary nucleus. They best suit to explain most features observed (e.g. jets, bursts and breakups).

Approaching the Sun: The Nucleus Awakes

To better understand the remaining characteristics of a nucleus, it is necessary to see what actually happens during a comet's approach of the Sun when it becomes active.

During an approach dramatic changes are triggered on a comet. While being in the outer solar system the incident solar energy is too weak to cause any changes. The nucleus stays cold and solid. However, the closer the comet comes to the Sun, the more solar energy arrives at the nucleus and heats it up. Usually in a distance smaller than 7 astronomical units (AU), the comet's nucleus turns active.

A process called sublimation begins. In this process, substances transit directly from their solid phase to the gas phase without going through the intermediate liquid phase. You can easily observe this phenomenon with dry ice which is nothing more than frozen carbon dioxide (CO_2). At room temperature it sublimes and results in a fine, gaseous nebula.

It is not difficult to understand that the warmer the nucleus gets, the easier more solid compounds can sublimate. At a distance of about 6.5 AU carbon monoxide (CO) is freed, at 6.0 AU ammonia (NH_3) and at distances lesser than 5 AU even less volatile compounds such hydroxyl-radicals (OH), molecular carbon (C_2) and methane (CH_4). However, there is a natural end to this. The temperature of the nucleus cannot arbitrarily increase. An effect which we may call “anti-greenhouse effect” works against it as the sublimation cools down the cometary nucleus. You can best understand this effect by considering an analogous process. Think about humans sweating. The evaporation of sweat from the skin (like the nitrogen gas leaving the surface) has a cooling effect because the evaporated sweat carries the body’s excess heat away. In the same way outgassing material takes away energy from the nucleus and cools it down. Thus, there is a natural limit to the temperature increase though this limit cannot be generalized as it varies from comet to comet depending on its composition.

The sublimation of water ice, which begins at approximately 3 AU, is considered to be the dominating factor for a comet’s ejection of dusty material. It is important to know that this process takes place on the nucleus’ surface as well as beneath it. Besides ices and gases a typical cometary core also contains more solid, mineral compounds such as silicates, iron oxides and chondrites (carbon). You can think of these compounds as dust covering parts of the nucleus and being bound in its interior. When water ice sublimates, the water gas also carries away some of the dust. The solar radiation and wind then blow it away from the comet and results in a dust tail pointing on a curved path away from the Sun.

Usually, the dust particles are extremely fine having only masses of about 10^{-16} g. However, larger fragments may also break apart. The latter ones then may form secondary cores with a very limited lifetime. These may be visible in telescopes thereafter. A complete breakup of the cometary nucleus may also be the case resulting in so-called “split comets”. One very famous representative of this group was Shoemaker-Levy 9 which was torn apart by Jupiter’s tidal forces exerted on it while approaching the planet.

The fate of the finer particles is uncertain. Whether they are completely freed from the nucleus and move into its coma and tail or whether they fall back on the surface depends on various factors such as gravity, the exact particle size, the density and velocity of the gas that carries them away.

If they do not make it into the coma and tail and fall back onto the surface, the dust particles will slowly cover the surface and form an isolating crust. This crust makes the nucleus less sensitive to incident sunlight. We can particularly observe this with old comets, i.e. comets having already orbited the Sun several times. Such old comets tend to exhibit a lower activity. The scientific community is not sure but it might be the case that some asteroids in our solar system may indeed be extinct comets. Only detailed analysis of their composition will tell us the truth.

Activity at Greater Distances?

So far, we have seen that sublimation is the driving factor for a comet’s activity. Contrary to most comets which only show activity from about 6.5 AU distance to the Sun, some are already active beyond this limit in the far outer regions of our solar system.

The great comet of 1996, Hale-Bopp, is merely one but remarkable representative of this small group. The activity of these comets, however, cannot be explained by classical sublimation processes as we have seen them before. In the regions of our solar system beyond 6.5 AU distance from the Sun, it is simply too cold to trigger sublimation for most volatiles. Thus, other sources for the activity have to be found. The scientific community cannot give a definite answer yet. It is simply very difficult to make observations of small objects like almost inactive cometary nuclei at these great distances.

A currently promising theory is the transition of amorphous ice to its crystalline structure. But what does that mean? Looking at Earth we know that the molecules in water ice are perfectly ordered in a hexagonal structure named the "crystalline form".

Contrary to this, in amorphous ice the molecules have no special order but are arranged irregularly. This type of ice further seems to be the dominant form of which water exists in the interstellar medium. The crystalline structure is merely an exception to this. Amorphous ice, similar to clathrate-hydrates, is capable of encaging other molecules. During the transition from amorphous to crystalline form, which can already take place at temperatures far below those required for sublimation, the embedded molecules can be freed and cause activity of the nucleus. The surface temperature of the nucleus in these far regions of the solar system is sufficient to trigger the transition.

Outbursts and Jets

Another phenomena often observed with comets are sudden increases in brightness. We can distinguish two different types. First, there may be a general increase in brightness which is more or less evenly distributed over the nucleus and coma (see Fig. 3.5). This type is usually referred to as "bursts" or "outbursts" of gas. Most observed bursts are directly related to the comet's activity and may occur while approaching the Sun and when departing from it. With only few exceptions, bursts tend to happen while the comet resides in the inner solar system. The likelihood for bursts further away decreases with increasing distance to the Sun.

Several causes for bursts have been discussed in the scientific community. Yet, it is still not clear what their actual nature is. We can, though, assume that not only one possibility is true. As we have seen, the nucleus' surface is subject to strong thermal and mechanical tensions caused by incident sunlight and sublimation. These tensions may then in turn lead to fissures and cracks on the surface. In case there is an embedded gas cavity below these cracks, further tensions may result in a sudden, explosive release of volatile gases caged in the cavities.

One can easily understand why this happens during the comet's approach to the Sun and the heating up going along with it. But then, how can bursts occur after the comet passed its perihelion and departs from the Sun? Should the nucleus not cool down in this case? This notion is only partially true. The warmth caused by incident sunlight moves down to lower layers of the comet's nucleus over time. The nucleus may hence act like a heat accumulator. In this way, after some time the warmth may

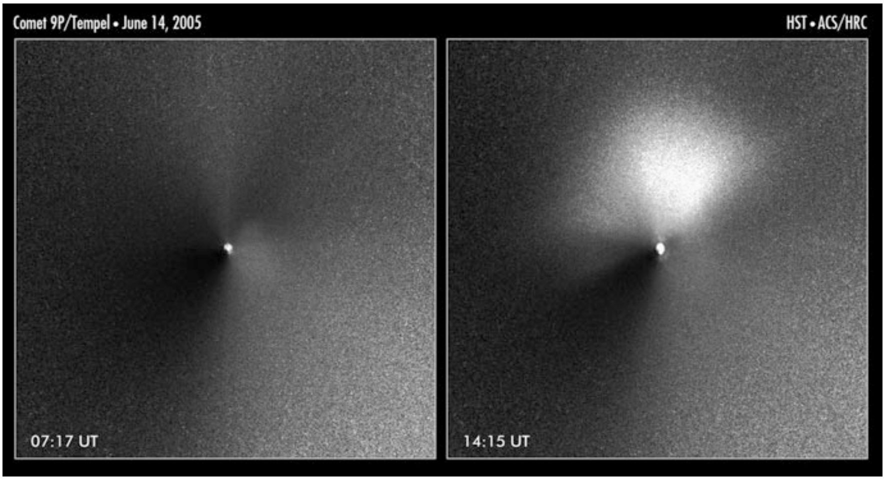


Fig. 3.5 The Hubble Space Telescope observes an outburst on comet 9P/Tempel on June 14, 2005 (credit: NASA, ESA, P. Feldman (Johns Hopkins University), and H. Weaver (Johns Hopkins University Applied Physics Lab))

reach deeper lying cavities and set their embedded gases free while at the same time the surface cools down.

Another explanation was given by Fred Whipple who suggested that collisions with formerly segregated smaller fragments of the core might fall back on the surface and cause the ejection of material.

In both cases, dust will be blown above the surface and into the coma. Sunlight is thereafter reflected by these dust particles and lead to a nonspecific increased brightness and visibility.

The second type of phenomenon is jets which describe a focused ejection of gas and dust (see Fig. 3.6). Depending on the rotation of the nucleus, their characteristics vary. In addition, the rotation of a comet, due to its irregular shape, may appear to be quite chaotic. Just imagine a fast rotating core and a focused jet streaming out of it. The jet will inevitably have spiral structure. Perhaps it is easier to understand with an analogy. Think about a water hose blowing out a strong stream of water. Now, imagine, what happens when you move the hose around. The stream of water will take several structures depending on the movement. This is what will happen to a cometary jet, too. The velocity of the jet strongly depends on the temperature causing the gas and dust to eject. As is the case with bursts, together with the ejected gas dust comes along and will reflect the sunlight.

The causes for jets are yet not understood. Most probably they emerge from smaller gaps or holes in the cometary surface which take the effect of a nozzle.

Active and Inactive Regions

The effects sublimation on the small nucleus provides are tremendous. Hence, is it the whole nucleus or at least its whole surface that is active and causes the phenomena we have seen so far?

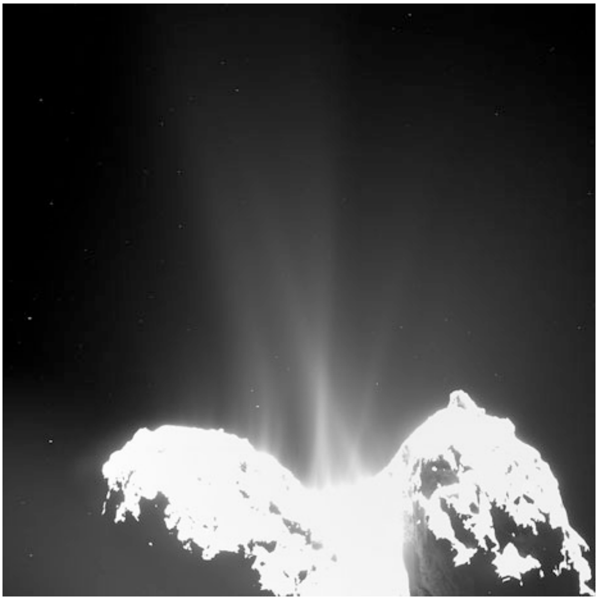


Fig 3.6 This image of comet 67P/Churyumov-Gerasimenko, taken by Rosetta in September 2014, from a distance of 7.2 km, shows jets of dust and gas streaming into space from the neck of the comet's nucleus (credit: ESA/Rosetta/MPS/UPD/LAM/IAA/SSO/INTA/UPM/DASP)

More or less until 1986, this was the widely accepted notion. Only back then, the year of the first direct close observations of Halley's comet by spacecraft of the Halley Armada, foremost Giotto and the Vega probes, changed the perspective.

One of the biggest surprises was that the first images taken revealed only selective parts of Halley's nucleus being active. These active regions showed strong outgassing including jets while the inactive regions remained totally frozen. Figure 3.7 depicts some of these active zones on images taken by Giotto.

Subsequent observation and analysis indicated that indeed only small parts of cometary nuclei are active (typically only 10–20 %). The table below shows a selection of some comets with their average amount of active regions.

Comet	Amount of active regions (%)
P/Wirtanen	25
P/Halley	10
P/Encke	0.5–10
P/Churyumov-Gerasimenko	1.4

It has yet to be fully understood what causes the distinction between active and inactive regions. Thick layers of dust covering areas of the nucleus may render these inactive.

Composition of a Nucleus or What Is It Made of?

Now that we have seen the main characteristics of a cometary nucleus and some forming elements, the question remains of what it is composed of and how one can find out.

Fig. 3.7 Halley's Comet as seen by the Giotto spacecraft on March 13, 1986. The nucleus reveals active and non-active regions (credit: ESA/MPAe Lindau)

The question, at first glance, seems to be quite easy to answer. However, having a closer look on the matter shows the difficulties involved. The nucleus itself is not directly observable from Earth and sending a spacecraft to each comet of interest might not be a feasible solution for obvious reasons.

Consequently, indirect techniques are required. Why not measure the volatile compounds and dust particles that were ejected from the core? Spectral analysis seems to be the method of choice and has been applied ever since the second half of the nineteenth century. First results showed that the observed spectra included the continuous spectrum of our Sun. This means that sunlight is reflected by the dust particles. Fortunately, also some spectral lines were found indicating the presence of some ionized molecules. Yet, capturing a comet's spectrum is a challenging task as it is continuously changing while approach the Sun and the amounts of detectable compounds may be faint.

It would be a fallacy to assume that the spectra alone could provide knowledge about the chemical composition of a comet's nucleus. Why that? If we have a closer look at the detected molecules, we can see that these ions and radicals in these spectra are unstable. This suggests that these unstable "daughter" molecules were derived from more stable "mother" molecules by some chemical processes. How could this work? The outgassed mother material from the nucleus is subject to the ultraviolet (UV) light emitted from the Sun which alters these molecules by ionization and dissociation processes.

Ionization may be caused by collision with other outgassed molecules or directly through the interaction with UV light. Typically molecules or atoms contain the same number of positive (protons) and negative (electrons) elements leaving them in a neutral state. Looking at the simplest chemical compound, hydrogen, it turns out that neutral hydrogen contains exactly one proton and one electron. In more complex atoms, neutrons are also included. Ionization is the process which causes a neutral molecule to become either negatively or positively charged.

There are basically two possibilities what can happen. If free electrons (negative) collide with a molecule and are subsequently trapped, it forms a negatively charged ion of the mother molecule. On the other hand, positively charged ions are produced by transferring sufficient amount of energy to a bound electron. As soon as the induced energy exceeds a certain threshold, the electron is freed leaving an energetically imbalanced atom or molecule, an ion, behind. In this case, since a negatively charged electron left, the remainder is positive charged.

Dissociation is the second process playing a key role. It is a general process in which molecules split into smaller particles such as atoms, ions or radicals by inducing energy e.g. in form of light or other electromagnetic radiation. Dissociation is usually reversible, i.e. when the induced energy is removed the original molecule forms again.

Material outgassed from the cometary nucleus is subject to these processes. Applying this knowledge helps to establish the original mother molecules from the observed daughter molecules. Thus, the daughter molecules CO^+, CN, CH were derived from the mother molecules CO, C_2N_2, CH_4 and CO_2.

These processes usually take place in the so-called cometary coma. Fortunately, besides these indirect methods, in the meantime some new techniques employing infrared and radio telescopes exist which allow direct observation of the cometary nuclei. Their results confirmed the previously gained indirect results.

3.3.2 Coma

As we have seen in the previous section, the sublimated volatiles and dust leave the cometary nucleus. These particles then form part of the comet's "atmosphere" or "coma". The coma, a word being derived from the Greek word "kome" meaning "hair", is a nebulous envelope around the nucleus and can be very large and extended, e.g. between 50,000 and 500,000 km. It gives the comet a fuzzy appearance when looking through a telescope. That is why, for a very long period of time, the coma was thought to be the actual core and thus the comets were supposed to be very large objects of several hundred kilometers in diameter. Nowadays, we know that this assumption is several magnitudes larger than the real size which is more closely towards several tens of kilometers.

The coma is generally made of water (H_2O) and dust. The force exerted on it by solar radiation and the solar wind forms the comet's tail.

The coma, together with the tail, is what makes the comet visible since the core is usually very dark with a very low albedo making it to faint for observations with telescopes. However, both the coma and the tail are illuminated by sunlight. The dust particles reflect the light and the gases glow due to ionization.

Formation of the Coma or How to Become Fuzzy

Since the coma is a consequence of the sublimation process taking place on the comet's nucleus, as we have just seen, it is not always present. For most of the lifetime of a comet, the comet merely consists of its core and nothing else. The coma begins to form approximately when the comet passes the orbit of Jupiter and subsequently enters the inner solar system.

The size of the coma typically grows while the comet approaches the Sun. The closer the comet comes to the Sun, the more material is sublimated and casted into the “atmosphere”. However, there exists a natural antagonist to the growth. The solar wind becomes stronger closer to the Sun and thus blows more gas and dust particles away from the coma. At a certain distance which differs from comet to comet more material is taken away from the coma then is inserted into it by sublimation. This even causes the coma to shrink. This threshold distance, yet, varies.

Every comet has a different composition and thus the sublimation process may vary accordingly. There exist more active comets, like C/1995 O1 Hale-Bopp, or less active comets such as 67P/Churyumov-Gerasimenko.

The coma typically takes spherical form around the comet’s nucleus. In its inner part, “inner coma”, the density of particles is relatively high. They, thus, collide with each other often and are scattered evenly in all directions which causes the spherical shape.

A comet’s coma consists of a multitude of different molecules. The following table shows an exemplary composition, the one of comet Hale-Bopp. The different components are normed to hydrogen (H_2O).

Hydrogen (H_2O)	100
Carbon monoxide (CO)	20
Carbon dioxide (CO_2)	6–20
Formaldehyde (H_2CO)	1
Methanol (CH_3OH)	2
Ammonia (NH_3)	0.7–1.8
Methane (CH_4)	0.6
Acetylene (C_2H_2)	0.1
Ethane (C_2H_6)	0.3
Hydrogen-sulfide (H_2S)	1.5
Formic acid (HCOOH)	0.06

This composition, especially the existence of hydrogen-sulfide, a molecule also created during decomposition, would cause a scent of rotten eggs if you actually tried to smell it.

Two different regions of the coma can be identified: the inner and the outer coma. The distinction comes about the different processes that take place in each of them, as we will see in the following. The inner coma is subject to the ejection of material from the core and ionization and photodissociations processes while the interaction with the solar wind is the dominating force of the outer coma.

Inner and Outer Coma

The “inner coma” extends up to about 1,000 km from the surface of the nucleus. Especially towards its lower levels, closer to the core, many uncertainties exist, as this zone is very difficult or even nearly impossible to observe from Earth. Most knowledge is derived from models, theories and observations of spacecraft.

What we know is that the inner coma is filled with neutral and ionized gases as well as with dust particles. They originate, as already discussed, from the nucleus and are transformed into daughter molecules by dissociation and ionization processes induced by the interaction with solar radiation. In addition, collisions between the particles may also play a major role since new compounds may be formed thereby. Molecular carbon (C_2), an important molecule with comets, may be formed in a reversible process when two carbon atoms collide.

$$C + C \Leftrightarrow C_2 + h\nu$$

where hν is an amount of energy (light quantum) that is set free in case of a collision.

The ionized gases form a plasma. A plasma is one of the four fundamental states of matter. The others are solid, liquid and gas. However, the characteristics of plasma are quite different from those of the other states. It is a medium of unbound positive and negative charges. Usually, the overall charge of the plasma is neutral ("zero"). Though being unbound, the particles are not free in the sense of being totally unaffected by each other. They still experience forces. When a charge moves it generates an electric current with a magnetic field. This field then influences other magnetic fields such that the charges are affected by each other's fields.

The plasma within the coma interacts with the solar wind in the "outer coma" which can extend to 10,000 or even 10 million kilometers. The density of particles in the outer coma is low compared to the inner coma. Thus, almost no collisions between particles take place. The dominating factor in this part of the coma is the interaction of the coma's plasma with the solar wind which in itself is a plasma. The clash of solar wind and coma causes in first instance a shock wave front. When the comet comes sufficiently close to the Sun, the solar wind's plasma and its magnetic field accumulate and enable a flow of particles around the comet. Thereby parts of the coma's plasma are torn away together with solar wind. This kind of "kidnapped" plasma then forms the comet's plasma or ion tail. The coma's dust particles on the other hand are the foundation of the dust tail. We will come to discuss both types of tails later on in this chapter.

3.3.3 Tails or What Makes a Comet Look Like a Comet

Probably the most impressive parts of a comet and what actually makes a comet to a comet in the view of most people are its tails. A comet's nucleus is usually quite small (several tens of kilometers in diameter), the coma already extends significantly to several hundreds of thousands kilometers. Some comets are known where their coma was larger than the Sun (roughly 1,400,000 km). However, this is nothing compared to the comet's tail which can stretch to several million kilometers. Comets with tails of up to 4 AU, i.e. about 596 million kilometers, have been observed.

The comet tails, as is the case with the coma, becomes only visible while the comet travels through the inner solar system. As we have already seen, when a comet crosses a critical distance from the Sun it awakes and becomes active. The warming caused by solar radiation allows cometary ices to sublimate into the coma. The ejected gases also take away dust particles present in the nucleus and blow them into the coma. The force exerted on the coma by the Sun's radiation pressure and solar wind then result in an enormous tail to form which, as we will see, points away from the Sun.

To better understand the mechanisms of cometary tail formation, we first need to know what "solar radiation pressure" and "solar wind" are.

Solar Radiation Pressure and Solar Wind

Our Sun continuously emits a huge amount of electromagnetic radiation. You can think about this radiation as a continuous flow of particles, the so-called photons. Photons, sometimes also called quantum of lights, stream through the solar system and exert forces on objects each time they hit an object and are subsequently absorbed or reflected. The momentum carried by the photon is then at least partially transferred to the object which is then pushed a bit away.

The force of a single photon hitting a particle is very small. However, since there are so many of them hitting on the particle, they exert enough energy to push the particle away from the Sun. This is probably better understood with an analogy from our daily life. Let us assume, the particle is resembled by a big heavy ball, such as basketball. Now you throw a tiny table tennis ball, the equivalent of a photon in our example, towards the basketball. As it hits the basketball nothing will happen. The small table tennis ball is simply not powerful enough to affect any changes to the much bigger and heavier basketball. However, if you start to throw many hundreds of table tennis balls on the basketball all at one time or with very short periods in between, the ball will essentially move a bit. This is what happens to particles in space. We call this effect solar radiation pressure.

The solar wind, on the other hand, is not related to photons or quantum of lights but is a continuous stream of plasma released from the Sun's upper atmosphere. It mainly consists of electrons, protons and helium nuclei, so-called alpha particles. Other ionized atoms and molecules are hardly found. The stream though being continuous is not constant but varies in density, temperature and speed over time. When a large solar eruption occurs, variations in the solar wind are present. Such solar eruptions are regularly covered in the media with stagy headlines such as "giant solar storm about to hit Earth and likely to cripple Earth's infrastructure".

It is true that the flowing plasma highly influences magnetic fields and interacts with other charged particles such as the ionized gases in a comet's coma. The effects of these interactions on Earth though are usually manageable. The plasma streams in outward direction through our solar system. Over time it slows down, also because of interstellar wind flowing in the inward direction and hitting the solar wind. At a certain distance in space both the solar wind and interstellar wind come to a halt and this is the zone, the heliopause, where our Sun's influence ends. It is not exactly known where this heliopause is. Measurements by the spacecraft Voyager suggest that this is at about four times the distance Sun-Pluto.

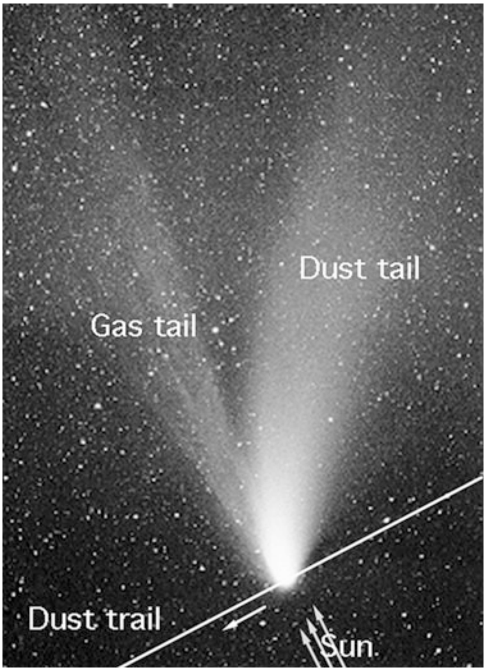


Fig. 3.8 The images shows the two different types of tails (credit: NASA/K. Jobse, P. Jenniskens)

Types of Tails

The tails of a comet are formed by gases and dust particles which are blown away from the cometary coma by the solar radiation pressure and the solar wind. Most frequently, two types of tails become apparent (see Fig. 3.8). Although in some cases other tails may exist, too.

The streams of gas and dust each form their own distinct tail, as we will see in the following. Each of these tails points into slightly different directions, as you can also see from Fig. 3.8.

The narrow and elongated tails of type I, the gas or ion tails, usually consist of ionized gas molecules torn away from the coma. Solar radiation pressure alone is not sufficient to explain the formation of gas tails. German astronomer and physicist Ludwig Biermann (1907–1986) postulated in 1951 that some kind of particle radiation emitted from the Sun is the most relevant factor in their formation. Nowadays, his basic idea is generally accepted. The particle radiation he assumed is found to be the solar wind.

Type II tails mainly consist of small dust particles pushed away from the coma by solar radiation pressure. These slightly curved and widely diversified dust tails are in general the most prominent and visible parts of a comet.

Very rare are type III tails which only exist under certain orbital constellations where a so-called antitail becomes visible (see Fig. 3.9).

On top of this “classical” categorization, only more recently discovered tails such as sodium and iron tails may exist with a comet.

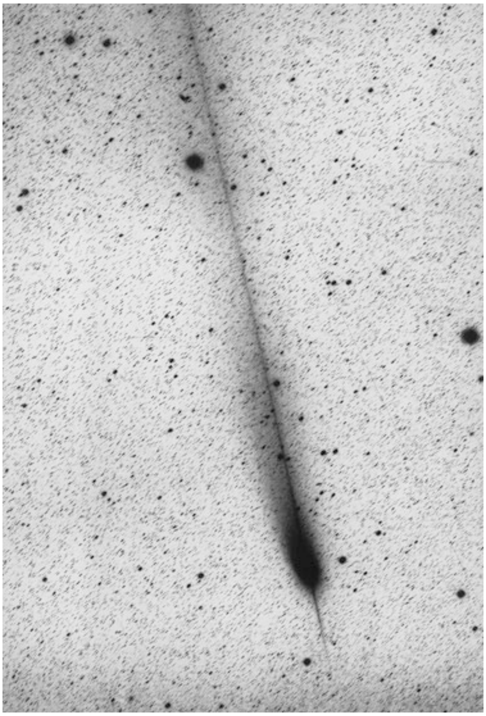


Fig. 3.9 Comet Hale-Bopp photographed on January 5, 1998. The photo shows the anti-tail of Hale-Bopp (credit: ESO)

Dust Tail

The dust tail is probably the most striking feature of a comet. It is usually very bright and curved, though being indistinct and structureless most of the time. The curved dust tails usually do not follow the comet's orbit as e.g. the gas tails do. This makes it possible to observe two distinct tails. Yet, how do dust tails form? The processes involved differ between the dust tails and gas tails.

We have seen that dust particles usually are ejected from the cometary nucleus into the comet's coma. There, they can be influenced by the solar wind. To better understand how the dust tails form, it is necessary to know which forces are involved and exhibited on dust particles. As soon as the dust particles leave the comet and are no longer influenced by it (gravitationally), they follow their own paths or orbits. Their initial velocity is somehow similar to the orbital velocity of the comet. They are only affected by basically two forces: the solar radiation pressure and the Sun's gravitational force.

We have seen that the smaller the dust particle the higher the influence of the solar radiation pressure. The radiation pressure is triggered by the absorption and or reflection of photons and increases with the surface of the particle by its diameter squared. For gravity, the mass of the particle is relevant. Both are two countering forces. The solar radiation pressure tries to push the particle away from the Sun while the gravitational force pulls it towards the Sun. This, however, means that the higher the solar radiation pressure, the lower the effect of gravity is and vice versa. It is thus the combination of both forces that defines a dust particle's path in space.

We can distinguish four different scenarios. First, if the dust particle is very large, the effects of the solar radiation pressure are negligible. It will thus remain close to the cometary coma or even fall back onto the surface. Secondly, if both solar radiation pressure and gravity are equal, no force is effective. The particles will leave into space radially from the comet. Their path is only determined by their velocities and size. Thirdly, if gravity overweighs the solar radiation pressure, the particle will behave as if the Sun was lighter. The gravitational influence is merely reduced by the radiation pressure. In such as case, the relative movement of the particle to the comet is slower. Fourthly, if the solar radiation pressure is stronger than gravity, the particles will move on hyperbolic orbits away from the comet.

Eventually, the acceleration of the dust particles is dependent on their surface (solar radiation pressure) and mass (gravity) can vary widely since the size of the dust particles can be very different. Based on the four scenarios, this will result in a widely spread dust tail. The sunlight is reflected by the small particles and the tail becomes visible.

Streamers and Striae

Dust tails are often featureless. Yet sometimes structures can become visible. This is often the case with larger comets. The most prominent features are the so-called streamers. They are caused by a sudden, explosive ejection of large amounts of dust from the cometary nucleus. The increased amount of dust then becomes visible as very bright straight elements in the tail. A streamer follows the curve of the tail and usually reaches to the coma.

Striae are more different and do not occur that often. Striae are wide, straight, almost parallel rays of dust. They originate from streamers (see Fig. 3.10).

Fig. 3.10 Comet McNaught reveals streamers and striae in January 2007 (credit: S. Deiries/ESO)

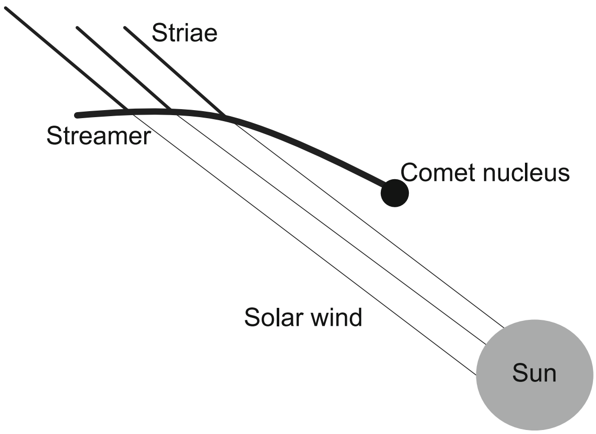


Fig. 3.11 Evolution of striae on the convex part of the dust tail

Therefore, in order to have striae it is important to have an explosive ejection of dust from the nucleus. This will most likely create a streamer.

The ejected dust, comprised in the streamer, consists of a large variety of differently sized dust particles. Each particle size will then follow its own path on space depending on solar radiation pressure and the gravitational force of the Sun. Hence, the particles will separate at the convex part of the dust tail appearing then as lined structures therein (see Fig. 3.11).

Gas Tail

The appearance of the gas tail differs dramatically from that of the dust tail. A comet's gas tail is formed by ionized gases blown away from the coma. The solar wind is the driving force behind its creation. Though, being usually not as impressive as the dust tail for the unaided eye of the observer, gas tails can be extremely long, even longer than dust tails. They may stretch over several million kilometers.

In 1996, the spacecraft Ulysses flew through the border region of comet Hyakutake's (C/1996 B2) gas tail which could be clearly measured by the probes instruments. Hyakutake's nucleus though had already been about 570 million kilometers away.

As we have seen earlier in this chapter, the gases and dust particles being ejected from the cometary nucleus consist of electrically neutral mother compounds (atoms and molecules). Influenced by the Sun's radiation the gases get ionized in the inner coma. These ionized gases form a plasma of low density. The plasma reacts sensitive towards magnetic fields and thus interacts with the solar wind. Its transported by the magnetic field in the "outer coma". Both this magnetic field and the coma influence each other. The solar wind is slowed down when encountering the coma and is folded around the comet's nucleus.

By this, the gas particles are torn out of the coma and carried away in anti-solar direction by the solar wind. The acceleration is extraordinarily high. By picking up the ionized gases, the solar wind gets slowed down a bit. Turbulences may occur on the far side of the Sun. These become visible as a structure in the tail. Through the interaction with the solar wind the gas tail illustrates Sun's magnetic field. Alterations in the field density or changed polarities may cause the gas tail to be disconnected, though being reformed shortly afterwards.

Most gas tails appear bluish to the human observer. The reason for this is the composition of it. The most common ion in the coma and therefore also in the gas tail is CO^+ (the ion of carbon monoxide) which scatters blue light better than red.

3.3.4 Orbital Characteristics or Where Do They Come from

Not only their special visual appearances due to their tails make comets interesting and different from other objects in our solar system. Already early observers noticed that the cometary orbits looked at first glance completely arbitrary. They seemed to come from all possible directions. This, indeed, is something that distinguishes them from the other solar system bodies.

Almost all of the other objects have similar orbital characteristics: their respective inclination, the form of their orbits, etc. only vary within very limited ranges. If we have a look at the planets for example, we notice that their orbits almost lie in one plane, the ecliptic. Only minor deviations exist. Additionally, they orbit the Sun on near circular paths, i.e. the eccentricity of the ellipses is very small. The smaller solar system bodies, such as asteroids, sometimes deviate from this more or less considerably but still within certain boundaries. These perturbations of their orbits are caused by the planets. Especially the gas giants have a strong influence on orbits of the smaller bodies. Just think about the formation of the Kuiper Belt.

What does it mean when these objects share similar orbital characteristics? We can be quite sure that they have a common background. They were supposedly formed by the same processes taking place in the protoplanetary disc (see Chap. 1).

The comets' orbits, however, are so different as to suggest another origin. If we have a closer look at their orbits, we see at first glance that these are often extremely elliptical and highly inclined.

They may have been scattered and thrown into their extreme orbits by the giant gas planets, first and foremost Jupiter. How does this work? Let us assume that a smaller object such as a comet approaches Jupiter. Events like this happen regularly even today. The comet's orbit is then gravitationally influenced by the gas giant. The closer its orbit brings the comet to Jupiter, the stronger the perturbations. Encounters with the "lord of the planets" only have a limited amount of results. Firstly, the comet may be captured and forced into orbit around it or to collide with Jupiter. A famous example for a collision with Jupiter is comet Shoemaker-Levy 9 about which we will talk in the later part of this chapter in more detail.

Secondly, the gravitationally influence may simply force the comet into another orbit but not capture it. Either, on the one hand, the comet can be bound closer to the inner solar system by reducing the eccentricity of its orbit and altering its orbital period, as was the case with comet Hale-Bopp whose orbital period was drastically shortened by its encounter with Jupiter during its previous journey into the solar system, just to mention a recent example. Or, on the other hand, the perturbations may force the comet on longer orbits leading them further away from the inner solar system. In extreme cases it may even be kicked out of the solar system. Since the perturbations depend on so many parameters such as distance to Jupiter, size and composition of the nucleus, you can imagine that each time a comet is affected, it will be forced on a different orbit. Over time comets will then be more or less evenly scattered in space. Such a process could indeed explain why comets appear to come from all directions.

Though we have seen why cometary orbits are different compared to other solar system objects, the question where comets originate remains of the moment unanswered. We will understand this when we have a more detailed look at their orbits.

What Do Their Orbits Look Like?

Comets, as all solar system bodies do, follow Kepler's laws with regard to their orbits. There are basically three possible types of orbits: elliptical, parabolic, and hyperbolic. These curves only vary in their eccentricities e. If the eccentricity is smaller than 1, the orbit is an ellipse (the smaller e the more similar to a circle). In this case, we can also speak of a closed orbit, since the comet will return again and again unless it gets extinct. If the eccentricity is larger than 1, the orbit is open and resembles a hyperbola. In this case, comets would only enter the inner solar system once and then leave forever and never return again. The borderline cases are parabolic orbits, having an eccentricity of 1. This type of orbit is often used to describe extremely elongated ellipses. In the following we will focus on elliptical and hyperbolic orbits.

Comets are commonly classified according to the length of their orbits. An arbitrary border line is drawn at orbital periods of 200 years. On the one hand, we speak about short-period or periodic comets when their orbital periods are below this threshold of 200 years. On the other hand, those comets having orbital periods of more than 200 years are called long-period. Though this decision seems to be completely arbitrary, there is also a physical background in it. If we have a look at the orbital inclinations of the comets, we can distinguish two populations. One group has orbits that almost perfectly lie in the ecliptic plane deviating only in a few degrees. The members of the other group have highly inclined orbits from almost all possible directions, i.e. isotropic origin. The first group corresponds to the short-period comets, the latter one to the long-period comets.

Long-Period Comets

Let's first have a closer look at the long-period comets. In addition to their arbitrary inclinations, they also have highly eccentric orbits, i.e. the value of e is close to 1. These orbits take them far beyond the outer planets to their aphelia. These comets further move around the Sun in all possible directions (retrograde and prograde). No preferences exist there. Their orbital periods go up from 200 years to several thousands or even millions of years. Despite their very elongated orbits, these comets are still gravitationally bound to the Sun.

These characteristics suggest that they originate from an "isotropic" reservoir, the Oort cloud. This cloud is named after the Dutch astronomer Jan Hendrik Oort

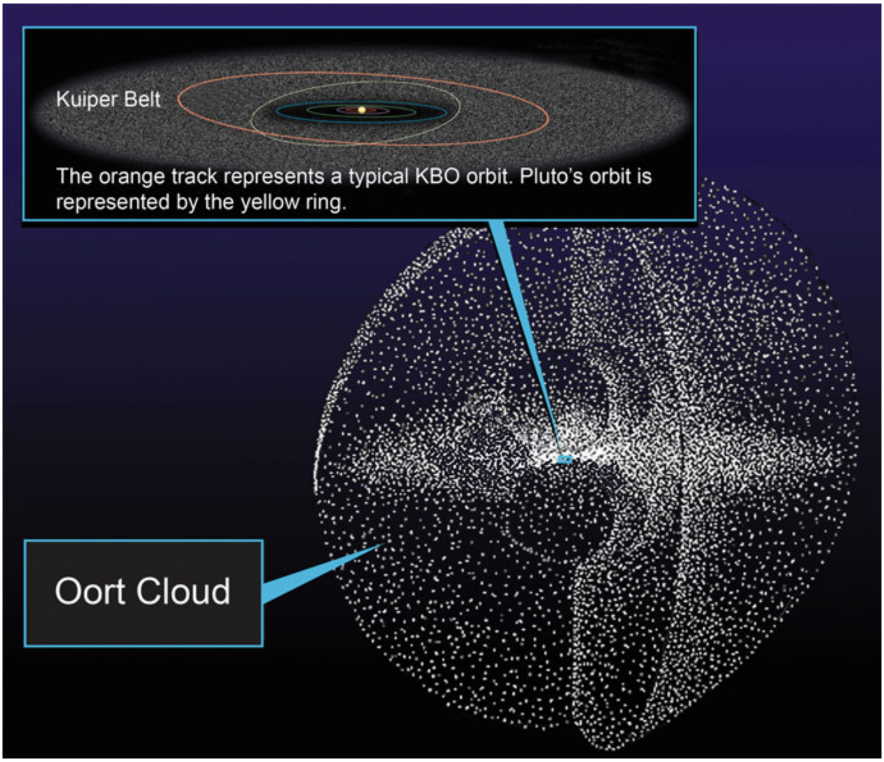


Fig. 3.12 An illustration of the Kuiper Belt and Oort Cloud in relation to our solar system (credit: NASA)

(1900–1992). He first postulated a spherical cloud around our solar system comprising a huge amount of icy planetesimals in it. The Oort cloud is believed to surround the Sun at a distance of up to around 100,000 AU (see Fig. 3.12).

Basically, the Oort cloud comprises two distinct areas. On the one hand, there is the doughnut-shaped inner Oort cloud ranging from about 2,000 to 20,000 AU. On the other hand, there is the spherical outer Oort cloud from about 20,000 AU to about 100,000 AU. Some astronomers believe the outer edge to be at even 200,000 AU which corresponds to roughly 3 light-years. This is a quite remarkable size. Just imagine Proxima Centauri, the closest known star to our Sun, is only 4.2 light-years away.

The outer Oord cloud is only loosely bound to our solar system, so that its objects are easily affected by passing stars or even the Milky Way itself. Occasionally, one of the icy planetesimals is dislodged from its almost circular orbit within the cloud and thrown into the inner solar system. A new comet is born. Though this may sound plausible it is difficult to prove it with hard scientific evidence. The outer Oort cloud is just too far away and its members are simply too faint and small to be observed by Earth-based telescopes. Space probes may be the only means to confirm the outer Oort cloud. Yet, Voyager 1, currently the fastest and farthest of all launched interplanetary probes will only reach the Oort cloud in about 300 years, a time by which its energy will already have been exhausted for a very long period of time.

Though it might contain billions of objects, the outer cloud is scarcely populated having tens of millions of kilometers between neighboring objects. The huge number of objects should also not deceive us about its total mass. New model predictions suggest it to be only about five times the mass of Earth.

The Oort cloud formed during the early stages of the formation of our solar system. Scientists believe that its objects, the icy planetesimals, formed much closer to the Sun and were then scattered far into space by the gravitational influence of the gas giants in the early days of our solar system either into what would form the outer Oort cloud or out of the Sun's influence. Indeed, current models of the Oort cloud suggest that the number of comets ejected into interstellar space outnumbers the ones retained in the cloud by a factor of at least three. This, however, raises some questions. By injecting comets into the inner solar system, the cloud gradually depletes. According to some models, and looking at the age of our solar system, the very thin outer cloud should have already run out of objects. There has to be source for resupplying it with new material.

In 1981, the American astronomer Jack G. Hills, proposed the existence of an inner part of the Oort cloud. The names Hills cloud and inner Oort cloud are used interchangeably in the scientific community. Other than the outer cloud, the inner one is disc-shaped, roughly like a doughnut. It is assumed that it has tens or hundreds of times more commentary nuclei than the outer cloud. The Hills cloud could be the origin for replenishment and thus explain the existence of the outer Oort cloud for billions of years.

Visitors from Interstellar Space

While a large number of long-period comets originates from the outer Oort cloud, there is assumed to be another source. If we consider the way our solar system formed to be representative for other solar systems, those should have gone through similar processes. Probably they are also surrounded by an isotropic reservoir of comets, just as the Oort cloud and some comets should have been kicked out of their respective solar system. These "interstellar" comets travel through interstellar space and are no longer gravitationally bound to a star. During their journey they might enter our solar system and behave like a long-periodic comet. But how can you distinguish an interstellar comet from a long-periodic one originating from the Oort cloud? The former ones have strongly hyperbolic orbits indicating to that they are not gravitationally bound to our Sun. In addition, these comets should have speed higher than Sun's escape velocity. Unfortunately, until now, no such comet has been discovered and confirmed to come from beyond the solar system. Comets with weakly hyperbolic trajectories have been observed but these could be identified as originating from the Oort cloud. So, do interstellar comets really exist? The question is difficult to answer but most probably yes. Their density in interstellar space may merely be too low in order to make such events happen frequently.

The hyperbolic trajectories will cause such comets to enter the solar system and then leave it again and never return. They merely drop by for a short “hello”. However, on rare occasions, it may happen that due to gravitational interactions with Jupiter, an interstellar comet is forced into a heliocentric orbit while passing through the inner solar system. Simulations suggest that this only happen once every 60 million years. However, comets Machholz 1 and Hyakutake are possible candidates as their chemical composition is atypical for comets of our solar system. While this may hint to an interstellar origin, it does not preclude other explanations for their odd makeups.

Short-Period Comets

The other major population comprises the short-periodic comet, hence all those having orbital periods of less than 200 years. It is interesting to see that among this group the inclinations of their orbital paths do not vary so much. Moreover, they orbit more or less in the ecliptic plane in the same direction as the planets do and their eccentricity is much lower compared to long-period comets.

It is characteristic for most of the short period comets that their aphelia lie in the region beyond Jupiter’s orbit. Comets having aphelia close to one of the outer planets are grouped together into a “family”. We will learn about this soon. However, not all short-period comets have orbits that reach out to the gas giants. Some, comet P/Encke seems to be representative for them, have very short orbital periods and their respective orbits do not take them in the realm of the gas giants, i.e. their orbits completely lie within the inner solar system.

Yet, how does it come that these comets have so similar orbital characteristics? As you can imagine, being so close to the gas giants subjects these small bodies to further perturbations of their orbit. The gravitational interaction with the major planets forces them into the ecliptic plane over time. Most relevant in this respect is Jupiter as it is the source of the greatest perturbations. This is owed to its mass which is about twice as much as the masses of all other planets combined. The gravitational interaction with it may affect a comet’s orbit drastically. The gas giants are able to capture long-period comets and force them into short-period orbits. Most of the family members are thought to be original long-period comets.

While some of the family members may be former long-period comets, the vast majority of the short period comets are supposed to originate from the Kuiper belt or to be more precise from the scattered disk. This relatively disk located beyond Neptune’s orbit stretches from about 30 to 50 AU and is thought to be the original formation zone of most comets, long- and short-period, in our solar system. Some of them were subsequently scattered far into space, e.g. into the region in which the Oort cloud formed.

Recently, a new group of short-period comets residing in the main asteroid belt between Mars and Jupiter were found. These main belt comets have almost circular orbits. They further blur the line between asteroids and comets which were initially thought to be completely different but now turn out to have more and more in common.

3.3.5 Comet Families

A huge amount of comets exist and there is a clear need for a refined classification. Distinguishing between short- and long-period is simply too coarse grain. Yet, how should we do that? Short-period comets are group into so-called families. Comets having similar elliptical orbits are put together in a family.

The classification is thus based on orbital parameters. It is one option to apply Tisserand's parameter for this. This parameter takes into account the gravitational influence and the resulting orbital perturbations on small solar system bodies. The orbital changes can be quite drastically, e.g. expelling a comet from the solar system or forcing it by far more eccentric orbit. Tisserand's parameter can be used to distinguish different kinds of orbits, such as comets. The value of this parameter is calculated from several orbital elements (semi-major axis, eccentricity and inclination) of a relatively small object (e.g. comet) and a larger perturbing body (e.g. planet).

$$\mathrm{T}_p = \frac{a_p}{a} + 2 \cdot \sqrt{\frac{a}{a_p}(1 - e^2)} \cdot \cos i$$

Several comet families are known today (see Fig. 3.13): the Jupiter family, the Halley family, the recently discovered main belt comets and the sun grazers. The latter ones are further subdivided into the Kreutz group and the Meyer group.

The Jupiter Family
The Jupiter family is the by far largest group of short-period comets. The family consists of over 400 known comets with orbital periods of less than 20 years. Their

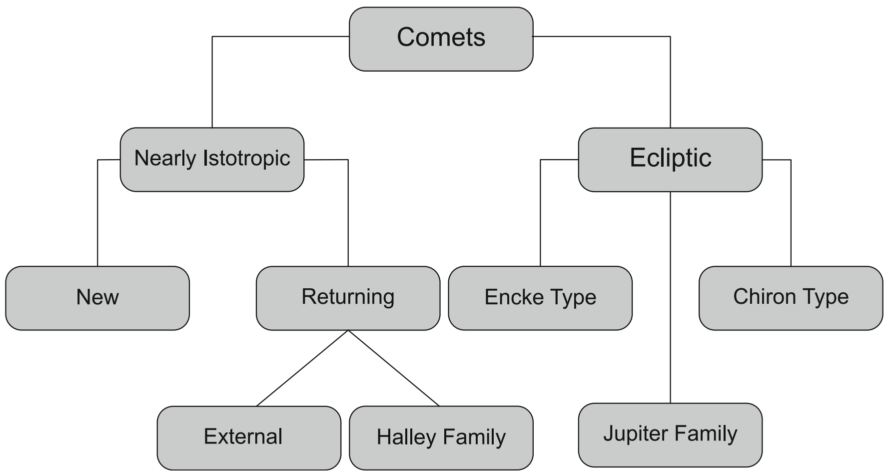


Fig. 3.13 Classification of comets based on their orbital parameters and applying Tisserand's parameter (based on Levison)

orbital inclination is usually less than 25° towards the ecliptic plane. They are named Jupiter-family comets since their current orbits are primarily determined by the gravitational influence of Jupiter.

Their orbits are contained within or do not extend much beyond the orbit of Jupiter having their aphelia close to Jupiter.

It was assumed that they originate from the Kuiper belt. However, nowadays, it appears that it is more likely that they originate from the Scattered disk which overlaps partially with the Kuiper belt but extends well beyond it. Yet, how are these distant and small objects influenced by Jupiter which is at best about 25 AU away. At such a huge distance Jupiter's gravitational influence appears to be rather negligible.

The orbits of Scattered Disk Objects (SDOs) or Kuiper belt objects (KBOs) can be perturbed by the gravitational influence of the ice giant Neptune and forced onto highly elliptical orbits. Some of the SDOs or KBOs have perihelia close to Neptune and can thus get into Neptune's influence sphere (Hill's sphere). In addition, this may also happen to SDOs that have more distant perihelia. The orbits of SDOs are relatively unstable and chaotic. It is thus likely that they may collide with each other and then be forced on new orbits which might bring them close to Neptune. In any case, the result is the same. As they approach Jupiter, the orbits of these small bodies may be perturbed further, resulting in a ellipse with lower eccentricities and a shorter orbital period. The orbits of Jupiter family comets bring them regularly close to the gas giant rendering their orbits unstable.

Famous Jupiter family comets are 9P/Tempel and 19P/Borrelly. Shoemaker-Levy 9 which crashed into Jupiter in 1994 may have also been a member of this family. Furthermore, 73P/Schwassmann-Wachmann, a currently dissolving comet, is also part of the Jupiter family (see Fig. 3.14).

The Halley Family

There are a few other periodic comets that have similar orbital characteristics as Halley. Their orbital periods are between 20 and 200 years, their inclinations may be up to 90°. These comets are classified as Halley-type comets, named after their most prominent representative. Their aphelia often lie between Saturn and Neptune.

Currently, slightly more than 60 Halley-type comets are known. However, they form a minority group within the short-period comets as they are clearly outnumbered by the Jupiter-family comets of which we known over 400.

Contrary to the Jupiter family comets which are supposed to originate from the Kuiper belt or the Scattered disk, the Halleys are believed to originally come from the spherical outer Oort cloud at the fringe of our solar system. A passing star or giant molecule clouds that pass close by may then send them on highly elliptical orbits towards the inner solar system as long-period comets. They may then be further influenced by Neptune or the other gas giants and forced on their Halley-type orbits.

Famous Halley family members are 1P/Halley, 12P/Pons-Brooks, and 13P/Olbers.

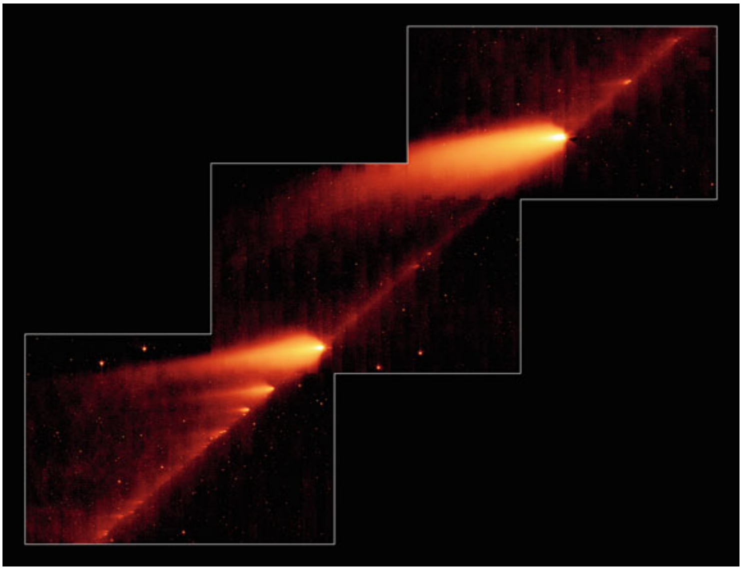


Fig. 3.14 Infrared image from NASA's Spitzer Space Telescope showing the broken comet 73P/Schwassman-Wachmann 3 skimming along a trail of debris left during its multiple trips around the sun (credit: NASA/JPL-Caltech)

Main Belt Comets

Main-belt comets (MBCs) are bodies orbiting within the asteroid belt that have shown comet-like activity during part of their orbit. They have only recently be discovered. The first main belt comet is (7968) Elst-Pizarro. It was discovered in 1979 and initially identified as an asteroid. Hence, it was given an official minor planet designation. In 1996, it was discovered that the asteroid had developed a tail, which is a clear feature of a comet. It was thus also given a cometary designation: 133P/Elst-Pizarro.

As the name suggest, they move on orbits that lie within the main asteroid belt between Mars and Jupiter. They orbit on near circular orbits and have low inclinations towards the ecliptic plane. Therefore, they do not show any major differences to other main belt asteroids.

It is not clear where the MBCs originated. It seems very unlikely that they are icy bodies coming from the outer solar system. No mechanisms are yet known that would bring such bodies into the currently observed orbits. Astronomers thus believe that they actually formed in the main belt close ego their current locations.

The reasons for their activity are also not known yet. Some believe that they are most likely not comets with sublimating ice but asteroids that exhibit dust activity. They should thus be relabeled as active asteroids. Several proposals regarding the cause of activity are currently discussed without any conclusion or preference so far. Firstly, the activity could due to sublimation of water ice. Secondly, some bodies could be rotating rapidly which may result in some kind of instability and could thereby loose mass from their surfaces which is then ejected into space. Thirdly, the activity may be due to microimpacts creating small impact craters from which dust and other material could be ejected into space. Fourthly, some of these bodies may get so close to the Sun that their surfaces heat up considerably. At such potentially high temperatures, the rocky material of these bodies may suffer cracks out of which dust can be ejected into space. Fifthly, radiation pressure competes with gravity to sweep unattached particles from the surfaces of asteroids. Near the sun, radiation pressure can eject dust very readily from kilometer-sized bodies and it can work in tandem with the other processes to leak material from the asteroids.

In October 2013, the breaking up of the main belt comet P/2013 R3 could be observed (see Fig. 3.15).

Sungrazers

Comets that come very close to the Sun during their respective perihelia are called Sungrazers. Sometimes the distance to the Sun's surface may only be as less as a few thousand kilometers. The smaller the comet is the higher is the probability of the comet being destroyed during its perihelion passage.

Based on similar orbital characteristics they can be put into subgroups. The subgroup having the comets that pass closet to the Sun is the so-called Kreutz group. This group named after the German astronomer Hienrich Kreutz (1854–1907) is also the largest subgroup. The Kreutz sungrazers all originated from one giant comet that broke up into many smaller pieces. Most of them have sizes of only 20–30 m. Due to their faintness and smallness it is almost impossible to observe them from Earth.

Only thanks to the Solar and Heliospheric Observatory (SOHO), a cooperation project of ESA and NASA, and launched in December 1995, most of these sungrazers became observable (see Fig. 3.16). Since its launch, SOHO has discovered more than 2400 sun grazers. About 85 % of them belong to the Kreutz group. The remaining 15 % distribute among the Kracht, Marsden and Meyer groups. The Kracht and Marsden groups are somehow related to the comet 96P/Machholz. The members of the Meyer group are typically very small, fairy and never have tails.

3.3.6 A Comet's Fate

In this chapter, we have discussed a lot about comets and their life cycles. Where do they originate? What happens during their trips into the solar system? Yet, we are missing one important aspect: their end! How do comets end? What is their fate?

In general we can distinguish three different end scenarios: ejection from the solar system; extinction; and Breakup or collision with another solar system body.

Ejection from the Solar System

A comet may be ejected from the solar system by gravitational interactions with another body of the solar system. Only the Sun and the gas giants, foremost Jupiter, are massive enough to kick out a comet into interstellar space. For this, the comet

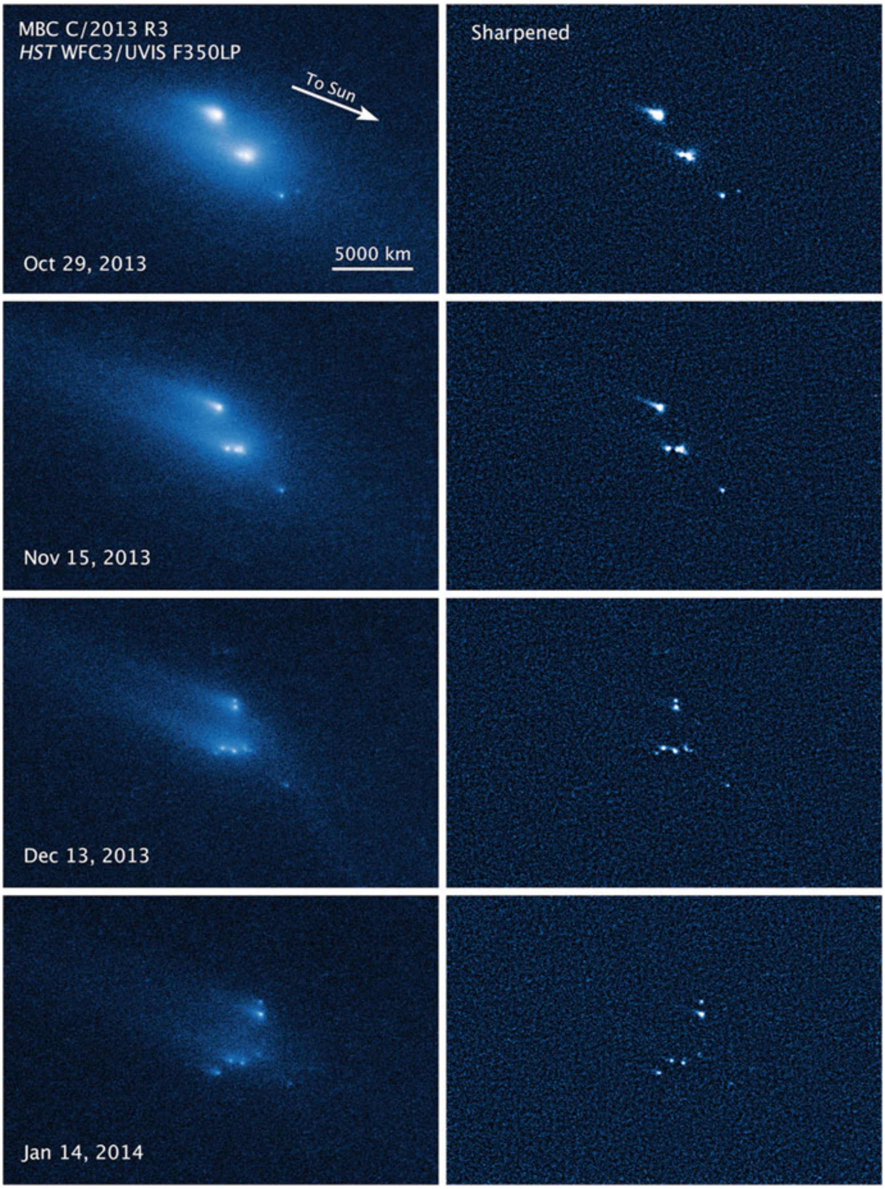


Fig. 3.15 This series of Hubble Space Telescope images reveals the breakup of an asteroid, a potential main belt comet (MBC) over a period of several months starting in late 2013 (credit: NASA, ESA, D. Jewitt (UCLA))

needs to be accelerated enough to reach the escape velocity necessary to leave the gravitational influence of the Sun behind. Furthermore, for this its orbit must be hyperbolic. An example of this is thought to be Comet C/1980 E1, which was

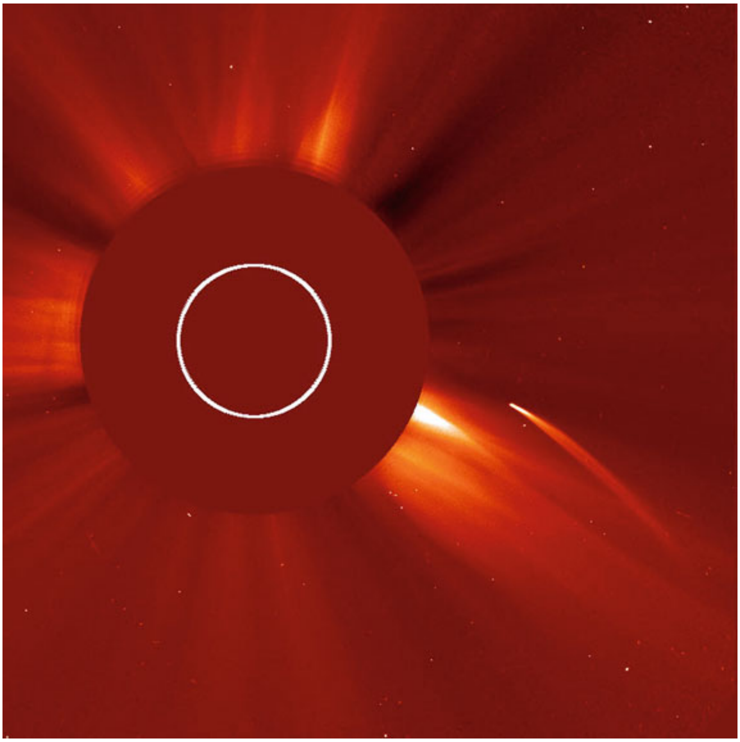


Fig. 3.16 A sungrazing comet as seen by the spacecraft SOHO as it dived toward the sun on July 5 and July 6, 2011 (credit: SOHO (ESA & NASA))

shifted from a predicted orbit of 7.1 million years around the Sun, to a hyperbolic trajectory, after a 1980 encounter with the planet Jupiter.

Extinction of Comets

This is a more common fate than being expelled from the solar system. During each passage into the solar system, a comet awakes and becomes active. While being active huge amounts of volatiles outgas. This, however, means that the comet loses material and will thus be exhausted after a certain amount of perihelion passages. The number of passages a comet can make s very variable and it seems to be different for long-period and short-period comets. Whereas long-period comets often have run out of volatiles after small number of passages (e.g. 50), a short-period comet may survive 1000 trips into the solar systems. The reasons for this have not been fully understood yet.

In any case, when the volatiles are exhausted the only thing that remains is rock or rubble. Comets in such a phase are no longer active and resemble more an asteroid rather than a comet. However, comets often do not go from active state to extinct state directly. In an intermediate state, the dormant state, they may just appear to be extinct but could wake up someday. How does this work?

The surface of a dormant comet is usually covered by a thick crust that may be several centimeters thick. The crust seals the volatiles below the surface and hinders them from outgassing. The comet appears to be extinct. Yet, 1 day, the crust may break up, the volatiles escape and the comet returns to active state. The comet (14827) Hypnos may be an example of a dormant comet. However, a comet's state is final if all the volatiles are exhausted.

Breakup and Collisions

This is probably the most interesting and fascinating end of life scenario for a comet. Comets can either breakup into several fragments or collide with another body in the solar system. The breakup can be either slowly, i.e. the comet is disintegrating over a longer period of time, or it can be sudden. It is particularly interesting to observe the fragmentation or breakup of short-period comets as this allows changes to be tracked over several perihelion passages. Smaller fragments often last only a few weeks, whereas larger ones can survive for years.

If we have closer look at the comet 73P/Schwassmann-Wachmann and its slow disintegration (see Fig. 3.17), we can see that in 1995, the comet broke up into four fragments. Then, during its return in 2001, only two out of the four fragments were still visible. In 2006, the comet further broke up into a total of 67 pieces. During its previous return in 2011, only the main fragment was still visible. All the otter 66 pieces had been disappeared.

We should differentiate on where on a comet's orbit such a breakup occurs. If it is close to Jupiter or the Sun, the explanation can often be easily found. These massive objects exhibit strong tidal forces on the comet. If the comet is not able to sustain this, it will break up. A spectacular breakup was that of the comet D/1993 F2 Shoemaker-Levy 9 which broke up in 1992 after a close encounter with Jupiter and

Fig. 3.17 Comet Halley during its perihelion in 1986 (credit: NASA/W. Liller)

then subsequently crashed into Jupiter in 1994. We will come back on this event later this chapter.

If the breakup occurs somewhere else on the comet's orbit, it is more difficult to find an explanation. The comets rotation and the exhibited centrifugal forces may cause a breakup of the comet or some parts of its nucleus may be lost into space. Also thermal stress could be responsible. A comet usually undergoes high differences in temperature during the course of its orbit ranging from hot (at perihelion) to very cold at its aphelion. This could, depending on the comet's structure and composition, cause thermal fractures which then in turn may be the predetermined breaking points. Collisions with other solar system bodies are of course also conceivable for breakups.

Some comets meet a more spectacular end—either falling into the Sun or smashing into a planet or other body. Collisions between comets and planets or moons were common in the early Solar System: some of the many craters on the Moon, for example, may have been caused by comets.

3.3.7 Naming of Comets

Now, that we have seen and learnt a lot about comets, it is time to pass on to some "administrative" staff. You probably have already wondered what the cryptic prefix to the comet's name means such as e.g. C/1995 Hale-Bopp.

This is part of the formal designation of comets. In 1994, the IAU agreed on a fixed naming convention for comets in order to properly identify them. This convention is similar to that of asteroids.

The naming is a two-step procedure. In the first step, a provisional name is given to the comet. During the second step, the comet is named after its discoverer. In the first step, the comet is designated by the year of its discovery followed by a letter indicating the half-month of discovery. A further number is added to indicate the order of the discovery within the respective half-month.

The letter of the half-month is easy to derive: The letter "A" is assigned to the first half of January, "B" to the second half and so on until the last half of December is designated "Y". The letter "I" is omitted as it could too easily be mixed up with the number "1".

Let us have an example. If we want to designate a comet which was the seventh to be discovered in the second half of March 2014, we will arrive at 2014 F7.

A number of prefixes are used to further characterize the type of the comet.

P/	Periodic comet
C/	Non-periodic comet
X/	Comet for which no reliable orbit could be calculated
D/	Periodic comet that has disappeared, broken up or been lost
A/	Mistake in identification as comet; actually an asteroid

There is a further refinement with short-period comets. After their second observed perihelion, they are assigned a number indicating the order of discovery. Halley's comet, e.g., is designated as 1P/1682 Q1 Halley as it was the first periodic comet that was discovered.

3.4 Prominent Comets

After so many details about comets in this chapter, let us now turn to some famous and remarkable comets.

3.4.1 The Famous Halley's Comet

Halley's comet is famous with regard to several respects. It is the first comet that had been recognized as a periodic comet. It is, at least until now, also the only short-period comet that can be clearly seen by the naked eye from Earth. Furthermore, it is the only naked eye comet so far that can potentially be seen twice in the lifetime of a human as it has an orbital period of 75–76 years.

Historical Observations

Although it had not been recognized as the same object, individual apparitions of Halley's comet have been observed for a very long time. How do we know this? Edmond Halley was able to determine the comet's orbit. He used this information to predict its next apparition. However, we can also use it for the opposite purpose and go back in time rather than into the future. If we take into account the possible perturbations of the comet's orbit due to the influence of the gas giants, we can derive when the comet should have visited the inner solar system in the past.

We can thus go back into the past and look for records of the comet. The earliest observations may go back at least until 467 BC. However, the records are not clear in that respect whether it was really Halley's comet that was seen. The ancient Greek described a comet between 486 and 466 BC. The timing, location, duration and associated meteor shower may indicate that it was Halley's comet. However, as you can see from the wide time frame of 20 years in which the records are varying, it could also have been a different comet.

The first certain observation of this comet dates back to 240 BC and can be found in the Chinese chronicle "Records of the Grand Historian" or simply Shiji. The descriptions therein including timing and appearance clearly resemble Halley's comet.

Babylonian records written on tablets date back to an apparition in 164 BC. The fragments of these written records can be found in the British Museum in London. Another Babylonian tablet describes a bright, hairy object in the sky that could be seen "day beyond day" for a month.

The next documented apparition dates back to 12 BC and is written down in the Book of Han. This was written by Chinese astronomers of the Han dynasty (206 BC–220 AD). The authors had been able to track the comet from August to October that year. Since this apparition is so close to the alleged birth of Jesus Christ, some scientists believe that Halley's comet may be considered to be the famous Star of Bethlehem. However, other explanations for this biblical phenomenon are also discussed such as a specific planetary conduction or simply another bright comet.

Further observations were made during Halley's apparitions in the years 66, 141, 347, 607 and 837 AD. None of them described something special except for the last one in the year 837. Halley must have been very bright during that apparition as it was only 0.03 AU away from Earth at his closest approach. Never again the comet has been so close to Earth again. The records describe a tail that stretched about 60° across the sky. The full moon for example merely covers half a degree. Thus, its tail was about 120 times larger than the full moon by then. The comet was observed in China, Japan, Germany and the greater Middle East. Most likely, it was also seen from other parts of the world. However, we lack written evidence for this.

A very famous apparition occurred in 1066, which have already touched briefly, and should further manifest the belief that comets are omens being either bad or good. During that time, Duke William II of Normandy, a Norman, tried to seize control of England. The English King, Harold II, an Anglo-Saxon, fought him. The year 1066 was a turning point in history. Harold died in the Battle of Hastings and William became the new king of England and was subsequently called William the Conqueror. That year, the year of the Battle of Hastings, Halley appeared again in the sky. As the by then people did not understand celestial mechanics, the comet appeared out of a sudden. Perhaps it was even send by God in their views. Hence, the apparition was later on interpreted as a bad omen for Harold and a good one for William. The comet is represented on the Bayeux Tapestry where it appears just above the battle grounds of Hastings (see Fig. 3.1).

Halley's comet and its alleged relation to the battle is also described in the Anglo-Saxon Chronicle, a collection of annals written in Old English about the history of the Anglo-Saxons in England. The first manuscript of it dates back to the ninth century. It has then been updated continuously until 1154.

In 1301, the comet was observed by the artist Giotto di Bondone (1266–1337). He was inspired by the appearance to assume that at least something similar had been the Star of Bethlehem which he put on one of his paintings.

When the Ottoman Empire invaded the Kingdom of Hungary in 1456 which led to the siege of Belgrade in July 1456, Pope Colixtus III ordered special prayers for the city's protection. A comet appeared in the sky.

The apparition in 1910 was special again. It was the first time that the comet could be banned on photographs during its journey through the inner solar system. This time, the comet came relative close again to Earth at a distance of only 0.15 AU. It hence gave a spectacular show in the sky. However, not all people were able to enjoy it. The comet's appearance was paired with a great panicking in the public. It had been found out by analysis of the orbit that Earth would pass through the comet's tail on May 19, 1910. Just before, astronomers had revealed that the toxic cyanogen was present in the comet's tail. The people were afraid, heated up by intensive media coverage, which lives would come to an end when Earth passed through the cometary tail. The sales of gas masks and quack "anti-comet pills" and "anti-comet umbrellas" and other weird stuff went up drastically. Astronomers worldwide tried to calm down the situation and explained that he density in the tail would be very low and the abundances therein very diffuse. At no point of time, any danger for life on the Earth existed. However, this did not work out. The panicking only stopped after the passage and when the people realized that nothing had happened.

Halley's last journey into the inner solar system culminated in its perihelion in 1986. This time Halley's appearance was not as impressive as during many of the earlier passages. During its perihelion Halley and Earth were on different sides of the Sun and the closest approach to Earth was merely at distance of 0.42 AU—by far more than during the last centuries.

However, not only the unfavorable distance to Earth was problematic. In 1986, it was the first visit of Halley in the inner solar system where light pollution on Earth played an important role and concealed most of the impressive experiences people in the centuries before had.

Yet, the situation was not as worse as you may think. Science and space technology had made a great progress and for the first time ever a whole armada of spacecraft was at ready for observing a comet, Halley's comet, with a close distance from space. We will deal with the Halley Armada in Chap. 6 in more detail. The five involved spacecraft, ESA's Giotto, the two Soviet Vega probes and the two Japanese probes Suisei and Sakigake, provided an unique view on the nucleus of Halley and delivered measurements and data that kept scientists busy for quite some time.

After its perihelion, Halley started its departure from the solar system. However, it should not be able to do that in secret as during all the centuries before. Its journey was further monitored.

On February 12, 1991, astronomers were able to observe several outbursts that lasted for several months (see Fig. 3.18). The outbursts started most probably in December 1990. By that time, Halley was at about 14.4 AU from the Sun roughly between the orbits of Saturn and Uranus. A dust cloud of more than 300,000 km in size was caused by about 1.4×1010 kg of dust ejected. The comet brightened from 24.3 to 18.9 mag, i.e., about 150 times brighter. The reasons for the outburst are not well understood. At the large distance from the Sun where the outbursts occurred, the sunlight is very faint and the surface temperatures on the nucleus' surface are only at about -200 °C. This is too cold for most sublimation processes to take place. At least for those that could explain such a dramatic outburst, e.g. the sublimation of water ice. Four possible explanations appear to be plausible. The outburst may be due to an impact with another small solar system body. The large amount of energy required was stored somewhere in the internal structure of the comet and then suddenly released. A different form of sublimation may have been active: the sublimation of carbon monoxide (CO). The last option is the interaction

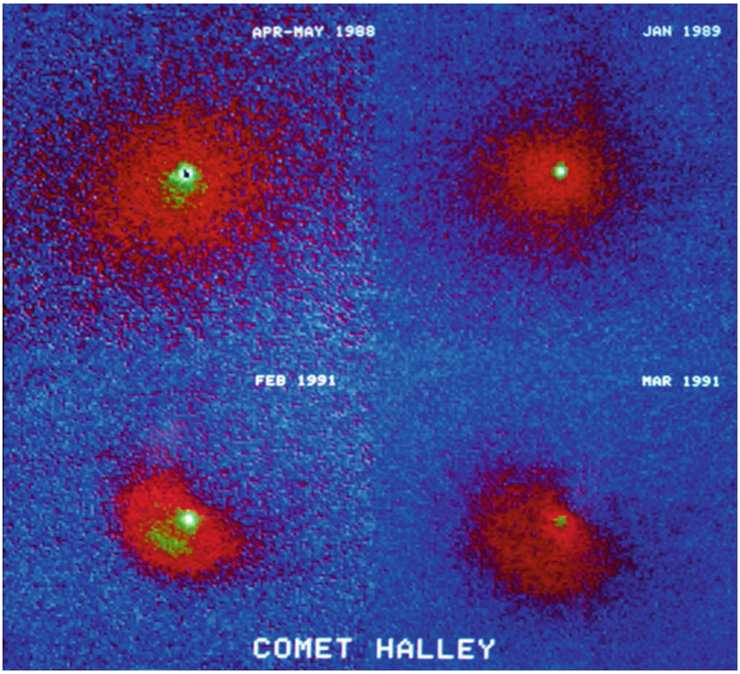


Fig. 3.18 A collage showing the evolution of Comet 1P/Halley as it moved away from the Sun. A major outburst took place in February 1991, just before the *third image*. The *last image* shows the dust released during the outburst started to disperse (credit: ESO)

with highly energetic particles in the solar wind. Options 1 and 3, the impact scenario and the sublimation of CO, are among the favored ones.

Observations continued. The Very Large Telescope (VLT) of the European Southern Observatory (ESO) located in Paranal/Chile (see Fig. 3.19) obtained the most recent images in 2003. The images were taken when the comet was on its farthest point from the Sun on its orbit (aphelion). You can imagine how difficult it is to identify Halley on this image. The successful observation means that Halley's comet is now observable and can be tracked throughout its whole orbit. This allows for more detailed studies about when a comet awakes and becomes active and how this process takes place.

Determination of Halley's Orbit

Halley's comet was the first comet that was actually recognized as periodic. Before apparitions of comets had been considered as single events. They simple come and go. This notion changed with the works of Isaac Newton and Edmond Halley. Newton laid the cornerstone for most of modern physics, in particular with his laws on gravity and motion.

In 1705, Edmond Halley applied these laws to calculate the gravitational effects of Jupiter and Saturn on comets and their orbits.

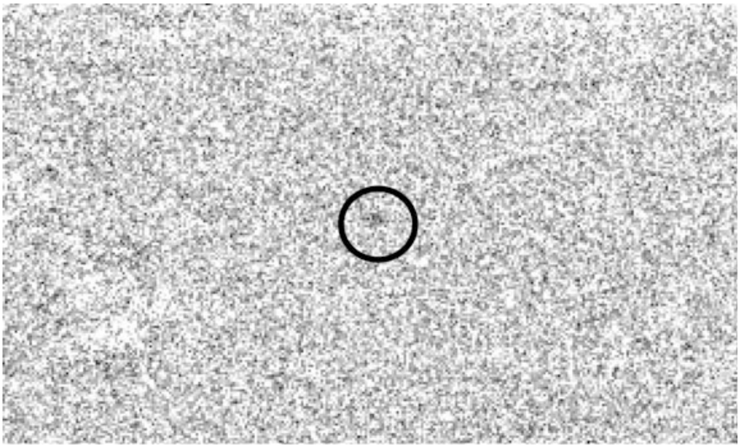


Fig. 3.19 ESO's Very Large Telescope in Paranal/Chile was able to detect Halley's comet close to its aphelion in 2003 (credit: ESO)

He then examined historical records of comet observations and found that the orbital characteristics of a comet observed in 1682 and the two comets that appeared in 1531 and 1607 in the sky respectively were very similar. Taking perturbations of their orbits by the gas giants into account (by that time only Jupiter and Saturn had been known), he concluded that these comets actually were the same object that had reappeared again. He then determined the orbital period of the comet to be 76 years. Nowadays, we consider this period to vary between 75 and 76 years. He then predicted the comets return for 1758.

This prognosis turned out to be correct, although the comet was not discovered again in the sky until December 25, 1758. The German farmer and amateur astronomer Johann Georg Palitzsch (1723–1788) was the lucky one who helped to confirm Halley's assumption by his discovery.

During that apparition Halley's comet passed its perihelion on March 13, 1759. Jupiter, Saturn and potentially Uranus had caused a slight retardation.

With his prognosis and approach, Edmond Halley laid the foundation of cometary science. "His" comet was named in order to honor him "Halley's Comet" and has ever since then fascinated people with its appearances.

Orbit and Origin

Halley's orbital period over the last three centuries has varied between 75 and 76 years. If we look back further in time, we can see that in earlier days since its first recorded observation in 240 BC, we can see that there has been a slightly larger variance in the range of 74 and 79 years.

Halley's orbit is highly eccentrically ($e = 0.967$) and about 18° inclined towards the ecliptic plane. The orbit takes Halley as close as about 0.6 AU to the Sun (perihelion) and roughly to the orbit of Pluto at 35 AU at its aphelion. The comet therefore passes again and again the realm of the gas giants and its orbit suffers slightly from continuous perturbations exhibited on it by the gas giant's gravitational influences.

The orbit is in one further respect peculiar. Halley's motion on its orbit is retrograde, i.e. in opposite direction as the planets and most of the other bodies in our solar system.

Though being classified as a short-period comet as its orbital period is less than 200 years, Halley is an atypical representative of this class. Its orbit is much higher inclined and the orbital period clearly longer than for most other short-period comets.

All this suggests that Halley's comet has not always been a short-period comet but does have a different origin. It is more likely that it initially used to be a long-period comet ejected from the Oort cloud. The gas giants, in particular Jupiter, would then have subsequently perturbed Halley's orbit while it travelled towards the inner solar system.

There are a few other periodic comets that have similar orbital characteristics as Halley. Their orbital periods are between 20 and 200 years, their inclinations may be up to 90°. These comets are classified as Halley-type comets, named after their most prominent representative.

Currently, slightly more than 60 Halley-type comets are known. However, they form a minority group within the short-period comets as they are clearly outnumbered by the Jupiter-family comets of which we known over 400. The latter ones, in contrast to the Halley-type comets, are considered to have originated from the Kuiper belt or the Scattered Disk.

However, Halley-type comets may also have a different origin other than the Oort cloud. In 2008, another trans-Neptunian object (TNO) was discovered: 2008 KV_{42}. It also has a retrograde and highly inclined orbit (104° towards the ecliptic plane) similar to Halley. Its perihelion lies just outside the orbit of Uranus (20.3 AU) and extends to roughly twice the orbit of Pluto (70.6 AU). It is the first TNO with a retrograde orbit that has been discovered. Its unusual orbit suggests that 2008 KV42 may have been perturbed inwards from its original source that might be the inner Oort cloud. This must have happened by a yet unknown gravitational disturbance. It may constitute a new population of small solar system bodies, which may be the origin of Halley-type comets. Some scientists believe that 2008 KV_{42} may simply be in an intermediate stage of becoming a comet.

Structure and Composition

Not much had been known about the structure and composition of comets before 1986. Part of the problem was that a comet's nucleus is usually concealed from our view by the coma and tail. As we have also seen, the chemical compounds in the coma (daughter molecules) are different from the original compounds (mother molecules) ejected from the nucleus due to chemical reactions.

Halley's last perihelion in 1986 gave a first, unique chance of getting more insight by the help of the Halley armada. In particular ESA's Giotto and the Soviet Vega probes delivered highly relevant results. So, what do we know about Halley's comet?

Is peanut-shaped nucleus has the approximate dimensions of 15 × 8 × 8 km and has a relatively low mass of 2.2 × 1014 kg and a very low density of 0.6 g/cm^3. In contrast, our Earth has a density of 5.5 g/cm^3 and even Pluto is even denser (2.03 g/cm^3). It is also one of the darkest bodies in the solar system we know having an albedo of only 0.04 and a slightly reddish color.

Data derived from Giotto suggest that only about 10 % of Halley's surface had been active during the observation.

3.4.2 2P/Encke

Comet 2P/Encke, sometimes also called Encke's Comet, is, as its name already indicates, the second short-period comet that had been discovered to be a short-period one. The naming deviates from the usual applicable scheme in that it is not named after its discoverer but bears the name of the person who determined its orbit, the German astronomer Johann Franz Encke (1791–1865). It follows thus the tradition of Halley's comet.

Found, Lost and Found Again

Comet Encke had been observed several times but was not recognized to be the same object. The first record observation dates back to January 17, 1786. The French astronomer Pierre Méchain (1744–1804) was searching for comets that night and found a relative bright fuzzy, yet unknown object in the sky. However, the object showed no signs of a tail and Méchain was only able to track the object for a couple of days. He was thus not able to determine an orbit. It is worth mentioning that Méchain is famous for his contributions to the early observation of deep sky objects and comets in cooperation with Charles Messier (1730–1817).

It was rediscovered on November 7, 1795 by Caroline Herschel (1750–1848), a German astronomer and sister of William Herschel. This time the comet was observable for 23 days. However, it was again not possible to determine its orbit. Most probably owing the fact that comet orbits by that time were considered to be parabolic and this did not match with Encke at all, as we will see in a moment. So, the comet was lost again.

On October 20, 1805, the French astronomer Jean-Louis Pons (1761–1831) found it again in the sky not knowing, like his predecessors, that the object he had found was an old friend. He was able to track the comet for about 1 month. Yet again, nobody could determine an orbit.

Johann Encke deviated from the common approach and assumed an elliptical and determined an orbital period of slightly above 12 years. At first glance, this appeared to be a success. However, the result should turn out to be incorrect. As we know today, the comet remained undiscovered during its following three perihelia.

It was only on November 26, 1818 that Pons found rediscovered it once again and observed it for 48 days. Encke had not given up and tried another time to determine the comet's orbit. This time, he applied the new method of least squares developed by Carl Friedrich Gauss. He came up with an orbital period of 3.3 years which corresponds to our current determination.

Encke subsequently had a look at historical records. He had been aware that periodic comets such as Halley's comet existed. Was this comet of 1818 another one of that type? Encke found out that the orbital characteristics of the current comet were very similar to the comet apparitions go 1786, 1795, and 1805. In the end, Encke was able to proof the identity of these objects and forecasted the next apparition to be in May 1822. Indeed it reappeared in the skon on June 2, 1822 and was discovered by the German astronomer Karl Rümke (1788–1862) while he was working from Australia.

The comet was then subsequently named 2P/Encke to honor the man who had been able to determine its orbit and recognize its periodicity. It was the second known periodic comet. However, more should follow.

Orbit

The short-period comet Encke orbits on an elongated ellipse that brings it as close at 0.339 AU to the Sun and takes it as far as 4.097 AU from it. Hence, it orbits lies in between the orbits of Mercury and Jupiter. Its moderately inclined orbit ($i = 11.8°$) takes it within 3.3 years once around the Sun. During its journey it frequently passes the planets of the inner solar system and the comet's orbit is slightly perturbed by their gravitational influence.

This very short orbital period actually renders it the comet with the shortest known orbital period. Today, we know Encke's orbit very well and can observe it throughout its whole orbit.

3.4.3 Shoemaker-Levy 9: The Jupiter Crash

Comet Shoemaker-Levy 9 and its collision with Jupiter in 1994 is a very prominent example for a possible fate of comets. Its story is fascinating and the detailed observations that had been carried out gave unforeseen insight into cometary structures and also on Jupiter's atmosphere. The crash also illustrates the importance and key role of Jupiter as a "cosmic vacuum cleaner" for the inner solar system. As we have already seen in this chapter, Jupiter's enormous mass and the resultant gravitational influence lead to many small comets and asteroids colliding with the planet or being massively perturbed in the orbits and possibly kicked out of the inner solar system. It is thus worth to tell its story again in more detail.

The Discovery

During the night of March 24, 1993, the astronomers Eugene Shoemaker, his wife Carolyn and their colleague David Levy pointed the 460 mm Schmidt telescope of the Mount Palomar Observatory into the night sky. They were working on a program whose goal it was to discover near Earth objects. Neither of them could anticipate how that particular night should change their lives.

The Shoemakers as well as Levy had already discovered several comets, eight together. Yet, the one found on a photo of that night, turned out to be special in many respects. Its discovery, following the usual procedure, was announced in IAU Circular 5725 and could be confirmed by many observers worldwide. Following an old tradition it was named after its discoverers Shoemaker-Levy 9 (SL9).

An Unusual Comet

Already the discovery image gave a first hint that SL9 was special. It showed an oddly elongated region of diffuse light about 50 arc seconds long and 10 arc seconds wide. Several peaks in brightness were apparent suggesting the existence of multiple nuclei lined up like a rope of pearls (see Fig. 3.20). Was SL9 a comet that had already broken up?

The comet's orbital elements revealed another peculiarity. They suggested that SL9 was orbiting Jupiter instead of the Sun unlike all other comets known by that time. Further studies hinted that Jupiter captured the comet in the early 1970s and forced it into a highly elliptic orbit with an eccentricity very close to 1. Its orbit around Jupiter was very loosely bound, with a period of about 2 years and an apojove (the point in the orbit farthest from the planet; corresponding to aphelion of comets orbiting the Sun) of 0.33 AU (49,000,000 km). Before the gas giant captured

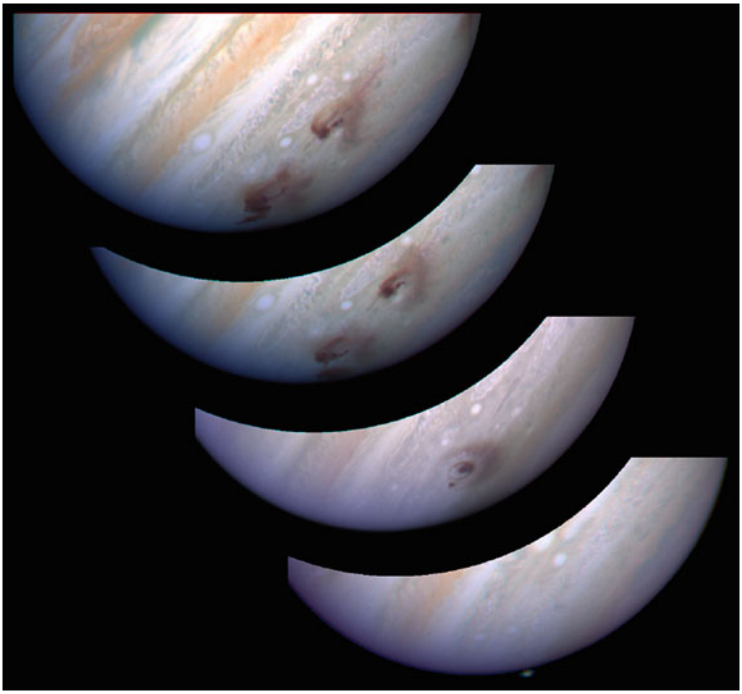


Fig. 3.20 Mosaic of the evolution of the impact of fragment G (credit: R. Evans, J. Trauger, H. Hammel and the HST Comet Science Team and NASA)

Fig. 3.21 The fragments of comet Shoemaker-Levy 9 on collision course with Jupiter. The image was taken on May 17, 1994 by the Hubble Space Telescope (credit: NASA, ESA, and H. Weaver and E. Smith (STScI))

SL9 it probably was a short-period comet with an aphelion just inside Jupiter's orbit and a perihelion somewhere interior to the main asteroid belt.

Apparently in 1992, SL9 passed very close to Jupiter with a distance of only about 40,000 km and well within the planet's Roche limit.

The Roche limit, named after the French astronomer and mathematician Eduard Albert Roche (1820–1883), is generally used in the assessment of the stability of a celestial body orbiting another one. A body A holds together due to gravitational forces. If A orbits another body tidal forces are exerted on it owing the fact that the two bodies A and B attract each other. However, the attractive force varies in its intensity on A: it is stronger on the side facing A and weaker on the far side. A remains in a stable state as long as the gravitational forces overweigh the tidal forces. As soon as the tidal forces become stronger as the gravitational ones, and exactly this happens when passing the Roche limit, A breaks up.

So, while passing Jupiter's Roche limit in July 1992, the comet's nucleus of assumed 4 km in diameter was torn apart into 21 single fragments having sizes between 50 m and 1 km (see Fig. 3.21). Each fragment of the comet was denoted by a letter of the alphabet, from "fragment A" through to "fragment W", a practice already established from previously observed broken-up comets.

Shortly after its discovery, Shoemaker-Levy 9's orbital elements were further refined and suggested something astonishing: the comet would pass within a distance of 45,000 km from the center of Jupiter. This buzzed the community of planetary scientists as this was well within Jupiter's radius (roughly 69,911 km). That meant SL9 was on collision course with Jupiter and would very likely hit the planet in July 1994 over a period of several days.

Collisions of celestial bodies are not rare in our solar system. Yet, never before scientists had been able to observe the collision of a larger comet and a gas giant with such a long lead time. Up to then this had not been a very active research topic. Scientists regularly focused on collisions of solid bodies. The reasons for this were quite simple. Such events were easier to observe and the results in form of impact craters were open to thorough analysis for a very long period of time.

The world was caught by surprise. Nobody exactly knew what would actually happen when a fragile cometary nucleus hits the dense atmosphere of a gas giant.

Soon this event got the attention of the mass media. The public looked forward excitedly to this event which was supposed to take place in the week of July 16–22, 1994.

Before the Impacts: The Observers Get Ready

The expectations grew with each single day the event approached. Unfortunately, further analysis of the orbital characteristics in July 1993 showed that, to the astronomers much regret, the impacts of the fragments could not be directly observed from Earth as all of them were supposed to occur on the far side of the planet. But, blessing in disguise, an armada of space probes was at the ready to have a direct view on the crash.

In particular, Galileo was in an excellent observation position. Galileo was launched in 1989 and was expected to arrive at Jupiter in 1995. The spacecraft was merely 1.6 AU away from the event.

Additionally, the probe Ulysses actually sent into space to observe the Sun was pointed at Jupiter and was supposed to glimpse at the impacts from a distance of 2.6 AU.

Even Voyager 2 was expected to be used for tracking radio emissions of the impacts. Just think about it, this probe had been in space since 1977 and ever since traveling on its lonely path out of our solar system. In 1994, it was already 44 AU away from Jupiter. Furthermore, the Hubble Space Telescope (HST) took aim at Jupiter.

What Will Happen?

What expected the scientific community to see from Earth? As they did not have a direct view on the events several phenomena were forecasted. First, reflections of the impacts in form of flashes should be visible at the moons of Jupiter. Secondly, after a few minutes the impacts should become visible from Earth after they rotated into the field of sight. Thirdly, the most optimistic prediction was that large, asymmetric ballistic fireballs would rise above the limb of Jupiter and into sunlight to be visible from Earth. Further effects such as seismic waves traveling across the planet, an increase in stratospheric haze on the planet due to dust from the impacts, and an increase in the mass of the Jovian ring system were suggested. However, given that observing such a collision was completely unprecedented, astronomers were cautious with their predictions of what the event might reveal.

The First Impact: It Begins

On July 16, 1994 at 20:13 UTC the waiting was over. The long awaited day had arrived: fragment A moving with an approximate velocity of 60 km/s hit the southern hemisphere of Jupiter. Galileo detected a fireball with a peak temperature of about 24,000 K which was clearly above the usual 130 K in Jupiter's atmosphere. Within 40 s the fireball cooled down to about 1500 K. The impact cloud, the plume, ascended to about 3000 km into the atmosphere.

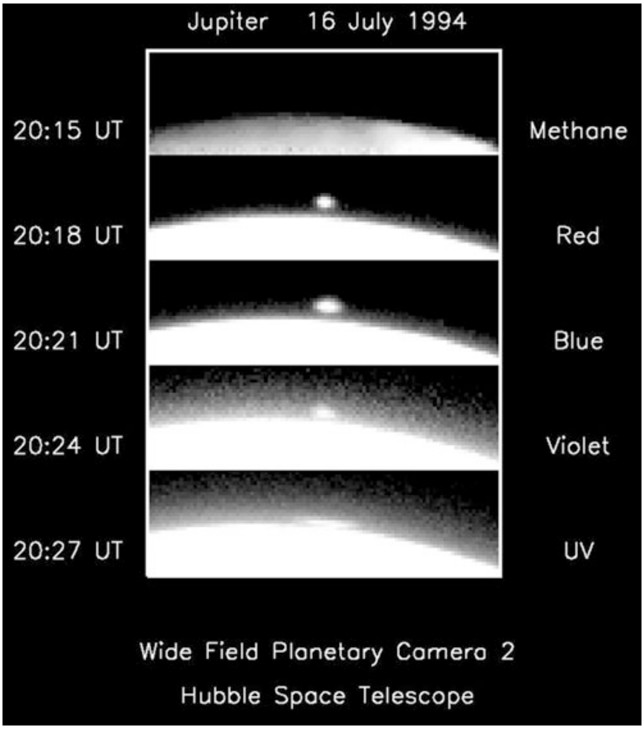


Fig. 3.22 Sequence of images showing a plume near the terminator of Jupiter at the time of the A impact (credit: HST Jupiter Imaging Science Team)

Thanks to the rapid rotation of Jupiter, the impact zone became visible to Earth-based observers and the HST after just a few minutes which still allowed a glimpse at the fresh impact. The images that had been taken sparked excitement (see Fig. 3.22). A few minutes after the impact fireball was detected, Galileo measured renewed heating, probably due to ejected material falling back onto the planet. Earth-based observers detected the fireball rising over the limb of the planet shortly after the initial impact. Furthermore, the dark spot that manifested at the impact zone appeared to be durable for quite some time. It had been one of the big worries before that these spots might fade away quickly.

From all around the world, observations were reported to the coordination center in Maryland/USA. The big event had begun and already now the expectations seemed to be excelled. How would the crashes of the other fragments develop? Fragment A was not the biggest one. Already fragment B, which was expected to hit Jupiter on Sunday 17.7. at 2:54 UTC, appeared much larger and brighter on photos than fragment A (Fig. 3.22).

The Series Continues

The observers excitedly awaited the second impact. This time, being nighttime in America and the Pacific region, ideal observation conditions existed for some of the world largest telescopes such as in La Sila (Chile), California or the Mauna Kea on Hawaii. Time passed but nothing happened. Even after the predicted impact time had long elapsed. “Nothing seen until 3:55 UTC”, the Palomar observatory in California reported.

What had happened? Hope, disappointment and excitement came together. Finally, the redemptive message arrived. At the Keck Observatory at the Mauna Kea on Hawaii, the 10 m telescope had been able to detect the impact. They had noticed the impact in a narrow L-band (3.27–3.44 μm). The plume had been weak but could be clearly detected at the predicted position. It began at 2:56 UTC and faded away until 3:13 UTC.

The astronomers were surprised. Why had this impact been so completely different from fragment A? Fragment B had looked so promising. It was the first big mystery of the impact week. Yet there was no time to take breath. The relentless staccato of the impact schedule left no room for that. Already at 7:02 UTC, fragment C was expected to plow into Jupiter’s atmosphere.

Indeed there was a confirmed sighting closely after the predicted time. First reports described two impacts. The first one was about 38 min too early. This caused a bit of confusion. Later one it became apparent that it was actually only one impact and the alleged second one merely had been a remainder of fragment A. The second impact was the real and expected one.

C’s plume rose above Jupiter’s limb at 7:20 UTC and showed a continuous increase in brightness over a period of several minutes. Only after about 10 min the brightness fell back to the level of impact A. It was the first time that changes in the spectrum of Jupiter’s atmosphere could be detected.

The media interest remained unbroken. Dozens of journalists gathered at the headquarters of the European Southern Observatory near Munich to gaze the first reports and images. Live coverages on several TV stations were held.

No time to pause. Next was fragment D’s turn. First reports arrived from Australian observers at 11:55 UTC. Astronomers were particularly excited about observations of the 600 mm-SPIREX telescope at the Amundsen-Scott station at the South Pole. Due to its geographic location it was able to provide uninterrupted observations of this event. Yet, shortly before the impact a heavy snowstorm covered the telescope. The astronomers first had to strip off all the snow at icy temperatures before they were finally able to see the effects of the impact.

The users of HST were in a much more comfortable position sitting in their air-conditioned offices at their computers. The HST took many photos, each photon being more impressive than the previous one. The viewers breadth caught.

While fragments E and F followed, the scientists waited excitedly for G. It had appeared on all photos as a very remarkable and bright object. Probably it had about five times the mass of fragment A, thereby releasing about 25 times of the latter ones impact energy. Since A had been so impressive, what about G then? Would it hold up the expectations or turn out to be a disappointment as B?

At 7:41 UTC, the first sighting was reported by SPIREX. The impact was clearly brighter and much more persistent than all previous ones. Hien Nguyen from the Amundsen-Scott station simply noted: “My God! It was extremely bright!”.

While the astronomers at the South Pole were cheering this time, the scientists at the Keck Observatory at the Mauna Kea were nervous. Dense fog hindered them from observing the event. Fortunately, just before the impact the fog vanished and allowed seeing the impact. The happiness, however, was only short-lived. They were able to take some photos of the impact and do some measurements before, shortly after the maximum brightness of the plume at 7:50 UTC, the fog returned. This time accompanied by heavy rain.

The amount of released energy was enormous. The 3.9 m mirror of the Anglo-Australian Observatory, e.g., had to be reduced to about 1 m in diameter. Otherwise it would not have been able to cope with all the infrared light coming from the impact.

On the evening of July 18, at 19:32 UTC fragment H collided with Jupiter. The next one was fragment K. If you wonder what happened to fragments I and J. Letter I was omitted while naming the fragments. Too easily it could have been mixed up with a cipher. The same, by the way, applied to O. Fragment J, though, existed at discovery time of SL9, but soon disappeared from later photographs. Most likely it disintegrated.

So K was next. SPIREX was once again the first to report that the impact had occurred at 10:33 UTC on 19.7. The maximum brightness was similar to that of fragment G. The astronomers had hoped that at least with fragment K, reflections of the impact at Jupiter's moons should be visible. During the previous impacts, the moons were either in direct sunlight or not visible from Earth. This time, Europa, one of the four Galilean Moons of Jupiter, was in an excellent position since it was visible from Earth but completely shadowed by the gas giant. However, once again they were disappointed. Nothing happened. The suspicion became substantiated that the fragments during the impact were first swallowed by Jupiter's atmosphere without giving any detectable traces. Apparently, the plume's brightness was not sufficient to cause any reflections.

There was still nothing to worry, the series of impacts continued. The impact of fragment L was detected at 22:15 UTC at 19.7. The amount of released energy was surprisingly high and the dark spot of the impact soon became a very easy target even for small telescopes. Fragment M, similar to J, had already been considered lost. The impact of fragment N that took place on 10:35 UTC at 20.7. was nothing special compared to the previous events. With each impact, the expectations grew. Fragment P, most likely broke up as images taken by the HST in May that year suggested. The remainders of it did not provide any visual traces during their impacts.

Now, the highlight of the week as predicted by several scientists was about to come: fragment Q. Something was jinxed. Its impact badly disappointed. Its plume was even less eye-catching as that of L. On 21.7. at 5:35 UTC, the Palomar Observatory reported the impact of R being of similar intensity as E. Everything seems to develop fine and the scientific community merely waited for the remaining impacts. Yet, a surprising message was issued by the Keck Observatory: an impact of fragment, that was presumed to be lost, did indeed take place in the early morning of July 20 (6:08 UTC). Impact S was reported at 15:29 UTC on 21.7.

Fragments T and U did not leave behind any detectable traces. Also with respect to V no positive reports were received. However, later on, the impact could be confirmed by an observation of the Lick Observatory.

Fragment W and with it the last remainder of SL9 bid farewell at 8:14 UTC on 22.7. Finally, it was time to take a deep breath. The crash was over and had the expectations more than fulfilled.

Scientific Results or What Did We Learn?

Finally, after the end of this special event, the huge amounts of data that had been collected needed to be analyzed. The big question was: what could we learn from the impacts?

However, the first aspect that attracted much attention was the fact that the fragments behaved so differently during their respective impacts. The best explanation for this seems to be the assumption that each fragment was composed differently. Because it was obvious: the size of the explosion was not directly correlated with the brightness of the fragment. Just think about fragments B and Q. However, this finding appeared to contradict conventional assumptions about the structure of comets.

Thus, the question arose whether SL9 had been a comet at all. Perhaps it was just another type of small solar system body such as an asteroid. This idea was soon quashed. The way the fragments were swallowed by Jupiter's atmosphere did not resemble the behavior one would expect with more solid bodies such as asteroids. Furthermore, the highly eccentric orbit of SL9 hinted to its cometary nature.

An interesting finding was that not all fragments had lied and moved in a direct line. There were some fragments that deviated from the main line, among them fragment B. This was a bit strange since they should be aligned after the break up of the comet. A possible explanation could be that these fragments only broke away thereafter. Their nuclei would then consist primarily of loosely bound parts, which then would simply fulminate, in Jupiter's dense atmosphere without major visible traces. Until now, this seems to be the best explanations for the sightings.

In addition, detailed spectral analysis was carried out for the impact zones by which numerous molecules were detected, among them carbon monoxide, ammonia, and hydrogen sulfide. Furthermore, large amounts of sulfur and carbon disulfide were detected which completely mystified the scientists in the beginning. The amounts were much larger than what had been predicted. Later on, it was found out that the surplus originated from Jupiter's lower atmospheric layers.

The presence of emission lines for iron, magnesium and silicon that were also found in the spectra showed that the impact energy was sufficient to vaporize metals.

As had been predicted beforehand, the collisions generated enormous waves which moved across the planet at speeds of 450 m/s and were observed for over 2 h after the largest impacts. The waves were thought to be traveling within the stratosphere.

The dark spots at the impact zones were visible for several months still and could easily be detected even in small telescopes.

Water on Jupiter?

Following recent results of T. Cavalié and his research team, SL9 is probably responsible for about 95 % of the water in Jupiter's stratosphere. Why is that? The water cannot come from deeper regions of Jupiter's atmosphere, such as the troposphere. The water steam present there is separated from the stratosphere by cold trap. When ascending into this cold trap, the water steam would freeze and would subsequently fall down again in form of ice crystals.

In the meantime, the processes that were going on during the first minutes after the impact can be well explained. First, the fragment dived into the atmosphere and remained almost intact. Only at its surface it disintegrated a bit. The fragment plunged into the atmosphere as deep as several hundred kilometers passing various cloud layers. Then, it exploded and triggered a fireball that rapidly moved upwards again and cooled down again after a few minutes. While cooling down it collapsed into a pancake-liked structure. The remainders then fell down again and heated up the stratosphere once again. This caused another flash to become visible. At the same time the characteristic dark spot developed. Chemical reactions were responsible for that.

An Isolated Case?

Was SL9 merely an isolated case? Was it merely Some strange body exceptionally crashing into Jupiter? We have seen that Jupiter with its enormous mass is capable of capturing smaller bodies that come close to it. Their respective orbits are elliptical and unstable due to Jupiter's influence. Statistical estimates, however, show that impacts do not occur too often. Objects of about 0.3 km in size seem to collide every 500 years and those with a diameter of 1.6 km only every 6000 years. Yet, these are only statistics. It is very difficult, if not impossible, to get reliable figures as the amount of observations that have been made are not sufficient.

For smaller objects, collisions seem to occur more frequently. Amateur astronomers have been able to detect various impacts over the last few years. The Australian Anthony Wesley was the first. He had been observing Jupiter with his 360 mm telescope for some time. Almost exactly 15 years after SL9, on 19.7.2009, he discovered an impact on Jupiter in form of a dark spot close to the South Pole. He quickly spread the new on the Internet and several people confirmed his discovery (Fig. 3.23).

The spot he saw measured approximately 5000 km in diameter. Its oval shape suggested that the impact trajectory had been inclined. The impactor, whose nature was still unknown, hit Jupiter on the far side between 9 and 11 UTC that day. Glenn Orton of the Jet Propulsion Laboratory (JPL) and his team believed to have identified the nature of the impactor after examining images taken by the HST and of three infrared telescopes. This time, most probably, a small asteroid, 200–500 m in diameter, hit Jupiter instead of a comet.

Soon after, Wesley and Christopher Go discovered independent from each other a further impact on June 3, 2010 which became noticeable in form of a flash of light. The determined light curves suggest a bolide of 8–13 m in size.

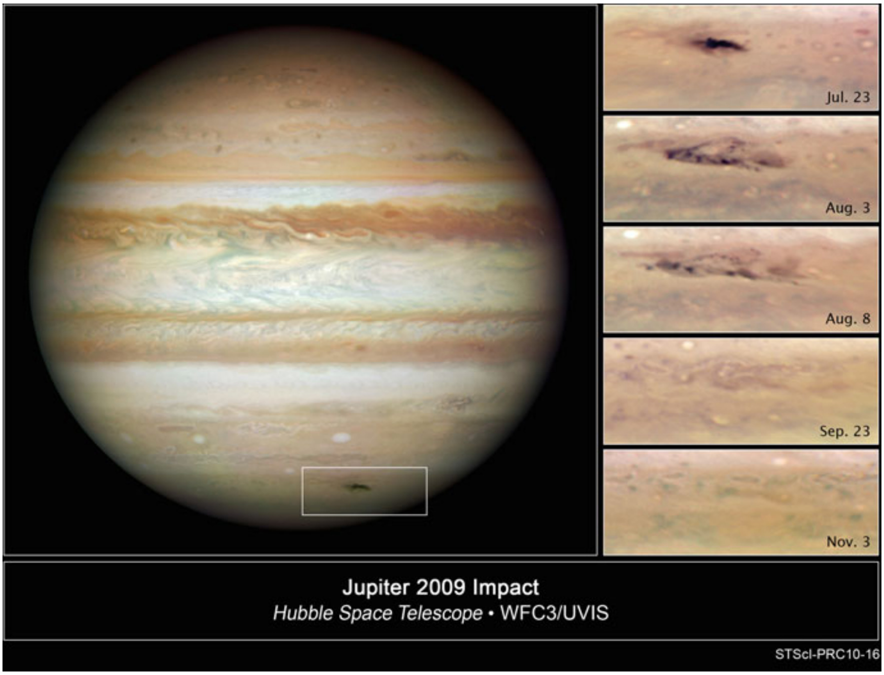


Fig. 3.23 In 2009, another impact occurred on Jupiter. An amateur astronomer first spotted the scar caused by the impact (credit: NASA, ESA, M. H. Wong (University of California, Berkeley), H. B. Hammel (Space Science Institute, Boulder, CO), I. de Pater (University of California, Berkeley), and the Jupiter Impact Team)

In August of the same year, three Japanese amateur astronomers M. Tachikawa, K. Aoki and M. Ichimaru discovered another one. The most recent impact was observed by the two Americans D. Petersen (visually) and G. Hall (photographically) on 10.9.2012.

3.4.4 C/1996 B2 Hyakutake

During the second half of the 1990s, a very rare event occurred: two extremely bright comets that were visible with the naked eye appeared on the starry sky: comets Hale-Bopp and Hyakutake. Both competed for the title “great comet”, but only one could win. Hale-Bopp, with which we will deal in the following section, clearly outplayed Hyakutake which nevertheless remained a remarkable appearance and delivered stunning new scientific results. Just to mention a few: It was the first comet ever where X-ray emissions could be detected and it had a record-breaking long gas tail. We will deal with this comet, which was eventually awarded the title “great comet of 1996”, in the following.

Discovery of Hyakutake

On the night of January 31, 1996, the Japanese amateur astronomer Yuji Hyakutake was observing the night sky on his quest to find new comets. He had been a dedicated comet observer for a very long time and had also hoped to find a new comet one night. Hyakutake even moved to the Japanese prefecture of Kagoshima where he had better observing conditions. That night, he used once again his large binoculars having 150 mm objective lenses. He preferred them to telescopes.

Already some weeks ago he had finally succeeded and discovered a comet, C/1995 Y1. Though this one never made it to a bright comet being visible to the unaided eye. That particular night at the end of January, Yuji was simply checking again the region where he found his comet. To his great surprise, there was another fuzzy object. After thoroughly consulting his star charts and lists of known comets, he was sure that this yet again was another comet. Shortly after this, he notified the National Astronomical Observatory of Japan about his discovery. Soon after, several astronomers confirmed his observations. Once again, he was able to immortalize himself in the list of comet discoverers. What he didn't know by the time of discovery was how magnificent his second comet should turn out to be.

At the time of discovery the comet had an apparent brightness of 11 magnitudes and was already pretty close to the Sun having only a distance of about 2 AU to our central star. Its coma already stretched over 2.5 arc minutes. Astonishingly, such a close and relatively bright comet had not been discovered earlier.

Orbital Characteristics

The first orbital elements obtained indicated that C/1996 B2 would pass Earth quite closely with a distance of about 0.1 AU (15,000,000 km) on March 26, 1996. Only four comets in the previous century have passed Earth closer. Hyakutake was the brightest Earth-approacher since the early eighteenth century, and in addition the 55 days between discovery and Earth approach is a record for a pre-perihelic Earth approach. As it also appeared to be a quite active comet during the first observations, it really had to the potential to become a great comet.

Hyakutake's orbit was determined to be an extremely elongated ellipse with an eccentricity very close to one and an inclination of about 125° towards the ecliptic plane. This all suggested that it was a visitor from far away: a long-period comet with an orbital period of about 17,000 years. Thus, it might have been originated from the Oort cloud or even be an interstellar comet. Yet, its orbit and the short period for an Oort cloud comet (usually several million years) suggested that it was not the comet's first journey into the inner solar system but had paid visits to it several times already.

On the one hand, this was a pity as new comets are usually more active. On the other hand, these are also by far more unpredictable in their behavior compared to "well-established" ones. Hence, the chances were very high that Hyakutake would at least maintain its activity and develop into a very bright comet.

Another particularity with this comet was its location in the night sky as seen from Earth. While most comets are close to the Sun in the sky when they are brightest and thus are best observable either right after sunset or at dawn, Hyakutake was different also in that respect. Its orbit suggested that it would be visible almost the whole night from the northern hemisphere as it would be passing close to the pole star.

Encounter with Earth, Perihelion and Farewell
C/1996 B2 became visible to the naked eye in early March, only a few weeks after its discovery which is also quite unusual. Its brightness continuously developed further. Yet, by mid-March the comet's apparent brightness was at a not very impressive 4 magnitudes. Such a brightness may at first glance well, however, you should not forget that this is not a brightness of a point object but an integrated brightness of the entire comet head which may well cover an area of several degrees over the sky. The actual comet, therefore, appears much fainter than its brightness value might suggest. At that time, its tail already stretched to about 5° long.

While coming closer to Earth, its brightness rapidly increased and the tail grew significantly in length. Thus, already the day before the comet's closest encounter with Earth, it was already among the brightest objects in the sky with an impressive tail of about 35°. The full moon in comparison merely covers an area of about 35 arc minutes. To illustrate it even more, look at the sky and try to hold your hand at right angles to a fully extended arm. If you now stretch your thumb and little finger apart, the gap in between projected on the sky covers about 20°. In addition to is impressive size, the comet also exhibited a clear bluish-green color even to the naked eye which set it quite well apart even from light polluted urban skies. This color was owed to strong emissions of diatomic carbon (C_2).

As previously mentioned, Hyakutake had its closest approach to the Earth on March 25, 1996. By that time, it already move so rapidly, covering approximately the diameter of a full moon every 30 min, that position changes compared to the background stars could easily be seen with the naked eye within a few minutes (Fig. 3.24).

It had an impressive brightness of about 0 mag and tail that covered about 80° in the sky. Already its coma extended to 2°, which is about four times the size of the full moon. Its location close to the zenith made it even easier to observe. Indeed a great comet! Yet, it lost the competition with comet Hale-Bopp, which turned out to be even brighter. Also the public was more fascinated by the latter one, as it was visible for several months to the naked eye whereas Hyakutake fainted rapidly after the closest approach and thereby merely remained visible to unaided eye for about 1 month in total. In addition, it had been an easy task to observe for European (amateur) astronomers as its closest approach fell in an extended period of bad weather considerably deteriorating the chances to spot it.

The Great Comet of 1996, as it was named afterwards, reach its perihelion on May 1, 1996 and passed the Sun in a distance of about 0.23 AU which is well inside Mercury's orbit. While approaching its perihelion it brightened again and started to clearly present its dust tail in addition to its impressive plasma tail.

However, it was very challenging to observe because it was so close to the Sun. Fortunately, SOHO, the Solar and Heliospheric Observatory launched by NASA and ESA in 1995 was available.

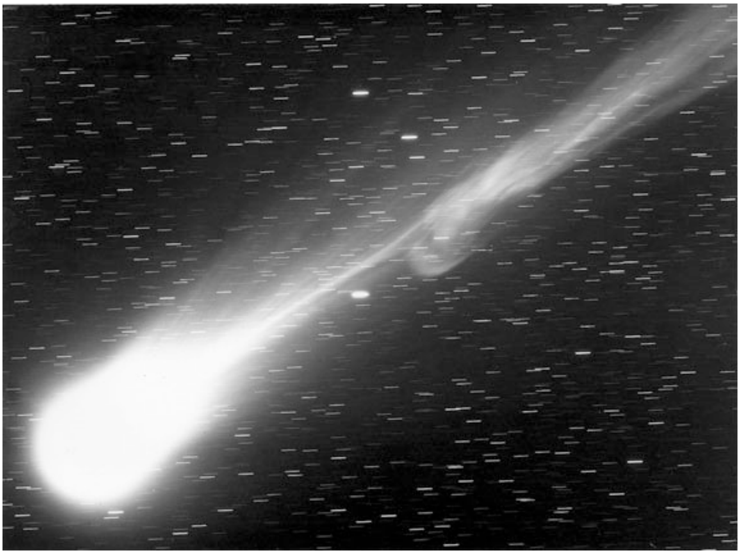


Fig. 3.24 Comet Hyakutake as photographed on March 25, 1996 (credit: E. Kolmhofer, H. Raab; Johannes-Kepler-Observatory, Linz, Austria, CC BY-SA 3.0)

After its perihelion, the fading continued. Its orbital path carried it rapidly into the southern skies, yet it was much less monitored there as after the perihelion it was much less impressive. The last known observation of the comet took place on November 2.

During its journey through the inner solar system, Hyakutake gravitationally interacted with the gas giants and the Sun. Its orbit was considerably stretched from an orbital period of about 17,000 years to an estimated 70,000 years.

Scientific Results or What Have We Learned

Hyakutake was not only a very bright and visually impressive comet, it also provided lots of new, fascinating information to the scientific community. We will learn about the most relevant ones in the following.

A question that arose quite early was whether the comet really originated from the Oort cloud. Examinations of its chemical composition revealed that it was atypical compared to other "normal" comets that had been known from our solar system. Hence, a possible interstellar origin was suggested. Did it really come from so far away? Was it the first interstellar comet to be actually observed?

The spectroscopic detection of ethane and methane significantly contributed to this hypothesis. The abundances of these two chemical compounds were found to be almost equal. Astronomers around Michael Mumma analyzed this and concluded that the comet may have formed far away from the Sun at temperatures of 20 K or lower possibly in interstellar space otherwise the Sun would have evaporated these volatile molecules.

While this hypothesis appears to be plausible, others come up that are at least equal to the former one. Some models suggest that the temperature distribution within the Oort cloud may not be equal. Thus, there may be very cold regions within the cloud with temperatures sufficiently low to allow Hyakutake to form.

Up to now, the question remains unanswered whether the comet really is of interstellar origin or simple an oddball from the Oort cloud.

Further spectroscopic analysis revealed the amount of deuterium in the comet. The ratio of deuterium and hydrogen was of particular interest. We have a specific ratio of these to molecules on Earth. It had been long assumed that a large portion of our oceans' water was originally provided by comets in the early phase of our solar system, when impacts had been much more likely. As both, hydrogen and deuterium are stable and not subject to radioactive decay, the ratio should stay the same over a very long period of time. If the theory of comets being Earth's water suppliers was true, the ratios measured in comets and on Earth have to be same or at least very similar. However, in the case of Hyakutake it was significantly different which suggests that there have to be various sources of Earth's water. Of course you could counter argue that the evidence was not really solid. C/1996 B2 was identified to be atypical in some respect why consequently not in this case of the deuterium-hydrogen radio, too? This is a valid argument. However, this finding of differing ratios have been confirmed with various other comets in the meantime, among them Hale-Bopp.

A great surprise was the discovery of X-ray emissions which had been detected by the satellite ROSAT (see Fig. 3.25). It was the first time ever that such emissions had been observed with a comet. In the beginning, this finding puzzled the scientists. Where did they come from?

They appeared to be ordered in a crescent shape around the nucleus and pointing away from the Sun. A combination of two mechanisms was found to be a good explanation of the emissions. First of all, there is an interaction between the energized solar wind particles and the cometary material sublimating from the nucleus, as we have seen earlier in this chapter. Subsequent observations of other comets, such as C/1999 S4 (LINEAR) with the Chandra satellite in 2000, suggested that X-ray emissions were produced predominantly by charge exchange collisions between highly charged carbon oxygen and nitrogen minor ions in the solar wind, and neutral water, oxygen and hydrogen in the comet's coma. Secondly, reflection of solar X-rays, as it is known from other objects of our solar system such as our Moon, may contribute to this. However, assuming even the highest X-ray reflectivity possible per molecule or dust grain is not able to explain the majority of the observed flux from Hyakutake, as the comet's atmosphere is very tenuous and diffuse. Yet, a combination of these two processes suits the observations quite well.

Interesting results were also obtained with regard to the nucleus, especially its size and activity. It turned out that Hyakutake's nucleus was small compared to other comets. Radio observations of the Arecibo Observatory suggested a size of about 4.8 km in diameter. Indirect infrared measurements later on more or less confirmed this magnitude of size. Other cometary nuclei are significantly larger.

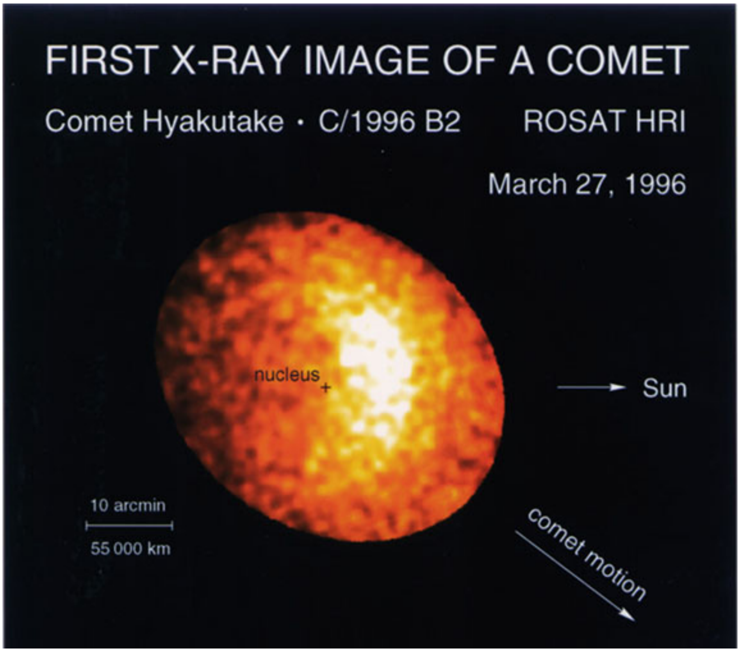


Fig. 3.25 An X-ray image of Comet Hyakutake, as observed by the High Resolution Imager on the ROSAT X-ray observatory. The X-rays arise primarily from a crescent-shaped region with a diameter of about 50,000 km, on the sunlit side of the comet (credit: C. Lisse, M. Mumma, NASA/GSFC; K. Dennerl, J. Schmitt, J. Englhauser, MPE)

The nuclei of comets Halley and Hale-Bopp have sizes of about 15 km and 60 km, respectively.

We have seen that usually only small regions of the nucleus are active. The vast part remains inactive. As the comet was very bright and its nucleus so small, most parts of its core had to be active in order to explain the observed pattern. The reasons for such a highly active nucleus are neither understood nor yet known.

In 1998 and 2000 several teams of scientists analyzed old data of the spacecraft Ulysses which had been launched into space in 1990 and was supposed to observe the Sun. In 1998, they found that on May 1, 1996, instruments on board the spacecraft detected a sudden large drop in the number of protons passing, as well as a change in the direction and strength of the local magnetic field. The scientists only had one explanation, Ulysses must had crossed the wake of an unknown object. The pattern suggested this to be a comet. However, the object remained unidentified.

Then, in 2000, two teams again analyzed the data related to this strange event. The first team of G. H. Jones and his colleagues were able to verify a change in the direction of the magnetic field that had been recorded by Ulysses' instruments. This pattern somewhat corresponded to the effects caused by a comet's plasma tail. Therefore, the started looking for likely suspects especially ones in the near neighborhood of the spacecraft. Yet, they didn't find one. Thus, they extended their search and identified comet Hyakutake. Could that be? It was so far away at a distance no one had expected to come into touch with a cometary plasma tail. Yet, all data indicated that Hyakutake was the right one.

The comet had crossed Ulysses' orbital plane on April 23, 1996. Taking into account the velocity of the solar wind, it would have taken 8 days for the tail particles to be carried to the spacecraft which perfectly coincided with the observations. In addition, the orientation of the plasma tail inferred from the magnetic field measurements agreed with the source lying in Comet Hyakutake's orbital plane.

G. Gloeckler and colleagues, forming the second team, independently from Jones analyzed the data and discovered a large spike in detected levels of ionized particles at the same time. The relative abundance of chemical elements detected indicated that the object responsible was definitely a comet and more precisely perfectly suited to the measured values of Hyakutake.

So, what did this discovery mean? Something unprecedented and completely unexpected had happened. Based on the Ulysses encounter, the comet's tail is known to have been at least 570 million kilometers (3.8 AU) long. Most likely it had been even longer. This is almost twice as long as the previous longest-known cometary tail, that of the Great Comet of 1843, which was 2.2 AU long.

3.4.5 Hale-Bopp: Comet of the Century

C/1995 O1, as is comet Hale-Bopp's official designation, may undoubtedly be called the comet of the century. It was one of the brightest comets ever seen and had a very impressive appearance in the sky in 1997 close to its perihelion and was visible for about 18 months to the naked eye. It topped the Great Comet of 1811, which hold the record of naked eye visibility at 9 months. For this reason and also influenced by an intensive media and Internet coverage, it was probably the most widely observed comet of the twentieth century. The comet was not only observed by professional and amateur astronomers but also by ordinary people on the street as Hale-Bopp was so bright and easy to spot in the sky. Even under heavy light pollution in big cities it was still remarkable and revealed not only a fuzzy coma but also tails to the unaided observer (see Fig. 3.26).

Yet, besides its impressive appearance, Hale-Bopp has been exceptional with regard to several aspects. Just to mention some of them briefly: it showed a quite unprecedented early activity and is the first comet ever a third tail had been observed, a sodium tail.

Discovery

Hale-Bopp was discovered independently by two American amateur astronomers, Alan Hale and Thomas Bopp, in the night of July 23, 1995. Hale, a passionate comet observer, who had already followed about 200 comets during their voyage

Fig. 3.26 Comet Hale-Bopp as seen from a larger, light-polluted city in Germany in 1997

through the inner solar system, looked at the night sky over New Mexico that night. He spotted a strange, potentially unknown object of an apparent magnitude of about 11 near the globular cluster M70 in the constellation Sagittarius just after midnight. Was this a new comet? He checked his star charts for possible close-by deep sky objects and a directory of known comets. Nothing! As he was also able to detect a movement of the object relative to the background stars, Hale was quite sure that he had indeed discovered something new, potentially a new comet. He, then, immediately emailed his discovery to the International Astronomical Union's (IAU) Central Bureau of Astronomical Telegrams at the Harvard-Smithsonian Center for Astrophysics.

In the meantime, Thomas Bopp was observing the night sky with some friends in Arizona. While peeking through a friend's telescope, he got a glimpse of the comet close to M70. Experienced as he was, he also consulted his star charts. Yet, he was unable to find any nearby deep sky objects. So, he also sent a Western Union telegram to the Central Bureau of Astronomical Telegrams.

There, in his small office, Brian G. Marsden (1937–2010) who had been in charge of the Bureau for almost 30 years handled these discovery notifications. He was a bit puzzled to actually receive a paper-based telegram in an era of electronic communication. However, the means by which a notification was transmitted did not matter. In the course of the night, the discovery of a new comet was confirmed and the comet was given its official designation C/1995 O1 and announced in the IAU circular 6187. Later on, following a long tradition, it was given the name of its discoverers: Hale-Bopp. Both of them had, by that time, no idea how famous "their" comet should become.

First Observations

Soon after its discovery it became apparent that Hale-Bopp was anything but a usual comet. Its orbital position after the discovery was calculated to be at 7.2 AU from the Sun. This is a distance where most comets do not show any activity yet and are thus extremely faint. Only a few distant comets are active in these regions of the solar system, such as Schwassmann-Wachmann, which raised some expectations. Most comets, as we have seen earlier this chapter, get active when they approach Jupiter's orbit, but Hale-Bopp was still well between Jupiter and Saturn and already showed a remarkable activity.

Furthermore, pre-discovery images such as one taken by the Anglo-Australian Telescope in 1993 showed the by then unknown comet at a distance of 13 AU, well beyond Saturn's orbit. At this distance, most comets are nearly unobservable due to their faintness. Halley's comet, just to illustrate this, was about 100 times fainter at the same distance from the Sun. All this started to nourish thoughts of Hale-Bopp being an exceptional comet.

As a comet's brightness is based on the reflection of sunlight from dust particles being driven from its nucleus to the coma by sublimating gases, it was tried to make some assumptions about the size of Hale-Bopp's nucleus. A size of 60 ± 20 km was derived, about six times the size of Halley. This assumption was further supported by the detection of a very large rate of outgassing carbon monoxide (CO). Indeed the initially suggested diameter was later on confirmed by further observations.

Will It Deliver: To Be or Not to Be

Though the expectations grew to be able to see a very bright comet, some comet scientists remained reserved. It is not uncommon that comets show heavy activity at greater distances to the Sun but then disappoint later on while approaching their perihelion. Either the development of a comet's activity slows down (e.g. most of the volatiles have been exploited) and stays behind the expectations or it simply breaks up. Comet C/1973 E1 Kohoutek and C/2012 S1 ISON are good examples for this.

In 1973, comet Kohoutek raised huge expectations and was already announced to become the "comet of the century" quite soon after its discovery. Due to its initially determined orbital path, it was considered to be an Oort cloud object making its first trip into the inner solar system. Thus, it was assumed that lots of volatiles should be available in its nucleus. A spectacular outgassing should be the result.

Yet, its appearance, though being visible to the naked eye, turned out to be a let-down. Scientists looked into the reasons for this and found at least two possible explanations. First, a partial disintegration of its nucleus might have happened during its journey caused by its warming up and the sublimation of gases. Second, further observations suggested, retroactively, that its origin was most probably not the Oort cloud but the Kuiper belt which would mean a different, more rocky and less icy composition and structure. This second point may explain the low level of outgassing.

The second example is the alleged bright comet ISON, formerly being designated as C/2012 S1. It was supposed to give a spectacular appearance in late 2013. Soon after the discovery of this sungrazing comet, media claimed that it would become a very bright comet, possibly even brighter than the full moon and thus turn out to be a new comet of the century. Yet, its activity stayed far behind the expectations. On the day of its perihelion passage (November 28, 2013), observations indicated that it fully disintegrated due to the Sun's heat and tidal forces.

This shows that one should be very cautious in predicting a comets development during its journey through the solar system. Too many unknown factors exist, being it the comet's actual composition or physical structure, which make it almost impossible to give reliable forecasts on the further development.

Fortunately, comet Hale-Bopp was in no respect like the two previously mentioned "failed" comets. It clearly exceeded all expectations. Nobody had assumed that the comet would be visible to the naked eye for a record period of 18 months! It allowed and still allows today intensive study of cometary behavior and characteristics, as we will further see.

The Hot Phase: Approaching the Perihelion

Time passed and Hale-Bopp came continuously closer to its perihelion, i.e. the closest point to the Sun on its orbit. As of May 1996, the comet was finally visible to the naked eye. By then, it had still a distance of 2 AU from the Sun and it was about to cross Mars orbit.

Scientists though remained cautious. Too often comets hadn't measured up to their expectations. Hale-Bopp's activity still looked promising though it had slowed down considerably over the last few months. Its brightness, however, further increased until the end of the year. Unfortunately, in December 1996, the comet was in line of sight with the sun and thus too close to it to be observable.

Its reappearance in January 1997 was all the more stunning. The comet was bright enough to be visible to anyone looking up into the sky. You could hardly miss it, even on highly light polluted skies (see Fig. 3.26).

Over the coming weeks, its brightness further increased while approaching its perihelion. In February 1997, it had already reached an apparent brightness of 2 magnitudes and showed two distinct tails on the sky, even to the unaided eye, a bluish gas tail and a yellowish dust tail. Though these tails were observable under urban conditions, their full beauty only revealed in rural areas far away from any disturbing light sources.

On March 22, 1997 Hale-Bopp was closest to the Earth with a distance of about 1.3 AU. Indeed a distance at which most comets would not be visible to the naked eye at all.

As the comet passed its perihelion on April 1, 1997, it gave a magnificent show. With its apparent brightness of −1 mag it was brighter than any star in the sky except Sirius and its dust tail stretched over 40° across the firmament. Each night it was already visible before full darkness and remained observable for most of the night in the Northern Hemisphere.

It triggered a hype on the Internet with thousands of web pages dedicated to the comet and where myriads of photos were posted. Such a public outreach had never been seen before.

After the Perihelion or Time to Say Goodbye

After its perihelion, Hale-Bopp bid farewell to the observers of the Northern hemisphere and moved on to the Southern hemisphere. Unfortunately, the comet was much less impressive there. Its brightness slowly but steadily decreased. But one could observe with the unaided eye how it faded away in summer and autumn 1997.

In summary, Hale-Bopp had been visible to the naked eye for 569 days, a new record and worth awarding it the title “comet of the century”.

The comet further faded while receding from the inner solar system. In spite of its increasing distance to the Sun, however, Hale-Bopp has been observable over all the years (see Fig. 3.27).

About 10 years after its perihelion passage, in October 2007, astronomers were still able to detect a CO-driven coma. This meant, Hale-Bopp was still active at an impressive distance of 25.7 AU. This is already pretty close to Neptune’s orbit, a distance where most comets have already been inactive for a long time.

Fig. 3.27 Comet Hale-Bopp, still active at a distance of nearly 2000 million kilometers from the Sun in 2001. Despite the very large distance from the Sun, the comet is still active as it continues to lose material (credit: ESO)

However, Hale-Bopp remained active. It was still detectable on August 7, 2012 at a distance of 33.2 AU from the Sun. Scientists expect that it will continue to be visible with large telescopes owing its faint brightness of about 30 mag until at least 2020.

Scientific Results or What Have We Learnt?

The unprecedented long activity and the involvement of many thousand observers lead to a flood of scientific results and new insights into the composition and structure of comets.

Orbit

Hale-Bopp's orbit resembles an extremely elongated ellipse with a high eccentricity value of 0.995 and is also almost perpendicular to the plane of the ecliptic with an inclination of 89.4°. The high eccentricity causes the comet to reach by far further regions of the outer solar system than e.g. Halley's comet. Soon after its discovery first calculations of its orbit were made. These suggested that Hale-Bopp orbited the Sun in about 4200 years. Its closest point to the Sun, the perihelion, was determined to be 0.914 AU and should go as far away as 525 AU in its aphelion. This, however, implied that the initial assumption that it was a new comet from the Oort cloud entering the inner solar system for the first time had to be wrong since this would require much longer orbital periods and a by far more distant aphelion. The Oort cloud is supposed to begin at a distance of about 2000 AU. Its origin from the Oort cloud appears to be sure since the aphelion is too distant for a comet originating from the Kuiper belt.

This, however, suggests that Hale-Bopp was not a first time visitor to the inner solar system in 1996. The high eccentricity of its orbit, though, suggests that there had not been many visits before. Usually, when a comet travels through the inner solar system, its orbit is perturbed by the gravitational interaction with the planets. Consequently, the comet should either be catapulted out of the solar system or forced on less eccentric orbits.

Though quite unlikely due its highly inclined orbit, during its last visit, Hale-Bopp had a close encounter with Jupiter in March 1996 coming as close as 0.77 AU to the gas giant. The gravitational influence of Jupiter's mass affected the comet's orbit tremendously by significantly shortening it to about 2540 years and reducing its previous aphelion from 525 to about 371 AU.

Further computations of its orbit revealed that Hale-Bopp had previously visited the inner solar system in July 2215 BC. By then, the comet almost had collided with

Jupiter. This near collision must have affected its orbit considerably. The effects were by far more severely than the orbital changes induced in 1996. Some scientists assume that, for this reason, 2215 BC may have been Hale-Bopp's first passage through the inner solar system.

Sodium Tail

As already indicated, the intensive observation of Hale-Bopp brought about some interesting and astonishing results. One of the most remarkable findings was the discovery of a third type of tail. Until then, it was thought that comets have two tails, a gas tail and a dust tail. Hale-Bopp, however, revealed the existence of another one, a faint sodium tail. Probably many other comets also have such a tail. But since it is very faint and can only be seen with powerful telescopes equipped with specialized filters, they have for sure simply been missed so far.

In the case of Hale-Bopp, the sodium tail lied between the gas tail pointing almost directly away from the Sun and the dust tail following the path of the comet's orbit.

Astonishingly, the sodium tail consisted of neutral atoms and not ions as one could have expected looking at the sources in the coma. It also covered a length of 50 million kilometers.

Though it was well known before that sodium emission existed with comets, discovery of a sodium tail being the potential source of these emissions was complete surprise.

Up to now, the mechanisms that cause these emissions are not fully understood. The sodium appears to come from the cometary inner coma. It is not clear, however, whether the actual source is the nucleus or not. Several mechanisms exist in theory that may cause the generation of source of sodium atoms. Currently, the most promising one is the collision of dust particles in the inner coma. Influenced by the Sun's UV radiation, sodium is then freed from these particles.

Organics

By using spectroscopic methods, Hale-Bopp showed the existence of many chemical compounds which did not come as a big surprise. Yet, many of the detected molecules had never been observed before with comets. It is further not yet clear whether these already existed within the comet from its beginning or had been synthesized by chemical reactions in the coma.

Deuterium Abundance

Basically there exist three national isotopes of hydrogen: normal hydrogen containing one proton, deuterium containing one proton and one neutron, and the radioactive tritium containing one proton and two neutrons.

Deuterium, sometimes in the form of heavy water, was found in Hale-Bopp along with normal hydrogen. Its abundance was found to be twice as high as that of Earth's oceans. This discovery caused some irritations since comets were thought to have brought significant amounts of water to the primordial Earth and thus should be accountable for most of ocean's water. If, however, Hale-Bopp's deuterium abundance was typical for comets, these cannot be the only source. Otherwise the abundance of Deuterium in Earth's oceans should resemble that of comets.

In addition to the existence of heavy water, Deuterium was also detected in many other compounds in the comet. Scientists analyzed the ratio of Deuterium and normal hydrogen therein. Surprisingly, the ratio highly varied from compound to compound. This suggested that different sources for these compounds had to be considered. Thus, the origin of the cometary ices could not merely be the solar nebula which would imply the same or at least very similar ratio as the nebula is thought to be homogeneous in its composition. Interstellar clouds are now believed to be the sources of the various cometary ices at least in the case of comet Hale-Bopp.

Detection of Noble Gases and the Place of Birth

The detection of Argon and other noble gases was another stunning result. One may, of course, ask what the purpose of detecting just another gas in a comet is. Nobel gases are very special with regard to their reactivity with other molecules. They are on the one hand highly volatile which means that they tend to sublimate very quickly after their threshold temperature is reached. Each of the noble gases sublimates at a particular temperature. They thus can be used as temperature indicators for the thermal history of the cometary ices as we will see soon. On the other hand, they are chemically inert. They thus do not actively participate neither in chemical reactions in the cometary ice nor in the coma. The latter fact also provides one speciality. We have seen that the compounds found the coma usually are daughter molecules derived from the original mother species that sublimated from the nucleus. This, however, means that direct inferences from the daughter molecules about the nucleus' composition is very difficult. In the case of noble gases it is different. Since these do not take part in chemical reactions, their abundance in the coma should somehow reflect the abundance in the nucleus.

If we assume the solar abundance pattern is also applicable to comets, then the abundance of noble gases in the comet should very similar to it. If they differ, we can assume that the temperature has varied and caused noble gases too be depleted.

Analysis of Hale-Bopp revealed that the noble gas Argon was enriched whereas the noble gas Neon, which is even more volatile than Argon, was more than 25 times depleted compared to the solar abundance. This, however, meant that Hale-Bopp's deep interior had never been exposed to temperatures of 35–40 K or warmer as Argon sublimates at these temperatures. In order to explain the depletion of Neon, Hale-Bopp must have been warmed above 20 K as Neon sublimates only at 16–20 K.

These temperature findings allow inferring the comet's origin. Several birth regions come into consideration, the Oort cloud, the Kuiper belt or the region between Jupiter and Neptune. The Oort cloud as its forming place is not possible, as we have already seen, with regard to its orbital characteristics. The region between Jupiter and Neptune is to warm to conform with the temperature findings. There, it would be warm enough for the sublimation of Argon and thus the detected abundance of Argon would differ. Thus, there remains only one location. The comet must have originally formed in the Kuiper belt and then migrated outwards to the Oort cloud, most probably as a result of Neptune's migration into the pre-Kuiper belt.

3.4.6 Comet Siding Springs: Cheek by Jowl with Mars

Comet Siding Springs is neither a large nor a really bright comet. It merely appears to be a typical smaller comet. The question you may thus ask is: why do we then cover it here in his book?

C/2013 A1, as the official designation of Siding Springs is, was special in one respect which made it more interesting to the scientific community than most other smaller comets: Siding Springs had a very close encounter with Mars on October 19, 2014. After its discovery, the first determined preliminary orbital elements suggested a collision with the red planet. Though, such an event was ruled out when its orbit had been further refined, the comet's flyby was still expected to be critical—not for Mars but for the many spacecraft currently operating on its surface or orbiting it. And there is currently a huge armada of space probes up there. The Martian sky is quite crowded with the NASA orbiters Mars Odyssey, Mars Reconnaissance Orbiter, and MAVEN, ESA's Mars Express and the Indian Mars Orbiter Mission. On its surface the two rovers Curiosity and Opportunity are still operating.

Discovery and Orbital Characteristics or Will It Crash?

Siding Springs was discovered on the night of January 3, 2013 by the professional astronomer Robert H. McNaught at the Siding Springs Observatory in Australia. At the time of its discovery it was still 7.2 AU away from the Sun. Its apparent magnitude restricted the observation to medium sized telescopes.

Precovery images, i.e. photos of the comet found on old archive photos that were taken before the discovery, revealed that Siding Springs had already been captured by the Catalina Sky Survey (CSS) on 8.12.2012 and by PAN-STARRS on 4.12.2012. Thus, observations could trace back much longer in time which allowed for an already more accurate computation of the comet's orbital elements.

Its orbit was found to be inclined by 129° towards the ecliptic plane and its movement to be retrograde, i.e. moving in opposite directions as the planets. The comet was supposed to arrive at its perihelion at 25.10.2014 at a distance of

1.399 AU from the Sun, which would bring him somewhere in between the orbits of Earth and Mars. The closest approach to Earth was determined to be on 5.9.2014 and coming as close as about 0.89 AU to our home planet. Its highly eccentric orbit, the eccentricity being very close to one, and the resulting extremely elongated ellipse on which it moved suggested that, Siding Springs was a dynamically new comet coming from the Oort cloud and paying its first visit to the inner solar system. Most likely it took him several million years to arrive in the inner solar system and it will take him another 1 million year outbound.

Initial observations by the Russian amateur astronomer Leonid Elenin in February 2013 revealed that the comet would be passing very close by Mars with a distance of only about 41,300 km from the center of Mars. This would be pretty close. Mars outer moon Deimos for example orbits the planet at a distance of 24,000 km. Due to the uncertainty of the orbital elements, the probability of crash could not be ignored. However, more refined later computations showed that the distance of would be about 139,000 km. Nevertheless, during its Mars flyby, comet Siding Spring will pass more than 16 times closer than the closest approach of a comet to Earth which had been comet D/1770 L1 Lexell with a distance of 2.3 million kilometers in July 1770.

Observations were made by NASA's Swift satellite, a multi-wavelength space observatory originally dedicated to the study of gamma-ray bursts (GRBs) and launched in November 2004, using its Ultraviolet/Optical Telescope (UVOT) and the Mars Reconnaissance Orbiter suggested the nucleus to be about 400–700 m in size.

The Encounter and Bidding Farewell

On October 19, 2014, the comet finally passed the red planet in a distance of 140,800 km, i.e. about one-third of the distance between Earth and the Moon. The space agencies were afraid that the dust and ions ejected from the comet could damage their spacecraft. Thus, they had to balance between the risk of damages and the unique opportunity given to observe a new long-period comet being formed of ancient material from the primordial solar nebula and proto-planetary disk in such a close distance far away from Earth and the Sun. Yet, they decided to manipulate the trajectories of their vehicles and maneuver them to the opposite side of the planet. The scientists assumed that this measure protected them from most of the particles while still allowing observations.

During the later course of Siding Spring's approach, however, it became apparent that the main part of the cometary tail would miss Mars by at least ten Mars diameter. So, a massive threat for the spacecrafts was not longer to be expected. Yet, higher-than-average-velocity dust particles ejected earlier during the approach of the comet remained a danger as these would collide with the Martian atmosphere, its moons and potentially also with spacecrafts being in orbit. Thus, the decision was maintained to keep the probes in their safe positions on the opposite side of the planet.

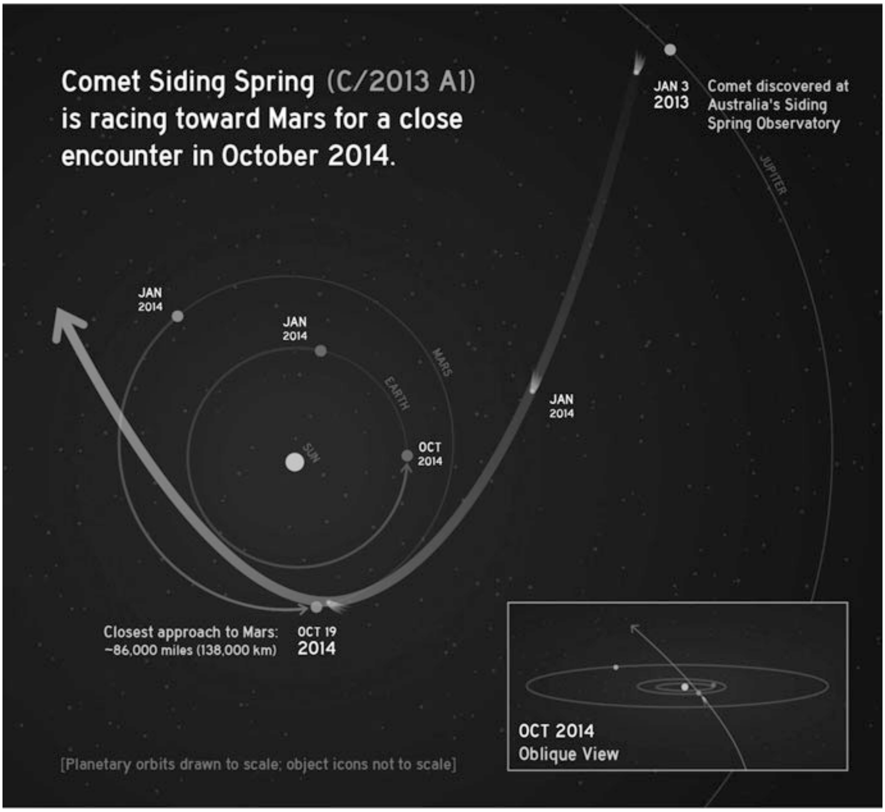


Fig. 3.28 Comet Siding Springs trajectory during its flyby of Mars (credit: NASA/JPL-Caltech)

As you can imagine the dust particles hitting the red planet's atmosphere were very small as they had previously been part of the comet's dust tail. They burned up in the upper atmosphere and created a meteor shower that would have been more than impressive for an observer on the planet's surface. Indeed the meteor shower was clearly detected by NASA's MAVEN and the two little observers on the Martian soil Opportunity and Curiosity (see Fig. 3.28).

Furthermore, debris from comet added a temporary, but strong layer of ions to Mars's ionosphere. This was the first time such a phenomenon had been observed on any planet. A few tons of cometary dust was vaporized high in Mars's atmosphere. Magnesium, iron, and other metals were observed to have had been deposited, too (Fig. 3.29).

Siding Spring's coma injected hydrogen into Mars' atmosphere and warmed it up by about 30 K for a few hours. An event that is also unprecedented in the history of cometary observations.

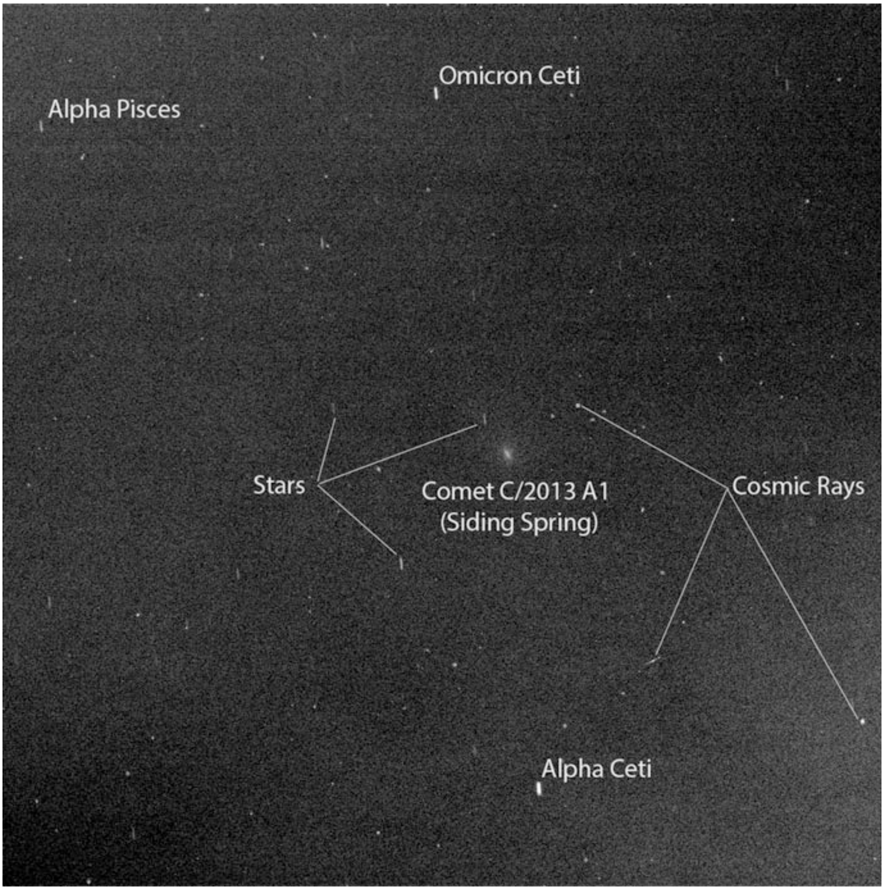


Fig. 3.29 NASA's Mars Exploration Rover Opportunity captured this view of comet Siding Springs from the surface of the red planet as it passed near Mars on October 19, 2014 (credit: NASA/JPL-Caltech/Cornell Univ./ASU/TAMU))

Thereafter, Siding Springs continued its journey back to the outer edges of our solar system. It can be taken for granted that the gas giants, and foremost Jupiter, gravitationally influenced the comets orbit. However, it is too early to clearly say how. Most likely its orbital period was drastically shortened though further observations are required.

Chapter 4
Trans-Neptunian Objects

For a very long period of time our solar system was believed to comprise nine planets: the terrestrial planets Mercury, Venus, Earth, Mars, followed by the gas giants Jupiter, Saturn, Uranus and Neptune. The exotic and tiny Pluto brought up the rear. Pluto was demoted from its planet status in 2006. Thus, only eight planets remain. In spite of this, the question remains what lies beyond the orbit of Neptune? Is it merely Pluto? Just plain, empty space? We have already come across the answer in the previous chapters. It is not empty space we can find beyond Neptune. A vast number of small bodies exist there. All of them together are referred to as trans-Neptunian objects (TNO).

In the simplest terms, we can distinguish two very different zones: the Oort cloud and a belt of small bodies. Whereas the Oort cloud is a spherical shaped envelope at the fringe of our solar system, the belt, on the contrary, is a relatively small and flat torus situated beyond Neptune's orbit.

We will see in this chapter that the structure of the belt is by far more complex than one might think. Two zones exist with different characteristics: the Kuiper belt itself and the so-called Scattered Disk. Both of these zones are quite different in certain aspects. The scientific community still debates whether the Scattered Disk should be considered as part of the Kuiper belt or treated separately. In the following, we will follow the latter opinion.

It may also occur to you to ask whether this belt zone beyond Neptune is just another asteroid belt, similar to the Main Asteroid belt between Mars and Jupiter. The answer to this is no, it is quite different from the one we know at the border of the inner solar system realm of the terrestrial planets. This distant belt is by far more massive (about 20–200 times than the main belt) and stretches over 20 AU from about 30 to 50 AU distance from the Sun. It consists of volatile materials compared to the rocky structure of the asteroids in the main belt.

In the following we will learn more about this zone and its members.

M. Moltenbrey, *Dawn of Small Worlds*, Astronomers' Universe,
DOI 10.1007/978-3-319-23003-0_4

4.1 Early Theories and Discovery of the Kuiper Belt

Soon after the discovery of Pluto by Clyde Tombaugh in 1930, speculations arose about a possible additional planet beyond the orbit of Neptune. The perturbations of Uranus' and Neptune's orbits led to the search for a Planet X that was postulated to cause these. However, soon after its discovery it became apparent that Pluto most probably could not be accountable for these perturbations as it was far too small. However, by that time, it was considered unlikely that another more massive planet existed out there. It was common belief that such a planet would have already been discovered.

Hence, several other theories were developed to tackle this issue. The American astronomer Frederick C. Leonard (1896–1960) was one of the first to throw his theory into the ring. He hypothesized about a trans-Neptunian population of smaller objects. Leonard stated that it was "likely that in Pluto there has come to light the first of a series of ultra-Neptunian bodies the remaining members of which still await discovery but which are destined eventually to be detected". Further, similar theories and postulations followed. In 1943, a more detailed theory was described by the Irish astronomer Kenneth Edgeworth (1880–1972) and further refined in 1949. He proposed a disk of icy bodies beyond the orbit of Neptune. Edgeworth assumed that the material in the original primordial solar nebula was only scarcely scattered in the region beyond Neptune. Thus, it was impossible to form a larger planet from these planetesimals as the probability of encounters between these bodies was very low. Consequently, the region, he suggested, should be populated by a vast number of small icy bodies. Some of these would, from time to time, wander into the inner solar system and form comets. Hence, the disk of icy bodies he proposed would not only help to solve the perturbation problem of the two ice giants, Uranus and Neptune, but also introduce a reservoir of comets lying close to the outer planets.

Further speculations followed. In 1949, the year when Edgeworth renewed and refined his theory, the Dutch astronomer, Jan Hendrick Oort (1900–1992) suggested that our solar system was enveloped by a spherical cloud of small icy objects, remnants left over from the formation of our solar system. He considered this cloud to be the source of long period comets. Oort thus contributed an important piece of information to our understanding of the modern solar system's structure.

In 1951, the American-Dutch astronomer Gerard Kuiper (1905–1973) also speculated about a structure of small bodies beyond Neptune. His theory, however, differed from previous ones, particularly the one introduced by Edgeworth. Contrary to Edgeworth, Kuiper assumed that a belt of small bodies had existed in the early formation of the solar system but dissolved soon after. As progressive as he was with his theory, Kuiper was also a child of his times. It was common belief by then that Pluto was not the large planet X that was originally expected but had at least the mass of the Earth. Having this mass, Pluto would have been able to clear its neighborhood by scattering nearby bodies away, potentially either further into the inner solar system or towards the Oort cloud.

As we know today, Pluto's mass is significantly lower, only about 0.002 times the mass of Earth. This mass is not at all sufficient for Pluto to properly clear its orbit of other bodies. This, as we will see in Chap. 5, is one of the main reasons why Pluto was demoted from a full-blown planet to a dwarf planet by the International Astronomical Union (IAU) in 2006.

In light of this, Kuiper's theory regarding the point of the postulated belt's present existence, becomes obsolete.

Although Edgeworth's and Kuiper's theories differ in several points, it is noteworthy that both described, independently from each other a sharp edge as the beginning of their belts at 30 AU with only Pluto being beyond Neptune.

The development of new theories did not stop with Edgeworth and Kuiper. In 1964, for example, Fred Whipple (1906–2004), one of the founding fathers of modern cometary science and well-known for establishing the term "dirty snowball" as a description of cometary structure, postulated that a large comet belt might exist beyond Neptune. He thought that this belt might be responsible for the perturbations in Neptune's orbit. Already then it was apparent that Pluto could not be accountable for this. Just 10 years after Kuiper, Pluto's mass was known to be clearly smaller than the mass of Earth.

Consequently, Whipple assumed that either another, tenth planet or his postulated comet belt would be perturbing Neptune's orbit. As already mentioned, the existence of a tenth planet was considered to be unlikely. Hence, Whipple suggested that the comet belt would have enough mass to be the cause for the perturbations in Neptune's orbit.

4.2 From Theory to Evidence

Everything mentioned so far had been mere theory. No evidence for such a belt had been established. If we exclude the observed comets, the only object known to orbit the Sun beyond Neptune was Pluto. Nothing else but empty space was known.

This changed suddenly in 1977, when the American astronomer Charles Kowal (1940–2011) working at the US Naval Observatory, discovered (2060) Chiron, an icy planetoid with an orbit between Jupiter and Uranus. Kowal had tried to find objects similar to Pluto and in his quest, had already searched half the ecliptic plane. He was in the fortunate position of having access to very sensitive instruments which were capable of detecting objects about 100 times fainter than Pluto. Despite this, he didn't find anything until November 1 of 1977.

Kowal didn't known at that time that Chiron would be merely the first of a new group of small icy solar system bodies, the so-called Centaurs. In 1992, another body, (5145) Pholus and in the following year (7066) Nessus, were discovered. Their orbits were similar to that of Chiron and all of them thus classified as a newly defined group of Centaurs. The name of this group is derived from Greek mythology where a centaur is a hybrid creature with the head, arms, and torso of a human and the body and legs of a horse. The newly discovered solar system bodies are also hybrids with respect to their nature and structure. On the one hand they have characteristics of asteroids and are thus classified in this group. On the other hand, they also exhibit features typical of a comet. Nowadays, it is assumed that Centaurs are extinct comets though their asteroid classification remains.

You probably ask yourself what the Centaurs, objects within Uranus' orbit, have to do with a small body belt beyond Neptune. For that we will take a brief look at Chiron's orbital characteristics. Chiron's orbit is highly eccentric (e = 0.37911) and inclined by 6.93° towards the ecliptic plane. With an average size of about 200 km it is not particularly small but also not a large object. The problem with the centaurs is that their orbits are highly unstable. This is because they are not protected by orbital resonances by at least one of the giant gas planets. This drastically limits their lifetimes. Either they collide with one of the giants or are expelled out of the inner solar system due to perturbations of their orbits by gravitational interaction with the planets.

This, however, gave the astronomers a problem. Assuming a fixed number of centaurs and their limited lifespans, the centaurs should all already have disappeared taking into account the lifetime of the solar system. As centaurs still exist, astronomers believe that their group is somehow replenished. Computer simulations suggested an origin from beyond Neptune's orbit, in the potential small body belt.

Although these simulations were not a definitive proof of the existence of the belt, they were nevertheless a strong indication that something was out there awaiting discovery. The study of comets also contributed to this notion. Comets, as we have seen, have a limited life time. During each journey into the inner solar system they lose quite a lot of material due to sublimation meaning that, 1 day, they will essentially disappear. The frequent discovery of comets suggests that these are also replenished. As we already known, we distinguish between two cometary groups—long-period and short-period comets. The long-period comets are supposed to originate from the Oort cloud. It was known that due to interactions with the gas giants and the Sun, orbits of long-period comets were perturbed. Thus, for some time, scientists believed that short-period comets were merely original long-period comets whose orbits had been drastically perturbed.

In the 1970s, however, astronomers realized that the number of short period comets that had been discovered was too high for all of them to come from the Oort cloud and hence another reservoir for these objects had to exist. How did they come to this conclusion? The basic idea behind it is as follows. Long-period comets move along extremely elongated ellipses and have very long orbital periods (in some cases up to several million years). In addition to this, as they arrive from the spherical Oort cloud, long-period comets may enter the inner solar system from any possible direction thus exhibiting the full range of inclinations towards the ecliptic plane. The orbits of short-period comets on the other hand are far less eccentrical and are only moderately inclined. We have already learnt about the transition between long and short period comets in Chap. 3. But for the sake of better understanding we will briefly repeat it. Long-period comets can be transformed into short-period ones by gravitational interaction with the gas giants, particularly with Jupiter. Each time long-period comets travel through the inner solar system, their orbits are influenced to some extent by these giant planets, e.g. By reducing the eccentricity of their orbits. Furthermore, the planets may drag the comets more and more towards the ecliptic plane. In certain cases and often after several passages, the cumulative effect of the perturbations may turn their orbits into less eccentric short-period ones with lower inclinations.

The problem astronomers now faced, however, was that due to the full range of inclinations present with long-period comets, only a few of them would interact with the gas giants at all and for each of the long period comets there was statistically a certain probability of being kicked out of the solar system or being forced on even more eccentric orbits rather than being transformed into a short-period comet. The total number of the number short-period comets observed could not be due only to travelers from the Oort cloud.

Based on these considerations, the Uruguayan astronomer Julio Fernández speculated in a paper in the Monthly Notices of the Royal Astronomical Society published in 1980 that a comet belt was needed to actually explain the observed number of short-period comets. He suggested that this comet belt should extend to between 35 and 50 AU from the Sun.

In 1988, the Canadian team of Tom Quinn, Martin Duncan and Scott Tremaine further investigated this together with further theories by running a number of computer simulations which confirmed that not all short-period comets could originate from the Oort cloud. On the contrary, assuming a comet belt as Fernández did, the simulations matched the observations. Tremaine subsequently named this “belt” the Kuiper belt. Legend has it Tremaine chose this name as Fernández had referred to Gerard Kuiper in the introduction of his paper. Tremaine was certainly not aware that his choice of name triggered a still-ongoing debate. There is a dispute in the astronomical community about whether the belt should be called Kuiper belt or not. Other names have been suggested and are used equivalently including the Kuiper-Edgeworth belt or the Edgeworth belt. It has become clear that Kuiper’s theory failed in at least one point. As we have seen, according to Kuiper, the belt should no longer exist today. As we will come to see, this is wrong. Edgeworth was closer to reality with his theory. Nevertheless, both Kuiper and Edgeworth contributed massively to the increased interest in such a belt and for this, they should both be honored.

4.3 Further Indications

Several further indications, some stronger than others, exist, which hint at the existence of such a belt. If we first look at the ice giant Uranus, we notice its awkward rotation axis of 97.8° which means the planet is basically rotating on its side. This is in stark contrast to the other planets. Scientists believe that such a tilt of the axis is only possible following a collision with an object of at least the size of the Earth. It can, however, be shown statistically that at least ten such objects are required for such a tilt of the axis to occur. This, again, suggests that a reservoir of bodies beyond the realm of the gas giants exists or at least existed at some time in the past.

The second indication comes when we think about the origin of the Pluto-Charon system. This is a strange system as the sizes of Pluto and its moon are so similar. The similar sizes of Pluto and its moon may be the result of a collision of the two bodies, in which Pluto captured Charon as its moon. Again, this requires several bodies to exist at least in the neighborhood of Pluto. We will learn more about this in Chap. 5.

The third aspect is related to Neptune's largest moon Triton. You may well ask what a moon of Neptune has to do with the existence of a small bodies belt beyond the planet's orbit. This will become apparent when we take a closer look at Triton's physical and orbital characteristics. The most striking anomaly we first encounter is its retrograde motion on its orbit around Neptune, i.e., it orbits in the opposite direction compared to the planet's rotation. This, however, implies that Neptune and Triton cannot have formed out of the same region in the protoplanetary disk. Hence, Triton must had come from a different location in the solar system. If we start looking for possible places, we quickly arrive at Pluto. Triton and the dwarf planet almost have the same size and a nearly identical chemical composition. This is strong evidence for a common origin. Yet, looking at the capture of Triton provides further indications. Previous theories postulated that the capture was the result of two objects that collided. Such as process in dimensions like this is nowadays often seen as very unlikely.

New theories discuss the possibility that Triton was part of a binary system. Such theories propose that at least one further body existed at the moon's cradle. When the binary system encountered Neptune it gravitationally interacted with the ice giant which in turn led to the dissociation of the binary. One partner was expelled into space, the other, Triton, was bound to Neptune and forced to orbit it for the rest of its existence.

You should, however, not forget that all of the aspects, we have considered, are merely indications. No evidence for the existence of a belt was available at the time. This suddenly changed in 1992.

4.4 1992 QB_1: Finally a Proof

The American astronomer David Jewitt and his student Jane Luu set an ambitious goal for themselves: to find another object beyond Pluto. Jewitt had already been puzzled about "the apparent emptiness of the outer Solar System" for quite some time and then convinced Luu to join him in his endeavor in 1987. "If we don't, nobody will.", was his valid argument.

The method they used was not too different in terms of its basic principle to that which Clyde Tombaugh had used whilst searching for Pluto. Jewitt and Luu employed a so-called blink comparator, which was invented to find the differences between two photographs. The photographs were inserted into the machine and these were then rapidly switched from one to another, i.e. blinking back and forth between these two images that had to be taken of the same region of the sky at different points of time. The idea behind is that a moving object will change its position during the course of time compared to the fixed background stars. You will then be able to see this object blinking at the two distinct positions.

Yet, in the actual application of the method there was a big difference. With the advent of CCD technology and thereby digital photographs, Jewitt and Luu no longer had to use photo plates. Taking analog photos of the night sky is a tedious and cumbersome task. CCDs are much faster and sensitive and also allowed for a much faster digital analysis on the computer screen. This, in no way, diminishes their achievements. It only helped them to cover much fainter areas in a much less time and to work through the tremendous amounts of data they had obtained.

It took Jewitt and Luu about 5 years until they finally succeeded in discovering of (15760) 1992 QB_1, as we know today is a classical Kuiper belt object and which then lent its name to these objects: cubewanos. Jewitt and Luu had identified an object slowly moving against the background stars. Initial measurements indicated that this object was located somewhere between 37 and 59 AU from the Sun and assuming a low albedo, i.e. reflectivity, had a diameter of about 260 km. So, there was actually an object out there. Yet, it was too early to rejoice. There was still the possibility that it was "merely" a newly detected comet. Months passed and more refunded orbital elements could be established which definitely ruled out the comet option. Today, we know that QB1 orbits at 43.7 AU from the Sun on a near-circular orbit ($e = 0.0654$) and only moderately inclined towards the ecliptic plane ($i = 2.19°$). Its diameter shrank to about 160 km.

About six months later, Jewitt and Luu struck again. This time, they discovered (181708) 1993 FW, a possible dwarf planet, orbiting the Sun at a distance of about 43.8 AU.

So, finally there was evidence for objects beyond Neptune's orbit and thus the alleged small bodies belt. The search continued and accelerated. By late 1994 already 15 objects had been found. In 2001, the first object with a satellite was discovered. Today, we know of more than 1,000 objects in this region.

Yet, the objects were initially given asteroid numbers and classified as asteroids. However, it soon became apparent that these objects were different to asteroids. They also did not match the profile of comets. They were more or less something in between. Hence, a new group was created: the so-called Kuiper belt objects (KBO).

Some of the KBOs are so large and similar to Pluto, that the planetary status of Pluto was called into question. Consequently, in 2006, the International Astronomical Union (IAU) introduced the new class of dwarf planets and demoted Pluto into it.

4.5 Origin of the Kuiper Belt

After the discovery of 1992 QB_1 and other objects in that region we can now be sure that a large belt of small solar system bodies exists beyond the orbit of Neptune and is awaiting further exploration. Yet, how did it form? This question needs to be answered before we can proceed to discuss the current structure of this belt and have a closer look at its members.

Unfortunately, the origins of the Kuiper belt are not yet fully clear and understood due to the lack of sufficient data. Too few objects have been discovered until now and the knowledge about their characteristics is still very limited.

The scientific community is eagerly awaiting the results of several wide-field survey telescopes among them Pan-STARRS and LSST. There are high expectations that new instruments will reveal many yet unknown Kuiper belt objects (KBO).

Several theories about the belt's origin are currently under discussion. One favored by many astronomers is that the Kuiper belt probably consists of planetesimals, i.e., fragments left over from the proto-planetary disc around the Sun that existed during the formation of the solar system. However, for some reasons these planetesimals did not make it to a full-blown planet. Several causes for this are conceivable. First, the planetesimals were so scarcely distributed that the likelihood of encounters were very low and due to this no further growth was possible. Secondly, it is also imaginable that similar to the main asteroid belt where Jupiter made the formation of another planet impossible, the influence of the gas giants in the direct neighborhood of the belt hindered the formation of a planet from actually happening.

Irrespective of which explanation is true, only small bodies formed, the largest of which were less than 3,000 km in diameter. The dwarf planets Pluto and Eris are currently found to be the largest known objects in the trans-Neptunian region, though it has still not been decided which of these two dwarfs is the largest. Anyway, we will see in the following chapter that the discovery of Eris in 2005 sounded the final death knell for Pluto's existence as a planet.

Still there is a difference between Pluto and Eris. The former lies within the Kuiper belt and is taken to be a prototype for a whole group of objects, the so called plutinos which are, as Pluto is, in 2:3 mean motion resonance with Neptune. That means for every two orbits of the plutinos around the Sun Neptune makes three. We will come back to this when we discuss the structure of the Kuiper belt.

The latter one, Eris, is not part of the (classical) Kuiper belt but is considered to be a member of the scattered disk which is overlapping with the Kuiper belt but extends much further into the outer solar system (up until 100 AU).

Thus, there are two different populations of objects in the trans-Neptunian region. Current theories suggest that both, the Kuiper belt and the scattered disk, developed from a proto-Kuiper belt which had initially been much closer to the Sun than they are today. Most likely they had also been much denser.

As computer simulations and the theories they are based on suggest, planetary migration played an integral role. The currently most promising theories are the Nice and Nice 2 model, named for the location of the Observatoire de la Côte d'Azur in Nice where they were initially developed by Rodney Gomes, Hal Levinson, Alessandro Morbidelli and Kleomenis Tsiganis.

One thing that is obvious is that both the Kuiper belt and the scattered disk were strongly influenced by the gas giants, and in particular Jupiter and Neptune.

4.5.1 Planetary Migration

So what did happen? After the dissipation of the gas and dust of the protoplanetary disk at the end of the formation of our solar system, the four gas giants Jupiter, Saturn, Uranus and Neptune were originally much closer together in a region stretching from about 5.5 and 17 AU from the Sun. Today their orbits cover an area of about 5–30 AU. Also the order is supposed to be different. The models indicate that initially Neptune had been closer to the Sun than Uranus and that these two planets later on swapped their positions.

Beyond the realm of the gas planets, a large disc of small rock and ice planetesimals existed. Its inner edge was just beyond the outermost gas giant between 15 and 20 AU and extended to about 35 AU. This disc, the proto-Kuiper belt was much more densely populated than the siblings, the current Kuiper belt and the scattered disk are. Scientists assume that the bodies forming the proto-Kuiper belt had a total mass of about 35 times the mass of Earth.

Even after the formation of the solar system, the orbits of the planets continued to change slowly due to interactions with a large number of the remaining planetesimals.

Some bodies from the disc's inner edge that were close to the outermost gas giant Uranus were gravitationally influenced by it. This caused perturbations of their orbits. In some cases, these small bodies were scattered inwards towards the inner solar system. Astronomers believe that this was true for the vast majority of small bodies. Only a few of them were sent outwards.

By scattering the small bodies inwards an exchange of angular momentum takes place. The preservation of angular momentum requires that when the small body is moved inwards, the planet is forced to slightly go outwards.

The inwardly moving objects then come closer to the next gas giants' influence zone, its Hills sphere. The process repeats and the planetesimal is cast further inwards. So, step by step, not only the small bodies move inwards, but also the gas giants change their orbits to farther distances from the Sun.

Of course the push the gas giants receive while interacting with the small bodies is very small, almost negligible and hardly influences the planet's orbit. Yet, the cumulative effect of many planetesimal encounters, shifts the orbits of the gas giants significantly.

Perhaps an analogy from our daily life, though not exactly the same, can help to better understand this. Think about a heavy basketball lying on the ground. Now, you throw a tiny, lightweight table tennis ball at it. What will happen as it hits the basketball? Probably nothing, though there is a tiny exchange of angular momentum. Let's change the situation and now cast many table tennis balls towards the basketball. Now, the latter will move slightly since the cumulative effect of all of the table tennis balls is sufficient to push it a bit away. This is what more or less occurred in case of the gas giants and the small bodies though they did not actually hit each other.

This process of moving orbits is what we call planetary migration. Yet, it does not continue endlessly. It comes to an end when the planetesimals encounter Jupiter, the by far most massive planet in our solar system combining in it more than two times the mass of the sum of all other planets together. Jupiter's enormous mass leads to a different kind of interaction. In most cases, the gas giant forces the small bodies on highly eccentric orbits or even kicks them out of the solar system. This is also considered to be the hour of birth of the Oort cloud. By scattering the small bodies outwards, Jupiter very slowly moves inwards due to the preservation of angular momentum.

4.5.2 Neptune and the Fate of the Proto-Kuiper Belt

Several hundred years (500–600 million years) of slow but gradual migration passed when Jupiter and Saturn reached their 2:1 mean motion resonance which increased their respective orbital eccentricities. This strong resonance caused a destabilization of the entire solar system in which essentially Jupiter pushes Saturn to its current position also due to mutual interactions with the two ice giants, Uranus and Neptune. Thus, Uranus and Neptune ended up with far more eccentric orbits and finally Uranus and Neptune swapped positions making Neptune the outermost planet.

In particular, the migration of Neptune had severe consequences for the proto-Kuiper belt in which it is now fully immersed (see Fig. 4.1). The ice giant approached the small bodies therein. Due to its gravitational influence, Neptune succeeded in capturing some of the small bodies into resonances while it pushed others into more or less chaotic orbits with higher eccentricities. This caused temporary chaos within the belt.

Many of the planetesimals, however, were cast inwards closer to the Sun. Their fates were then finally decided by their encounter with Jupiter. This process thinned out the proto-Kuiper belt drastically. Some scientists believe that more than 90 % of its small icy bodies were lost in this way.

The surviving bodies, either captured into resonances or thrown into more chaotic orbits, were consequently pushed outwards. Some of them formed what we call the Kuiper belt nowadays; others, especially those with more inclined and eccentric and potentially unstable orbits, established the scattered disk.

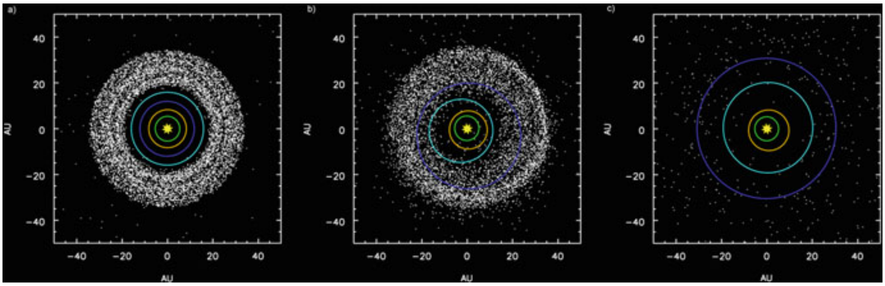


Fig. 4.1 Simulation based on the Nice model showing the migration of the outer planets and their effects on the Kuiper belt: (**a**) Before Jupiter–Saturn 2:1 resonance; (**b**) scattering of Kuiper belt objects into the solar system after the orbital shift of Neptune; (**c**) after ejection of Kuiper belt bodies by Jupiter. Planets shown: Jupiter (*green circle*), Saturn (*orange circle*), Uranus (*light blue circle*), and Neptune (*dark blue circle*) (credit: AstroMark, CC BY-SA 1.0)

However, it was not only the planetesimals which were affected but also the two ice giants. Friction within the belt made their orbits more circular again leading to the solar system as we know it today.

4.6 Structure of the Kuiper Belt or Bringing Order to the Chaos

We have just seen how the migration of the planets allegedly formed the Kuiper belt and the scattered disk. We have learnt that some of small bodies of the proto-Kuiper belt had been captured into resonances whilst others thrown into more chaotic orbits. Hence, there seems to be some underlying structure. Is there more? In the following we will first deal with the structure of the Kuiper belt before turning our attention to the scattered disk in the later parts of this chapter.

We know by now that the Kuiper belt consists of a huge number of small icy and rocky bodies. Yet, it is more than a mere chaotic aggregation. Astronomers were able to identify clear structures within the belt, which we will consider in the following sections.

4.6.1 Problems, Problems,...

The general problem we have is that large parts of the Kuiper belt are yet to be explored. By now, we know of some thousand objects in the Kuiper belt though it is assumed that many more exist. It is very difficult to discover, observe and analyze these objects as they are very small and faint. Even with the largest telescopes, which we have at our disposal, like the 10 m Keck telescopes at the Mauna Kea on

Hawaii or the Very Large Telescope of the European Southern Observatory (ESO) in Chile, it is hardly possible to resolve these bodies let alone to see any structure on a planetary disc. We even struggle to observe Pluto and this Kuiper belt object (KBO) is one of the closest and largest. Only relatively recently, with the help of the Hubble Space Telescope (HST) has it been possible to derive some very coarse surface maps of this dwarf planet. You probably now have some understanding of how difficult it is to observe even smaller and more distant bodies.

Hope lies with space probes visiting these outer regions of our solar system. The current New Horizons mission is a good example of this. Another option will be available in the next few years with the development of the next generation of large telescopes such as the Thirty Meter Telescope (TMT) and the European Extremely Large Telescope (E-ELT).

4.6.2 What We Know…

In spite of all these problems, what do we actually know of this almost undiscovered land? Scientists are sure that the belt stretches from roughly 30 to about 55 AU. Its clear borders are given by objects being resonant with Neptune.

The inner edge of the main body of the belt is determined by the 2:3 resonance with Neptune at about 39.5 AU while its outer limit is defined by objects being in 1:2 resonance with the ice giant at about 48 AU. Other resonances exist within the belt and are populated by smaller groups of objects.

The Kuiper belt itself is quite thick, resembling more a torus or doughnut than a belt. The main concentration of objects merely extends to about 10° outside the ecliptic plane a factor, which also distinguishes the Kuiper belt from the scattered disk whose objects are far more highly inclined. In addition to the main concentration, we can find a diffuse distribution of a few objects extending several times farther (see Fig. 4.2).

Neptune still influences the orbits of the KBOs. Hence, the structure of the belt is not final but changing continuously. Some objects may be sent to the scattered disk or into the inner solar system. However, this applies only to objects in certain regions. Many parts of the Kuiper belt are considered to be more or less stable. The unstable regions have thinned out over time. This results in observable gaps in the distribution of objects similar to the Kirkwood gaps of the main asteroid belt between Mars and Jupiter.

Basically, we can distinguish two parts of the Kuiper belt: the classical Kuiper belt and the resonant areas.

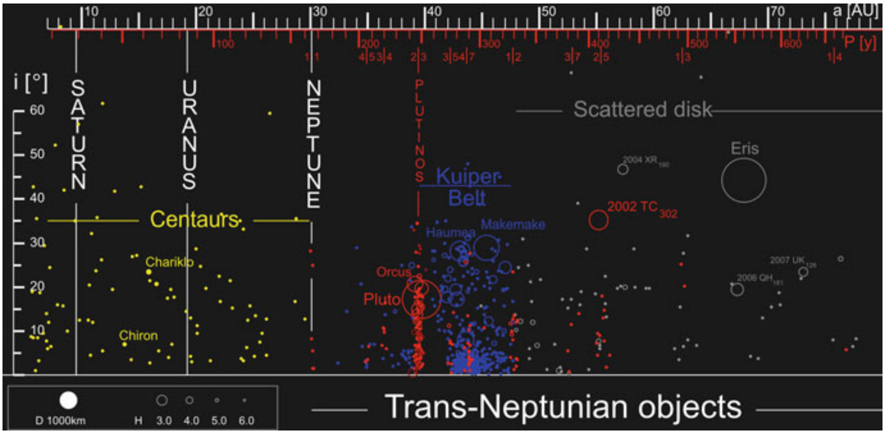


Fig. 4.2 Distribution of the trans-Neptunian Objects (credit: Eurocommuter, CC BY-SA 3.0)

4.6.3 The Classical Kuiper Belt

The classical belt or main body of the belt lies between the 2:3 and the 1:2 orbital resonance with Neptune (approximately stretching from about 40 to 48 AU) as depicted in Fig. 4.3. In this part of the belt, the gravitational influence of Neptune is negligible. Thus, the orbits of its members are relatively stable. About two thirds of the known KBOs are known/thought to reside in the classical belt.

1992 QB1 was the first classical belt object that was discovered and by chance also the first KBO at all if we exclude Pluto. That is way the classical belt objects are also often referred to as “cubewanos” (Q-B-one).

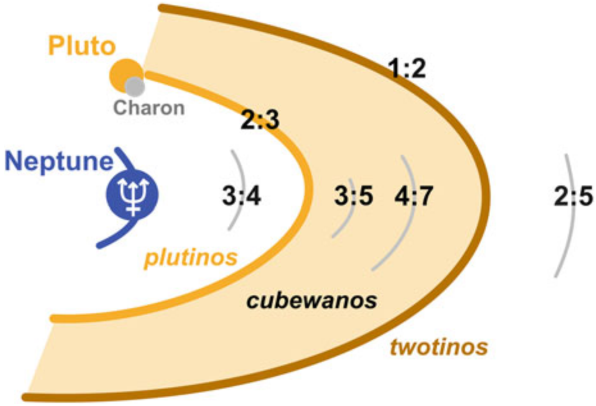
The cubewanos are characterized by low eccentric orbits and by the fact that they are not governed by any orbital resonance with Neptune. Astronomers were astonished to discover that these are not a homogenous group. Two populations appear to exist within the classical belt. They are called the dynamically hot and the dynamically cold populations.

4.6.4 Hot vs. Cold

These two populations differ in a number of characteristics. You should, however, note that the terms “cold” and “hot” do not refer to the temperature of these objects, as temperature is roughly the same for all objects in the Kuiper belt. These terms were chosen in analogy to gas particles and their respective speeds to each other. Thus, cold KBOs have a relatively lower speed to each other than the hot do.

The cold objects usually have lowly inclined, near circular orbits only deviating slightly from the ecliptic plane ($i < 10°$; $e < 0.1$). In this respect they move in a similar fashion to the planets.

Fig. 4.3 Structure of the Kuiper belt and overview over the resonances (credit: Eurocommuter/Lilyu, CC BY-SA 3.0)



The hot objects are by far more inclined (up to 30° towards the ecliptic plane). Furthermore, their orbits are clearly more eccentric.

However, it is not only the orbital characteristics of hot and cold objects in the Kuiper belt which differ. Surprisingly, they also exhibit different colors. The cold cubewanos are markedly redder than the hot ones which are more blueish in color.

The reasons for these colour differences are not fully understood. Some astronomers think that the two populations originally formed in different regions of the solar system. On the one hand, the cold objects? are considered to originate from the place where they currently reside or have just moved a bit outwards from due to planetary migration.

On the other hand, the hot cubewanos most likely formed closer to Jupiter and were then pushed to their current position by interaction with the gas giant.

Though this theory appears to be promising, no evidence to substantiate it is available yet. It is hoped that with the Deep Ecliptic Survey will in time reveal the actual origin of hot and cold cubewanos.

4.6.5 Distribution Within the Classical Belt

One observation that is also striking is that the cubewanos are not distributed evenly in the classical belt. The semi-major axis of the known objects indicates that there is a certain preference for the middle part of the belt. The closer you come to the edges of the classical belt, the fewer bodies you will find there.

How does this come about? The reason for this is really quite straight forwards. Smaller objects residing at the fringes of the belt are close to the limiting resonances, i.e. 2:3 and 1:2. Due to slight gravitational influences from other objects, especially Neptune, these small objects tend to be captured into one of the resonances or their orbits. This results in them being modified and in them being lead away from these zones.

4.6.6 Families or We Belong Together

We know that asteroid families do exist, e.g. in the main asteroid belt. The question is whether such families can also be found in the Kuiper belt. For a long time this was not clear. Nowadays, observational evidence provides the answer to this question.

The families in the Kuiper belt are supposed to have formed by the breakup of a single larger body. The remnants of this body share similar orbits and physical characteristics.

Astronomers claim to have found the first family in the Haumea family, which comprises the dwarf planet Haumea, itself, its moons, 2002 TX 300 and seven other smaller bodies. It is true that this family exists but Haumea had first been falsely classified as a classical KBO and later turned out to be a resonant object.

4.6.7 Resonant Objects

About one third of all known KBOs are in one of the mean motion resonances with Neptune. In particular, the 2:3 resonance group appears to be well populated. This is also the group we know most about and which is particularly famous for its most prominent member, the dwarf planet Pluto. Yet, other resonances exist where we can find objects: 1:2, 3:5, 4:7 and 2:5.

The objects being in 2:3 orbital resonance have semi-major axes of about 39.4 AU, i.e. for every two orbits they make, Neptune orbits three times. Up to now, we know of about 200 members. Pluto as the largest and most prominent one also gave the “trivial” name to this group: plutinos. You should, however, be careful not to confuse the term “plutino” with “plutoid” which also exists and is used frequently. The latter refers to all dwarf planets, the so-called ice dwarfs, in the trans-Neptunian region. It is in no way related to resonances.

In the following we will stick the term plutino when we refer to the objects being in 2:3 resonance. Many plutinos, yet not all of them, follow the example of Pluto and cross Neptune’s orbit. It takes them approximately 250 years for one orbit. However, due to being locked in resonance with the Neptune, they will never collide with the ice giant.

In general, the perihelion of the plutinos is inside or at least close to Neptune’s orbit whereas their aphelion takes them far out close to the outer edge of the classical belt near the 1:2 orbital resonance.

Their extremely eccentric and inclined orbits ($10° < i < 25°$; $0.2 < e < 0.25$) suggest that they originally did not form there but had been scattered into this regions by Neptune during planetary migration and subsequently captured in this resonance. If they formed there, they would probably exhibit more near-circular orbits as for example the members of the cold population of the classical Kuiper belt.

However, some plutinos deviate from this general pattern. 2005 TV189, for example has a highly inclined orbit with 34.5° towards the ecliptic plane. The plutino 2007 JH 43 follows a quasi-circular orbit. This shows that many things still need to be explored in order to properly understand the trans-Neptunian region.

The 2:3 mean motion resonance is quite stable one. However, you should not assume that the plutinos stay on their orbits forever. There are other factors present that may affect their journeys. The resonance band in which the objects are captured is narrow so that they may easily be kicked out of resonance. Now, think about Pluto. Its mass is relatively low, so the effects of its mass should usually be negligible and its zone of gravitational influence, its Hill sphere, is small. Yet, if a plutino approaches Pluto, the dwarf planet may perturb the plutino's orbit. In some case, this may essentially drive some plutinos out of the resonance and their fate will depend on Neptune.

The plutinos host some famous members such as Pluto, Orcus, Ixion and Huya. These are some larger objects and we will come back to them at a later point in this chapter.

The objects being in 1:2 resonance at the outer edge of the classical belt are often referred to as twotinos. The semi-major axes of their orbits are at about 47.8 AU and their orbital periods are at about 330 years. The twotinos move on moderately inclined and eccentric orbits ($i < 15°$; $0.1 < e < 0.3$). The group of the twotinos, however, is only sparsely populated and consists of far fewer members than the plutinos. Additionally, the 1:2 resonance is less stable than the 2:3 resonance. This may suggest that original twotinos and plutinos were equal in number but that the twotinos were more easily scattered.

The bodies in 3:5 resonance orbit the Sun at a distance of about 42.3 AU in about 275 years. It is quite a small group of bodies of which only ten members are known so far. The most prominent representatives of this group are (126154) 2001 YH140, (15809) 1994 JS and (143751) 2003 US292. Not much information is available about them, except their orbital characteristics.

The same holds for the 4:7 resonance objects. The 20 or so known resonance objects are located in the middle of the classical belt and orbit the Sun in about 290 years at a distance of about 43.7 from the Sun. Many things remain unknown. The only information we have is that the members of this group are quite small compared to other KBOs and that their orbits lie almost in the ecliptic plane.

The last group, we know about are the 2:5 resonant objects. With their orbital period of 410 years and a distance of 55.4 AU, it is questionable whether we should still consider them as belonging to the Kuiper belt.

4.6.8 The Strange Case of the Kuiper Cliff

One striking aspect is that beyond the 1:2 resonance edge only very few objects have so far been found. This is contradictory to the scientists' assumption and models. When estimating the primordial mass in the protoplanetary disk required to form the ice giants and the larger objects of the Kuiper belt, such as the dwarf planet Pluto, models suggested that the number of large objects should increase by a factor of about two beyond 50 AU distance from the Sun.

So, the question was whether cliff was really the edge of the Kuiper belt or merely the beginning of a broad gap. At least some objects were found at the 2:5 resonance as we have just seen.

Bernstein et al., however, could confirm that the cliff was indeed real and not due to observational problems.

The reasons for its existence remain unknown. Various speculations, some being more plausible than other, have arisen in the last few years. One of the more realistic ones suggests that the primordial material was simply too scarcely scattered in order to be able to form larger objects. The astronomer Patryk Lykawka recently speculated that an as yet unidentified planet, perhaps of the size of Mars or Earth, orbits the Sun in the outer regions of the solar system. The gravitational influence of this hypothetical planet would have hindered the formation of larger objects. Until now, no such planet has been found and nor are there any indications of its existence. Hence, the scientific community considers it very unlike that such a planet actually exists.

4.6.9 Composition of KBOs

In the previous sections we have learnt quite a lot about the structure of the Kuiper belt and the different types and populations of KBOs that reside within it. Yet, one thing we have not come across is their chemical composition. What are the KBO made up?

As we have already discussed it is very difficult to observe and measure the KBOs due to their small sizes, faintness and because they are so distant. Unlike many other objects of our solar system, scientists have not had the chance to directly observe or visit these KBOs using a space probe. The very first space probe to attempt this is the New Horizons mission. Its probe is currently approaching Pluto and will allow detailed analysis of the dwarf planet and its moons before it continues its journey to other KBOs. The exact KBOs to be visited have yet to be selected.

So far, analysis of chemical composition has been mostly restricted to spectroscopy. However, this is extremely difficult to carry out due to the faintness of the objects. Astronomers have only succeeded in spectroscopic analysis of a few KBOs. For the remainder, just general information can be derived. The data reveals that the KBOs exhibit a wide range of colors ranging from neutral grey to deep red. Their surfaces, and this is the only thing that can be studied by spectral analysis, showed a plurality of different chemical compounds, such as ammonia (NH_3), water ice and light hydrocarbons like methane (CH_4). The composition is strikingly similar to comets.

Scientists were quite surprised to discover such a plurality of different surface characteristics, as they had expected the KBO to be uniformly dark. They had thought that the volatile compounds on the surfaces of KBOs would have evaporated a long time ago due to electromagnetic radiation of the solar wind and cosmic rays.

What caused this diversity? One common explanation was resurfacing due to impacts of smaller bodies and outgassing. However, many astronomers have now disregarded this theory. The variations in chemical composition are just far too extreme to attribute them to randomly occurring impacts. This problem still needs to be solved.

Scientists hoped that they might find a solution to this question as soon as they achieved analysis of single KBOs. One of the first objects besides Pluto that revealed its secrets in more details is 1993 SC. In 1996, scientists were able to have a closer look at this KBO. Its surface characteristics are very similar to those of Pluto and Neptune's moon Triton which suggests a common origin.

Subsequently, it was possible to detect water ice on 1996 TO60.

4.6.10 The Mystery of Quaoar's Surface

50000 Quaoar is a large KBO with a diameter of 1260 ± 190 km and orbits the Sun at approximately 43.3 AU in 284.5 years. It thus belongs to the classical KBOs or cubewanos. Some scientists think it should be classified as a dwarf planet. Yet, the discussions continue. The astronomers Chad Trujillo and Michael E. Brown discovered it on June 4, 2002.

Quaoar has been under continued observation and analysis since its discovery and has caused some headaches to the scientists involved. Spectral analysis revealed ammonia hydrate (NH_3H_2O) and crystalline water ice on its surface. Both of which should not have existed. It also differs significantly from Pluto which is basically covered by methane (CH_4), solid nitrogen N_2, and carbon monoxide (CO).

The problem with these two compounds is that they should have been destroyed a long time ago by the influence of cosmic rays and electromagnetic radiation from the solar wind.

In addition, crystalline water ice, for example, requires that the surface temperature had at some point in time exceeded the critical temperature range of 105–125 K which is required for the crystallization process to occur. Yet, how was this temperature achieved? The average temperature in the Kuiper belt is only about 50 K.

Several possibilities have been discussed so far in order to explain the existence of these ammonia hydrate and crystalline water ice. A first attempt was made in describing radioactive decay in the KBO's interior as the driving force. This idea was, however, soon dispelled as the process may indeed heat up the interior but not the surface. The temperature on the surface is eventually limited to the radiation in space which is by far lower than the required critical temperature. Secondly, a heavy bombardment of micro-meteorites could be responsible for heating up the surface and leading to the crystallization of water ice. Yet, this is not compatible with the detected amount of ammonia hydrate. The impacts most probably would have caused a massive loss of ammonia into space.

Thirdly, if we assume that crystalline water ice and ammonia hydrate are unstable when bombarded by energetic particles (such as the ones from cosmic rays or the solar wind), there might be a solution. Usually, in such a case, crystalline ice should be transformed into its amorphous form when the induced energy breaks up the crystal bonds. The astronomers David Jewitt and Jane Luu, the discoverers of 1992 QB1, provide two explanations and their assessment. On the one hand, there may be a magnetosphere or atmosphere protecting the ice from this bombardment. This, however, seems to be not very realistic looking at Quaoar's size and knowledge we have about other KBOs of comparable size. On the other hand, the two astronomers propose resurfacing of the KBO and thereby a replenishment of the two compounds within the last 10^7 years. The reasons for this are again not yet known. Cryovolcanism may be the origin. Cryovolcanos, sometimes also called ice volcanos, are volcanos that erupts volatiles such as water, ammonia or methane, instead of molten rock. These substances are collectively referred to as cryomagma. After eruption, cryomagma condenses to a solid form when exposed to the very low surrounding temperature in the Kuiper belt.

Hope lies again with the New Horizons mission when the space probe, after visiting Pluto, will continue its journey towards other KBOs and deepen our knowledge of these strange small worlds.

4.7 The Scattered Disk

We have already seen that more than the Kuiper Belt exists in the outer regions of our solar system. Briefly, we mentioned the Scattered disk, which will be the subject of this section. It is a disc, or you may call it in general terms a structure, which overlaps with the Kuiper belt but extends much farther away.

4.7.1 Scattered Disk vs. Kuiper Belt

There is an ongoing debate in the scientific community about whether the scattered disk should be considered part of the Kuiper belt or independent of it. The distinction is not really clear cut, as we will see in the following sections.

We already know that the Kuiper belt is a relatively thick belt, resembling a torus or a doughnut that stretches from approximately 30 to 50 AU from the Sun. The two basic types of KBOs, the resonant ones and the cubewanos, have relatively stable orbits and are hardly disturbed by Neptune's gravitational influence.

The scattered disk objects (SDO) differ at least in this respect from the KBOs. They are neither locked in an orbital resonance with the ice giant nor far enough away to be out of reach of it. SDOs usually have very eccentric orbits with their perihelia being close to Neptune's orbit and their aphelia far away. We will come back to these details shortly. The bodies of the Scattered Disk can thus be subject to gravitational influence by Neptune if their perihelia bring them close to about 30 AU the Sun. In the other parts of their orbit they are mostly protected from Neptune's influence.

Another difference between KBOs and SDOs lies in the actual orbits. KBOs often have moderately inclined and eccentric orbits. The SDOs orbits are, on the contrary, highly inclined, often going up to 40° towards the ecliptic plane. Additionally, compared to those of the KBOs, the orbits of SDOs are also far more eccentric.

At first glance, the differences in orbits and sensitivity to Neptune's influence seem to provide a clear distinction between these two trans-Neptunian populations. This suggests that we should treat them separately.

Yet, on closer look, the situation does not remain as simple. It is true that the orbital characteristics between KBOs and SDOs are different, but they still may be subject to change from one into the other. On the one hand, a KBO may, as we have seen, be scattered by collision with another object or by Neptune's influence. Just think about the cubewanos that are close to the resonance bands. The slightest instabilities caused by exterior influences can lead to perturbations of the KBO orbits which in the end could throw them into the scattered disk. The KBO will then continue its existence on a more chaotic orbit as a SDO.

On the other hand, a SDO's orbit after being perturbed by Neptune when being close to its perihelion may be captured into an orbital resonance with the planet and thereby continue its life as a KBO. Passing from one population to another and back is statistically possibly several times during the lifetime of one of these bodies.

Some astronomers thus prefer to use the term "Scattered Kuiper Belt Object" (SKBO) instead of SDO. Others, like the Minor Planet Center at the International Astronomical Union (IAU) classify these populations into different groups. They distinguish those with stable orbits (KBOs) from those with unstable ones (SDOs and centaurs). In this book, we follow the latter point of view. In any case, most of it is just definition and naming and has no influence on the actual observations, measurements and amount of new knowledge we gain.

4.7.2 Discovery of the Scattered Disk

As you can assume, SDO are even more difficult to discover than KBO since they are more distant from the Sun. The standard technique frequently applied is the blink comparator that had proved so successful in detecting Pluto and many other small bodies. For the SDO, the advent of the CCD technology again, meant a big step forward in our ability to detect such small and faint objects.

The discoverers of 1992 QB_1 again proved to be successful. Jewitt and Luu discovered the first object to be classified as a SDO in 1996: 1996 TL_{66}. Three others followed soon after: 1999 CV_{118}, 1999 CY_{118} and 1999 CF_{199}. Today we know of more than 200 SDOs, including the possibly largest trans-Neptunian object, Eris. Brown, Trujillo and Rabinowitz had discovered Eris in 2005. This object, later on classified as dwarf planet finally pushed Pluto from its throne as a planet.

Statistically it is assumed that, looking at the origin of both the Kuiper belt and the scattered disk, similar numbers of KBOs and SDOs should exist. The fact that we know of so few SDO (slightly above 200) and far more KBOs (more than 1000) is probably owing to the observational difficulties involved in observing SDOs.

4.7.3 Origin of the Scattered Disk

The origins of the scattered disk are thought to be almost the same as those of the Kuiper belt. The scattered disk formed during planetary migration when Neptune immersed into the proto-Kuiper belt and started scattering its members. Some of them formed the Kuiper belt and others that were expelled into more chaotic orbits were the core of what we today consider to be the scattered disk.

However, the scattered disk is still in, continuous evolution. We touched upon this aspect already with respect to the distinction between SDOs and KBOs. If KBOs are captured in weak orbital resonances with Neptune or are at the border of the narrow stronger resonance bands, they develop some weak orbital instabilities over millions of years which then in turn can eventually shift them into the scattered disk. These are only isolated events that do not happen that frequently. However, over a long period of time, more and more KBOs will wander into the scattered disk and enlarge it gradually.

4.7.4 Composition of SDOs

What are the members of the scattered disk made of? Do we have any information about, particularly given that the scattered disk is so difficult to observe and analyze?

SDO have very low densities and are mostly composed of frozen volatiles such as water ice and methane, as is true for most of the known trans-Neptunian objects. Very often similarities between KBOs and SDOs can be found which do not come as a real surprise when considering their similar formation histories and the on-going interchange of objects between these two populations.

Yet, there is one major difference whose cause still needs to be found: the colors. Originally it had been assumed that all trans-Neptunian objects should have a red surface color. We have already seen that assumption was turned upside down when the cold reddish-colored and the hot blueish-colored populations of the cubewanos were discovered.

What color do SDOs then have: red or blue? Simply put: neither of them. Indeed most SDOs have a white or grayish appearance. What could cause these colors? Firstly, sublayers of the surface may have more whitish colors due to the chemical compounds being frozen there. Impacts of smaller bodies or micro-meteorites may then expose these sublayers to us. Secondly, the American astronomer Mike Brown proposed, that due to the great distance of SDOs to the Sun and the thereby the very low temperatures of SDOs, the methane atmosphere of the SDO may be entirely frozen and cover the surface of the body as a thick layer of whitish material.

Similar explanations have been found in the case of Pluto. Yet, since the dwarf planet is much closer to the Sun and the temperatures are much higher, the methane will eventually only freeze onto cooler, high-albedo regions. The albedo describes the ability of an object to reflect light, in this case sunlight. If the reflection rate is very high, as for high-albedo regions, most of the sunlight is reflected. Hence most of the energy is lost, as it is not conserved by the body but re-emitted into space. These regions do not heat up as low-albedo regions. This may cause the contrasting black and white zones on Pluto that have been detected.

4.7.5 Orbital Characteristics of SDO or How to Cope with Mere Chaos

The orbits of SDO are by far more chaotic than those of most other objects in our solar system, with the notable exception of the Oort cloud. The scattered disk is a very dynamic environment in which many things occur. The most prominent contributing factor to the apparent chaos is, of course, Neptune and its tremendous mass. SDOs are in permanent danger of being disrupted by it.

The orbits of SDOs are characterized by medium and high eccentricities with their semi-major axes lying beyond 50 AU. Yet, they come as close as about 30 AU to the Sun at their perihelia. Their aphelia, on the contrary lie much farther outwards.

Neptune plays a very important role here. It continuously perturbs the orbits of these bodies. In addition, the SDOs also influence each other. All this results in more or less chaotic, random motion. Very weak, temporary orbital resonances with Neptune are known at 1:3, 2:7, 3:11, 5:27 and 4:79. These resonances, however, are so weak that it is relatively easy for an object to leave or join what?.

The decisive question remains: What happens when Neptune perturbs the orbit of an SDO? Basically, we can think about three different possibilities. Firstly, if the perturbations are minor, the SDO will simply continue its journey on another orbit in the scattered disk. Secondly, the gravitational influence of the ice giant may kick an SDO farther outward of our solar system towards the Oort cloud. The third option is probably the most interesting one, but also the option which still raises the most questions: Neptune may, in this case, send the SDO inwards into the inner solar system. The SDO will then, from a classification point of view, eventually turn from a trans-Neptunian object to a cis-Neptunian object. Astronomers believe that the centaurs, which have their semi-major axes between those of the outer planets, are the result of this process. Centaurs have unstable orbits that will further cross the orbits of one or more of the giant planets. Since these centaurs are in no way protected by orbital resonances by one of the planets, their orbits remain unstable and will further be perturbed by at least one of the gas giants. Thus, the centaurs will eventually have dynamic lifetimes of a few million years. Some of the centaurs will evolve into Jupiter-crossing orbits whereupon their perihelia may fall into the inner solar system. The former SDOs and now centaurs may then be reclassified as active comets in the Jupiter family if they display cometary activity. The fate of these objects is already determined. Either they will collide with the Sun or a planet or else they may be ejected into interstellar space after a close approach with one of the planets, particularly Jupiter.

How can we be sure about such a development? There are dynamical studies that support this theory. In any case, the centaurs are a good starting point for being an intermediate body, as they typically behave with characteristics of both asteroids and comets. We know that many SDOs and also KBOs have icy compositions similar to comets. Due to their large distance to the Sun, the energy from solar radiation is not sufficient to trigger activities on these bodies. Yet, if centaurs come closer to the Sun, it is assumed that they may exhibit comet-like activity.

4.7.6 Kuiper Belt and Scattered Disk as a Reservoir for Comets

In order to conclude our trip through the Kuiper belt and the Scattered disk, we will come back to the original problem that triggered the search for these structures. Where do all the observed short-period comets come from? We have seen that long-period comets from the Oort cloud cannot be accountable for all of them.

Initially it was believed that the Kuiper belt would be the origin of most of the short-period comets. Yet, soon after its discovery it became apparent that the KBO's orbits are too stable and that these comets have to originate from a more dynamic environment: the scattered disk.

In particular, the Jupiter family short-period comets are considered to originate from the scattered disk. As we have seen, the centaurs are considered to be the intermediary between the SDO and the final comet.

The problem is that several differences exist between SDO and Jupiter family comets. Although the centaurs share a reddish or neutral color with many SDOs, their nuclei are bluer, indicating a fundamental chemical or physical difference. One hypothesis is that comet nuclei are resurfaced as they approach the Sun by subsurface materials, which subsequently bury the older material.

4.8 Detached Objects or Strollers in the Middle of Nowhere

The scattered disk, however, does not seem to define the end of the regions in our solar system inhabited by small solar system bodies. It had been thought that between the scattered disk and the inner border of the Oort cloud, a large area of emptiness would exist. This notion was disproved by the discovery of Sedna (2003 VB_{12}) in 2003 by the Mike Brown and his colleagues. The newly discovered object is strange in many respects. It has very eccentric orbit and is far away from the Sun. Sedna reaches its perihelion at about 75 AU from the Sun. Sedna is no longer subjected to the influence of the planets and is in some way "detached" from the rest of the solar system.

Sedna, thus, forms the prototype of a new class of objects in our solar system, the so-called detached objects. These are considered to be a very dynamical class of trans-Neptunian objects. They are basically characterized in that their orbits have perihelia sufficiently distant from the gravitational influence of Neptune that they are not or only moderately affected by the ice giant or the other planets.

So far, at least nine detached objects have been discovered among which Sedna is clearly the largest and most distant. It is so special that it stands as prototype for a subclass among the detached objects—the sednoids.

Sednoids are defined to be trans-Neptunian objects with a perihelion greater than 50 AU and a semi-major axis greater than 150 AU. As of 2014, only two sednoids are known: 2012 VP_{113} and Sedna itself. Both have perihelia greater than 75 AU and it takes Sedna, for example, about 12,000 years to orbit the Sun. Scientists believe that more of these bodies exist and are awaiting discovery. The problem with both sednoids is that they spend only a very small fraction of their orbital periods near the Sun. Only during that period are they observable. This limitation is due to their size and faintness. This suggests that a larger number of similar orbits could exist out there which are currently undetectable.

4.8.1 Orbits and the Problem of Their Origin

Detached objects have very highly eccentric orbits which take them far out into the outer limits of our solar system. They have semi-major axes of up to a few hundred astronomical units. Their perihelia are so distant from the Sun that the gas planets, in particular Neptune, do not or at least hardly gravitationally influence them.

And herein lies a big problem. Looking at the peculiar orbits of the known detached objects, it becomes apparent that such orbits could not possibly be the result of scattering by the gas giants. But how did the detached objects get there and what caused their extreme orbits? The answer to these questions has not yet been definitively answered, though several speculations exist.

One possibility could be, as Morbidelli and Levison suggested in 2004, the passing of a nearby star. Our Sun is supposed to have existed in a young star cluster during the early formation of our solar system. The members of this cluster would have formed in processes similar to the formation of our Sun in the interstellar cloud. During this period it was likely that our Sun encountered other stars in the cluster.

Two things may have happened during such an event when a star of this cluster having a protoplanetary disk like ours at that time passed. First, if the passing star was small relative to our Sun, such as a low-mass star or a brown dwarf then, during such an encounter, a significant amount of material could have been transferred. Our Sun may have captured some of the extrasolar planetesimals and forced it into eccentric orbits. The eccentricity and inclination of these orbits would then depend on the inclination and speed of the passing star.

Secondly, a more massive star passing our Sun may have lifted the orbits of some SDOs to greater semi-major axes due to the Suns gravitational influence.

Another possibility that may have caused the present results is the existence of a distant planetary-mass solar companion. Gomes, Matese, and Lissauer proposed that a Neptune-mass companion at less than 2000 AU from the Sun or a Jupiter-mass planet at less than 5000 AU could be held responsible for tearing SDOs out to their present location and orbits. Trujillo and Sheppard argued that a “super-Earth” of between 2 and 15 Earth masses near the ecliptic plane at a distance between 200 and 500 AU would be sufficient to explain the orbits of detached objects.

So far, none of these theories could be verified. The Wide-Field Infrared Survey Explorer (WISE), a space telescope launched in 2009, conducted a survey on “nearby” objects of planetary size. The preliminary result, published in March 2014, stated that no planet of Jupiter’s size had been found within a distance of 26,000 AU from the Sun. Taking these findings into account, the existence of another planet orbiting our Sun at greater distances becomes less attractive and realistic. Yet, further surveys may help to enlighten our knowledge in this area.

4.9 Prominent Trans-Neptunian Objects

As we have learnt a lot about trans-Neptunian Objects in this chapter, it is a good point in time to show some prominent examples of these small solar system bodies. Figure 4.4 shows an overview of the largest TNOs.

4.9.1 Sedna: A Remote, Detached Object

The astronomers Michael (Mike) Brown, Chad Trujillo and David Rabinowitz discovered a new large trans-Neptunian object (TNO) on November 14, 2003: (90377) Sedna (see Fig. 4.5). This object was unique and in some respect awkward. Nothing like it had been discovered before. With respect to the naming, Brown said, “our newly discovered object is the coldest most distant place known in the solar

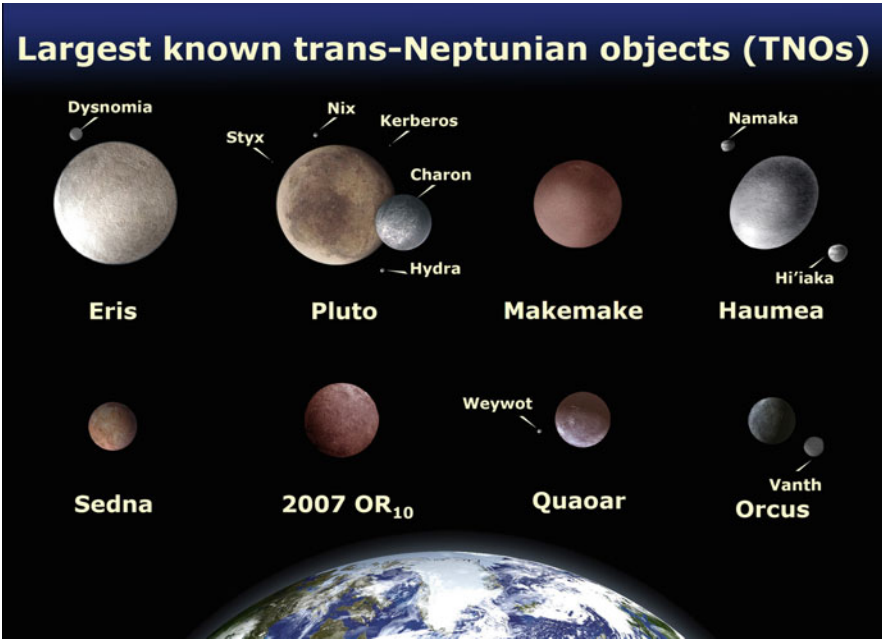


Fig. 4.4 Overview over the largest known trans-Neptunian Objects (credit: Lexicon, CC BY-SA 1.0)

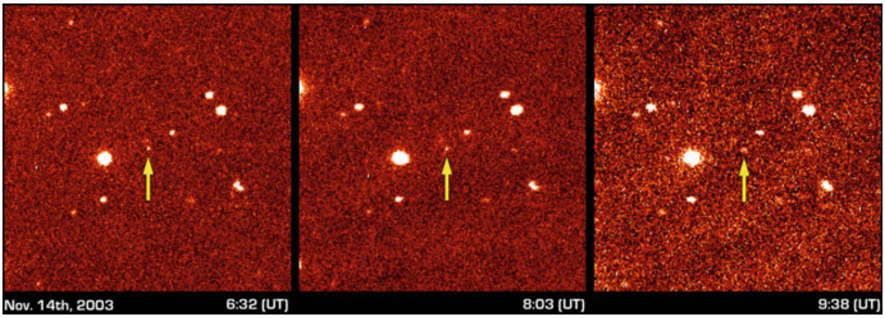


Fig. 4.5 Discovery images of (90337) Sedna (credit: NASA/Caltech)

system. So we feel it is appropriate to name it in honor of Sedna, the Inuit goddess of the sea, who is thought to live at the bottom of the frigid Arctic Ocean."

Orbit

Sedna's orbit is remarkable. This TNO has the longest orbital period of all known large objects in our solar system. Long-period comets originating from the Oort cloud sometimes having orbital periods of several million years are excluded, as they are not considered to be "large" objects.

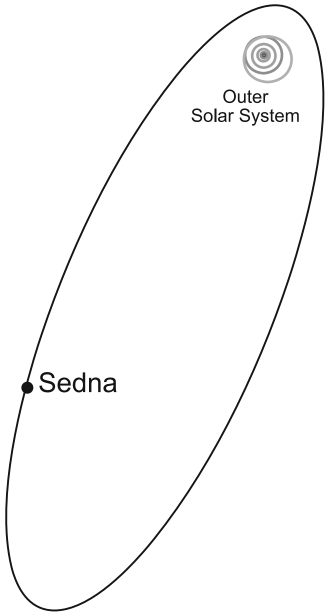


Fig. 4.6 Orbit of Sedna put in relation to the outer solar system planets

It takes Sedna about 11,400 years to orbit the Sun once. Its orbit is highly eccentric (e = 0.855) and moderately inclined (i = 11.93°) as depicted in Fig. 4.6. Its perihelion lies at about 76 AU which. This is one of the largest perihelia of any known solar system object. Its aphelion is much farther away at approximately 937 AU while its semi-major axis is at 524.4 AU.

When it was discovered, Sedna was about 89.6 AU away from the Sun, rendering it the most distant object that had yet been observed by that time. Later on, the dwarf planet Eris was discovered at a distance of 97 AU from the Sun. Eris is only temporarily farther away. At the time of its discovery, the dwarf planet was close to its aphelion while Sedna was approaching its perihelion.

Observations made at the Multiple Mirror Telescope (MMTO) at the Fred Lawrence Whipple Observatory located in Arizona (USA) in 2005, suggest a rotation period of about 10 h.

Physical Characteristics of Sedna

As for most TNOs it is very difficult to obtain reliable information about Sedna. It was possible to determine its distance. Furthermore, Sedna's albedo is estimated to be about 0.32. Together with its absolute magnitude (brightness) of about 1.8, it is possible to arrive at a rough estimate of 1,000 km for its diameter (Fig. 4.7).

The Hubble Space Telescope carried out a search for potential moons in 2004. The result was negative. The lack of any moons confronts the astronomical community with a significant problem. As long as no moons or any other close objects exist that gravitationally interact with Sedna, it is impossible to determine its mass without sending a spacecraft there.

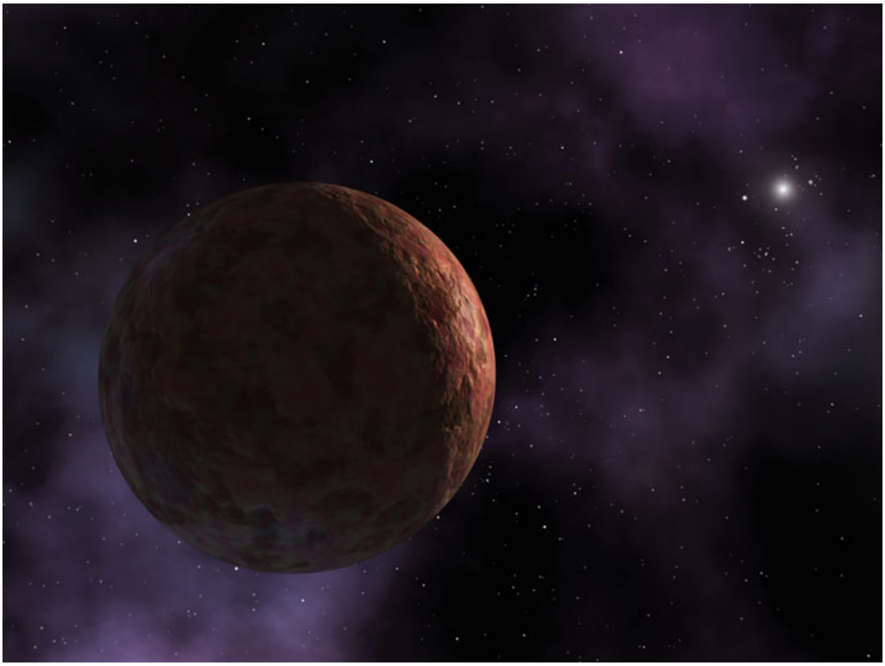


Fig. 4.7 Artist's visualization of Sedna (credit: NASA/JPL-Caltech/R. Hurt (SSC-Caltech))

What else do we know about this very distant TNO? When observed in visible light it has a very reddish color. In fact, it appears to be one of the reddest objects in the solar system and is almost as red as the red planet Mars.

Trujillo and his colleagues propose that the red color may be due to a cover hydrocarbonate sludge or tholins formed by space weathering, i.e. by long-time exposure to ultraviolet radiation. Sedna's surface also appears to be homogenous in color and spectrum. At least photometric analysis, e.g. light curves, did not show any differentiating features. This may be owed to Sedna's huge distance. Astronomers assume that this region of our solar system is only very scarcely populated though we do not know for sure. This was also probably the case during the formation of the solar system. It is thus very unlikely that collisions with our bodies occurred rendering Sedna's surface indistinct.

In 2005, Trujillo and his colleagues were also able to set upper limits in Sedna's surface composition of 60 % for methane ice and 70 % for water ice. The presence of methane further contributes to their assumption that tholins exist on its surface. Further analysis also revealed the existence of nitrogen. M. Barucci and his colleagues suggested, after comparing Sedna's spectrum to that of Neptune's moon Triton, that the surface may be composed of 24 % tholins, 7 % amorphous carbon, 10 % nitrogen, 26 % methanol and 33 % methane.

Sedan may have a tenuous atmosphere induced by the presence of nitrogen ice on its surface. During a period of about 200 years close to its perihelion, surface temperatures may rise above the critical temperature to trigger sublimation of the frozen nitrogen on its surface. An atmosphere as is known from Pluto or Triton created by the sublimation of methane is considered to be unlikely due to the very low temperatures on Sedna's surface.

Origin: Wrong Orbit, Wrong Place?

Sedna has a very peculiar orbit. Astronomers believe that it is not very likely that Sedna formed with its present orbit. However, its perihelion is also too far away from Neptune to get influenced by its gravity. Therefore, it cannot have been scattered there by the gas giant's gravitational influence. Sedna might be the first discovered inner Oort cloud object. The inner Oort cloud is a hypothetical disc extending from the fringes of the Scattered Disk to the spherical outer Oort cloud, which envelopes our solar system.

So, let us assume Sedna formed at its current location. This implies that the protoplanetary disc must have extended to at least 75 AU from the Sun. Sedna's initial orbit further must have been nearly circular. If that were not the case, its formation would not have been very likely if at all. In case of a more eccentric orbit there would probably have been larger relative velocities between the planetesimals in the protoplanetary disc. This would have been too disruptive, e.g. when collision occurred. These disruptive forces would have overweighed the accretion process.

Hence, we can see a clear mismatch between Sedna's current orbit (highly eccentric) and the presumed initial one (near circular). What does this tell us? Simply spoken, Sedna must have been brought to its current orbit by some gravitational interaction with another unknown body.

Several scenarios have been discussed in the scientific community. Yet, no final conclusion could be made. Nevertheless, we will briefly go through the currently discussed scenarios.

Firstly, an unknown planet beyond the Kuiper belt and the Scattered disk may exist. Secondly, a binary companion of our Sun could be the reason and thirdly, a passing star could have exhibited gravitational force on Sedna and brought it to its current orbit. Fourthly, Sedna may be of extrasolar origin.

Let us first look at scenario 1. This hypothetical planetary sized body would probably exist in the (inner) Oort cloud. Numerical simulations show that indeed Sedna's orbit could be explained by an approximately Neptune-sized object at a distance of 2,000 AU or less. Assuming a Jupiter-mass object a distance of 5,000 AU would be sufficient and even an Earth-mass object at 1,000 AU could have caused the perturbation of Sedna's original orbit. However, there is one major issue with this scenario: no such object has been found so far. As of March 2014, a sky survey carried out by the Wide-field Infrared Survey Explorer (WISE) telescope has ruled out the possibility of a Saturn-sized object out to 10,000 AU and a Jupiter-sized or larger object out to 26,000 AU (see Fig. 4.8).

Scenario 2 suggests that Sedan's orbit is the result of the influence by a large binary companion to the Sun, perhaps thousands of AU away from our central star. Several of these hypothetical companions have been proposed over the last few decades. One of them is named Nemesis and is suggested to be a brown dwarf.

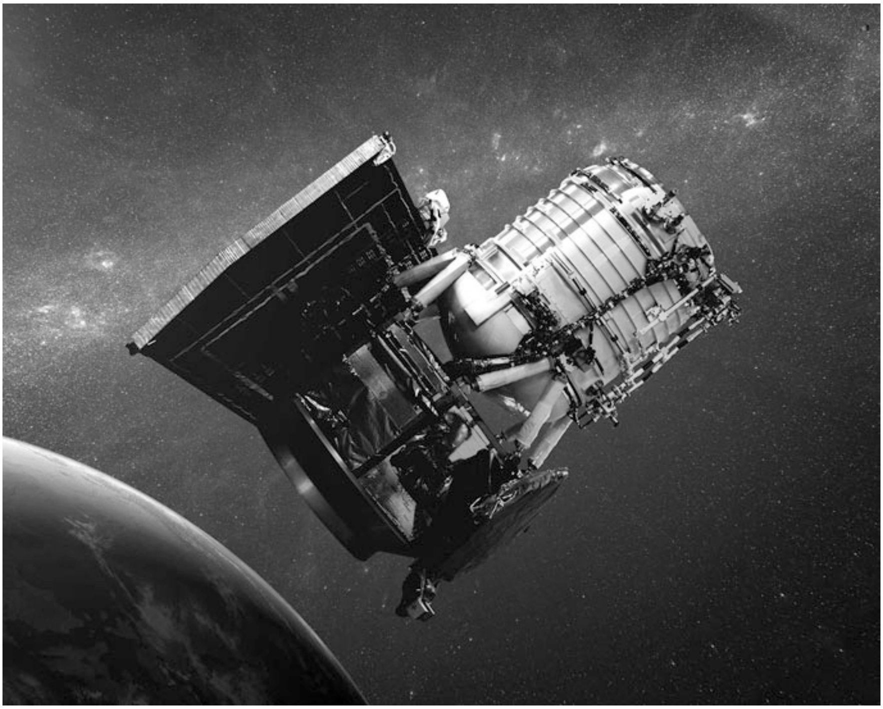


Fig. 4.8 This artist's concept shows the Wide-field Infrared Survey Explorer, or WISE spacecraft, in its orbit around Earth (credit: NASA/JPL-Caltech)

However, no such companion has been found up to now. There are also no indications for its existence. The results of the WISE telescope also do not increase optimism of finding such an object.

If we now assume that a young star from our Sun's birth cluster passed close enough to our solar system, Sedna's orbit could be explained by such an event (Scenario 3). According to this scenario, it is necessary that our Sun formed in an open cluster, e.g. similar to the Pleiades, several stars of this cluster may have gradually dissociated over time. Why is a young star of the Sun's birth cluster necessary? The aphelion of Sedna is not distant enough to be affected by any passing star at their current distances. Hence, the best explanation would be a young star in a dissociating open cluster, which temporarily could get closer to our solar system.

Alessandro Morbidelli and Scott Jay Kenyon have suggested scenario 4 according to which Sedna is of extrasolar origin. It was captured by the Sun from a passing star that was surrounded by a planetary system similar to ours. The more massive the passing star would have been, the more difficult or even impossible it would have been for our Sun to capture Sedna. Hence, scenario 4 would require that the passing star would have about one-twentieth of our Sun's mass, which would imply a brown dwarf.

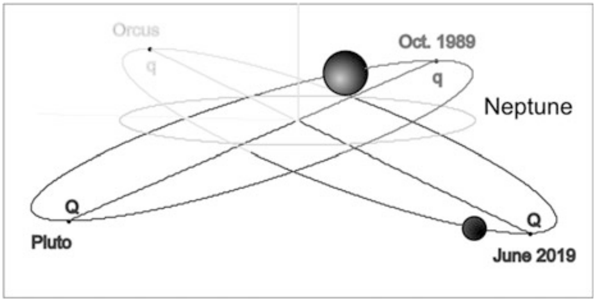


Fig. 4.9 Orbit of (90482) Orcus and Pluto (credit: Eurocommuter/Lilyu, CC BY-SA 3.0)

4.9.2 Orcus

Orcus is a large trans-Neptunian object (TNO) having the official minor planet designation (90482) Orcus. A large moon, Vanth, accompanies it. Michael E. Brown, Chad Trujillo and David Rabinowitz discovered it on February 17, 2004.

Orcus' Orbit

Orcus is a plutino and hence locked in 2:3 mean motion resonance with Neptune similar to the prototype of all plutinos, Pluto. It revolves around the Sun in 247 years. Orcus' orbit is very similar to that of Pluto (see Fig. 4.9). This is partially owed to the aspect that both are locked within the same resonance with Neptune. However, there are further aspects that render the orbits of these two objects similar, e.g. both have a perihelion above the ecliptic plane. Orcus' perihelion is at 30.27 AU (Pluto: 29.66 AU), its aphelion at 48.07 AU (Pluto: 48.87 AU). The semi-major axis of Orcus is at 39.17 AU and that of Pluto ay 39.26 AU. Both have similar eccentricities with $e = 0.227$ (Orcus) and $e = 0.245$ (Pluto). Orcus' orbit is inclined by 20.75° towards the ecliptic plane, Pluto's by 17.15°.

As you can see, the orbits are similar but oriented differently. The mutual resonance with Neptune thereby constrains Pluto and Orcus to remain in opposite phases of their otherwise similar orbits. For this reason, Orcus is sometimes also called "anti-Pluto" and might be a candidate for becoming a dwarf planet. Further analyses are, however, required for that.

Physical Characteristics

Looking at its brightness, Orcus appears to be smaller than Pluto. It is very difficult to obtain detailed and reliable information of Pluto. So you can imagine how difficult it is to observe an even smaller object although it is located at a similar distance from us. By using the Spitzer Space Telescope (2008) and the Herschel Space Telescope (2010), it was possible to determine Orcus diameter to be 850 ± 90 km. This is considerably smaller than Pluto with its approximate 2,300 km. Its albedo is estimated to be in the range of 0.22–0.34.

It is a great happenstance that Orcus is accompanied by a moon. This allows us to determine the mass of the total system. For this we need to measure a few parameters such as the distance of the Orcus from Earth, the orbital period of its moon, Vanth, circulating Orcus and Vanth's distance from Orcus. Then, it is possible to derive the common mass of the double system by employing Newton's law of motion and gravity. All these parameters are known to a more or less reliable degree. Hence, the mass of the Orcus–Vanth system has been determined to be $(6.32 \pm 0.05) \times 10^{20}$ kg. This is much less than Pluto ($(1.305 \pm 0.007) \times 10^{22}$ kg) and Eris ($(1.67 \pm 0.02) \times 10^{22}$ kg).

Now, we know the total mass of the Orcus-Vanth system. How can we determine the individual masses of Orcus and its moon Vanth from that? How this mass is partitioned between Orcus and Vanth depends of their relative sizes. The ratio of masses of Orcus and Vanth is assumed to possibly be anywhere in the range from 1:33 to 1:12. Even with the largest telescopes it is very difficult to resolve the two bodies and measure their relative sizes independently. The sizes thus have to be estimated. Based in assumed albedos of Orcus and Vanth. If we assume Vanth's albedo to be 0.12, which would be typical for small reddish Kuiper belt objects, its diameter would be 378 ± 100 km. If we assume the albedos of Orcus and Vanth to be equal, we arrive at an estimated diameter of 276 ± 17 km. The mass of Vanth thus would be 2–6 $\times 10^{19}$ kg.

Orcus also appears to be different from other TNOs. The first spectroscopic observations in 2004 showed that the visible spectrum of Orcus is flat, i.e. neutral in color, and featureless. Most TNOs and especially KBOs have a reddish color when observed in visible light. Further Infrared observations in 2004 by the European Southern Observatory (ESO) and the Gemini telescope suggested its surface to comprise mixtures of water ice and carbonaceous compounds, such as tholins. Tholins are usually made accountable for the reddish color. They are formed from more simple chemical compounds such as methane when these are exposed to long-term electromagnetic radiation (e.g. ultraviolet radiation coming from the Sun). The water and methane ices can cover no more than 50 % and 30 % of the surface, respectively.

Vanth: A Companion of Orcus

Orcus' moon Vanth was discovered using observations of the Hubble Space Telescope on November 13, 2005. It orbits Orcus in a nearly face-on circular orbit with an eccentricity of about 0.007, and an orbital period of 9.54 days. Vanth orbits only 9030 ± 89 km from Orcus and is too close to Orcus for ground-based spectroscopy to determine the surface composition of the satellite.

Its discoverer Michael Brown and several other astronomers assume that, similar to the Pluto-Charon system, the Orcus–Vanth system is tidally locked. This means that it takes for Vanth just as long to rotate around its own axis as it does to revolve around Orcus. By this, Vanth will always show the same hemisphere towards Orcus. It is the same mechanism we experience with our Moon. We will never see the backside of it without technical means such as spacecraft.

Where does Vanth come from? The first idea we can think of is a origin similar to that of our Moon which is supposed to be the result of a severe collision of the young Earth with another large body. The latter one grazed along Earth and a huge amount of its mantle was blown into space. The debris then accreted to what we call our Moon today.

Is this a likely scenario for Orcus? It seems not. The spectra of Vanth and Orcus, though being extremely difficult to separate, appear to be very different from each other. We can thus assume that it is quite unlikely that Vanth formed from debris of that had been excavated from Orcus. In that case, the composition and by this also the spectra should be similar.

It might be possible though that Vanth had a small Kuiper belt object that was then captured by Orcus and forced into an orbit around it.

Chapter 5
Dwarf Planets

As we have already seen in previous chapters, there are a vast number of objects that form part of our solar system. These objects have been continuously categorized and re-categorized after new scientific findings had emerged. In 1801, when Ceres was discovered in the "gap" between Mars and Jupiter, it was initially seen as the "missing planet" that was supposed to be there following the Titius-Bode-Law. It was therefore assigned a planetary symbol. However, during the following decades more and more objects like Ceres were discovered in that region. Thus, the astronomers decided to re-classify Ceres into the newly created category of asteroids ("star-like objects"). Science made huge progress and with it the need and wish to re-classify Ceres resurfaced again. This time promoting it into the class of dwarf planets. These had recently been introduced by the International Astronomical Union (IAU).

In 2006, the IAU introduced this latest classification according to which basically three classes of objects exist in our solar system: planets, dwarf planets and small solar system bodies. This chapter is devoted to the newly created class of dwarf planets.

One of them, Pluto, has undergone an unprecedented rise and fall in its perception within the astronomical society and the public: once being celebrated as the long sought Planet X and then being demoted several decades later. A separate section of this chapter will deal with him.

5.1 Why a New Class?

The astronomical community, represented in the IAU, considered it necessary to introduce the new class of dwarf planets since more and more objects (mainly trans-Neptunian objects) have been discovered that raised doubts about Pluto's status. These newly discovered objects share many of Pluto's key orbital characteristics. Objects such as Eris or Sedna are not being acknowledged as being planets. Yet,

M. Moltenbrey, *Dawn of Small Worlds*, Astronomers' Universe,
DOI 10.1007/978-3-319-23003-0_5

they are also substantially larger than the other small solar system bodies. So what is a planet? There has never been a clear definition.

During a meeting in Prague in August 2006, the IAU defined three criteria an object has to meet in order to being considered as a planet:

1. The object must be in direct orbit around the Sun.
2. The object must be in hydrostatic equilibrium (i.e. massive enough to be in spherical shape by its own gravitational force).
3. The object must have cleared the neighborhood around its orbit.

The first condition is clear and undisputed, although the interpretation of restricts the term “planet” to objects in our solar system (“orbit around the Sun”) and thereby neglecting exoplanets orbiting distant stars.

The second condition requires the object to have enough mass to form it into a spherical or at least almost spherical shape. Several forces are present in a (planetary) body, the most important being the gravitational force and the internal pressure. The more massive the body is, the higher its internal pressure and the more rounded its shape is. For smaller bodies, such as asteroids, gravity usually does not play the key role but other non-gravitational forces become more dominant. This causes the body to be irregularly shaped. This is the case for most asteroids and the other small solar system bodies.

The third condition for an object being considered a planet, as set out by the IAU, is also the most disputed one. A planet should be dominant enough to be able to clear the neighborhood around its orbit from smaller objects. Planets tend to clear their orbital region over time (over many orbital cycles) by means of gravitational interactions with these bodies. The planet’s gravity will cause a smaller body either to accrete (i.e. collide with the planet), capture it as a satellite, force it into a resonant orbit or disturb it into another orbit. All these events have happened in the solar system and can be observed.

It is the third condition dwarf planets fail to achieve. They are not massive enough to clear their neighborhood. However, as previously stated, this condition is disputed in the astronomical community. Applying a strict interpretation of this condition would also raise problems with the remaining planets. Earth, Mars, Jupiter and Neptune, for example, have not cleared their neighborhood neither. Jupiter and Neptune, for example, have several thousand Trojan asteroids in their neighborhood. We came across these Trojans in Chap. 2.

5.2 How Many Dwarf Planets Do Exist?

Since Pluto also fails to fulfill the third condition, it was demoted to a dwarf planet by decision of the IAU in 2006. However, it was recognized as a prototype of a new category of objects sometimes called plutoids or ice dwarfs, though these terms are not widely accepted. All dwarf planets beyond Neptune’s orbit (trans-Neptunian objects or TNOs) are considered to belong to this group.

Currently, the IAU accepts five official dwarf planets, all but one, Ceres, belonging to the group of plutoids or ice dwarfs. These four ice dwarfs are Pluto, Haumea, Makemake and the one which got the whole ball rolling: Eris. The current only exception is Ceres which unperturbedly orbits the Sun in the main asteroid belt. Many more dwarf planets are suspected beyond Neptune. Several candidates such as 2007 QR_{10}, Quaoar, Sedna, Orcus, 2002 MS_4 and Salacia have been named but not yet been officially recognized. The remaining part of this chapter will deal with the official dwarf planets. The story of Pluto will be discussed in more detail since not many other objects in our solar system (except for our Moon and Mars) have stirred the emotions of people. Some of the candidate dwarf planets are subject in Chap. 4.

5.3 Pluto: Rise and Fall of a Planet

Pluto, the ice planet, the sentinel of the outer solar system has ever since its successful discovery in 1930 been subject of a long lasting, emotionally heated debate about its planetary status in the astronomical community. However, not only there. This small world in spite of its physical unimportance made it into the mass media. Hardly any other object in our solar system besides the main planets has drawn so much public attention. All began with its exciting discovery story.

5.3.1 The Quest for Planet X

Soon after the discovery of Uranus in 1781 by William Herschel, it became clear that something else had to be out there. Perturbations in its orbit were observed suggesting that another big planet should be out there, somewhere beyond Uranus' orbit. The French astronomer Urbain Le Verrier had a closer look at these perturbations in its orbit and mathematically predicted the position of a possible planet that he thought was the reason for causing the perturbations. He gave his calculations including the predicted positions of the potential planet to the German astronomer Johann Gottfried Galle.

On the evening of September 23, 1846, the day when Galle received Urbain's letter, he set out to search for a planet and indeed he found Neptune pretty close to the predicted position (deviating only about 1° from the calculated position).

Yet, regrettably, the discovery of Neptune did not solve all problems. Considering its estimated mass and orbit, the new planet was not possible to be held accountable for explaining the full range of perturbations to Uranus' orbit. Additionally, it became apparent that Neptune's orbit was perturbed itself. Again, the question resurfaced: is there something else out there? Beyond Neptune?

This gave fresh impetus to earlier speculations about planets beyond Uranus. Already before Neptune's discovery, ideas were around in the early nineteenth century that one planet alone could not be enough to cause the observed perturbations of Uranus. In the 1830s, the German astronomer Peter Andreas Hansen (1795–1874), by that time director of the Seeberg Observatory in Gotha/Germany, proposed that a single planet could not adequately explain Uranus' motion and thus suggested that at least two planets had to be behind Uranus.

Furthermore, Jacques Babinet (1794–1872) being in an intellectual dispute with his colleague Le Verrier, objected the latter's calculations after the discovery of Neptune. He claimed that Neptune's observed mass was smaller and its orbit larger than what Le Verrier had predicted. He, thus, proposed that another planet, which he called Hyperion, was lying behind Neptune. This allegedly other planet would have roughly 12 Earth masses, he assumed.

However all these were mere theoretical speculations. Though some astronomers set out to find another planet, Planet X, none of them was successful. It should take about another century until the hunt for Planet X was crowned with the discovery of Pluto in 1930, an achievement closely coupled to the names of Percival Lawrence Lowell (1855–1916) and Clyde Tombaugh (1906–1997).

Lowell, a wealthy Bostonian famous for his engagement in astronomy, founded the Lowell Observatory in Flagstaff, Arizona in 1894 and dedicated his remaining lifetime to astronomy. Besides his personal dedication to find Planet X which he thought to be his lifetime challenge, he also saw it as one way to rebuild his reputation which had suffered from some opinions he had had expressed previously on Martian canals (as first described by Italian astronomer Giovanni Schiaperelli).

So, in 1906, he started an extended search for Planet X where he focused his observations on the ecliptic as he thought Pluto's orbit should be close to it. This was the case for the other planets. Why should it not be similar for the new planet? He invested lots of time and money into this lifetime challenged as he dared to call it.

Unfortunately, today we know that Pluto has a highly eccentric and inclined orbit. At that time Pluto was way too far above the ecliptic to be captured by his survey. Though being unsuccessful, a new search stage began in 1914. Always being open to unconventional—sometimes not scientifically founded speculations—he assumed in 1915 in his "Memoir of a Trans-Neptune Planet", that Planet X had about seven times the mass of Earth and a mean distance of 43 AU from the Sun. He further suggested Planet X would be a large, low-density object with a high albedo, like the gas giants. As a result, it would show a disc with diameter of about 1 arcsecond and an apparent magnitude of 12–13 mag. This would render it enough to be spotted.

Lowell unexpectedly died in 1916. A close friend to him later on stated that failing to find that planet "virtually killed him". With his death, the search came to a halt and did not continue until the end of the 1920s. This delay was primarily caused by a legal feud between the Lowell Observatory and Lowell's widow, Constance. She tried to secure a big part of the million dollar portion of Percival's legacy which Lowell had foreseen for the observatory.

In 1929, Vesto Slipher, the back then director of the Lowell observatory, was impressed by astronomical drawings a 23 year old farm boy from Kansas had made.

Subsequently, he offered him a position at the observatory. Soon after his arrival, Clyde Tombaugh was entrusted with the hunt for Planet X.

Tombaugh started systematically scanning the sky by taking pairs of photos each pair with a couple of weeks in between. This approach was needed in order to identify moving objects. He used a device called "blink comparator" for this. This device was invented to find differences between two photographs inserted into it and by rapidly switching from one to another, i.e. blinking back and forth between two images that had to be taken of the same region of the sky at different points of time.

The idea behind is that a moving object will change its position during the course of time compared to the fixed background stars. You will then be able to see this object blinking at the two distinct positions.

Classic blink comparators are no longer in use today but the basic idea is still applied in some modern image differencing algorithms. Nowadays, most of the process is completely automated relieving the astronomer from the burdensome comparing of thousands of images.

Coming back to the hunt for Planet X, Tombaugh decided not to arbitrarily image each relevant region but to do so near its opposition point, i.e. having an angle of 180° from the Sun. The idea behind is simple but ingenious. At this point, the apparent retrograde motion of objects beyond Earth's orbit is at its fastest. Thus, the moving object will cover a larger distance than at any other point on its orbit. This provides for an easier identification of the object on two consecutively taken photographs since blinking point will be more apart.

So, he started scanning the sky and focused on the entire zodiac, his survey thus had a much broader scope than Lowell's one. On February 18, 1930, his cumbersome work came to an end. He had already looked at about two million stars captured on many photo plates. The search had appeared to be never ending. However, on this day, he discovered a moving object on plates he had taken on January 23 and 29, 1930. This object was merely about 6° from two locations of Planet X, which Lowell had proposed. So, it seemed likely that Planet X had finally been found. After photographic confirmations, it was officially announced on March 13, 1930 that Planet X had been discovered.

Soon after the announcement, it became apparent that there had been several prediscoveries, i.e. Planet X had been captured on photo plates without being recognized as it. Tragically, Planet X, was later on found on photo plates taken at the Lowell Observatory by Percival himself in March and April 1915.

After the discovery, the Lowell Observatory called for name suggestions as with all astronomical discoveries, the right to name lies with the discoverer. By far more than 1000 suggestions were received from all over the world. Three names made it on a short list among which the employees of the observatory could vote on: Minerva, Cronos and Pluto. Pluto, originally proposed by 11-year old Venetia Burney from Oxford/England, received the most votes as the other two names soon dropped out of the game. The name Minerva had already been given to an asteroid (93 Minerva discovered in 1867). Cronos had been proposed by a quite unpopular astronomer, Thomas Jefferson Jackson See (1866–1962). Thus, Pluto was officially named on March 23, 1930. It was decided to give him as astronomical symbol a stylized P + L, partly in recognition of Perceval Lowell.

5.3.2 The Fall of Planet Pluto

Soon after, a long lasting debate began about Pluto's planetary status. Was it really the long sought Planet X? It was so different from what the astronomical community had expected. It was not even close to any resemblance with the ice giants, Uranus and Neptune. Lowell had expected an object of about seven times the mass of Earth, something more of the size of Uranus. Was this faint little dot lacking any resolvable disc really the expected Planet X? It appeared to be too small and doubts increased and should never disappear. The debate was further nourished since estimates of its mass were revised downwards throughout the following decades.

The Problem of Pluto's Mass

In 1931, shortly after its discovery, it was estimated that Pluto's mass roughly resembled that of Earth. Lowell's estimation had been soon dropped due to Pluto's faintness. Then, in 1948, new estimations seemed to suggest about the mass of Mars. Yet again, in 1976, when astronomers more closely considered Pluto's albedo, it was further downsized to about 1 % of that of Earth. It was only in 1978 when Pluto's moon Charon was discovered which made it possible to actually measure its mass to be about 0.2 % of Earth's mass. This was so much lower than anything expected previously.

Again, was this really a planet? How did it come that Pluto's mass had been drastically revised continuously after its discovery? No other planet has undergone such a development. This lies in the very specific nature of Pluto: for a long period of time it was just too faint and small to make reasonable observations. It was simply impossible to resolve Pluto's planetary disc in ground-based telescopes. As science and technology developed, it became possible to use more elaborated methods and models like considering its albedo and its apparent magnitude. We have touched this in Chap. 2.

Finally, with the help of Charon (see Fig. 5.1), Pluto's almost equal sized companion, it became possible to make direct measurements. The idea behind this type of measurement is not difficult to understand and only requires basic math. First, we need to measure a few parameters such as the distance to Pluto from Earth, the orbital period of Charon circulating Pluto, and Charon's distance from Pluto. Then, it is possible to derive the common mass of the double system by employing Newton's law of motion and gravity. Usually, the moon's mass is negligible as it is by far small than that of the planet. Though, the situation is exceptional with Pluto. Charon is quite large compared to Pluto (about 1,212 km vs. 2,368 km).

Consequently, Charon does not exactly orbit around Pluto but they influence each other and orbit around a common point, the center of mass or barycenter (see

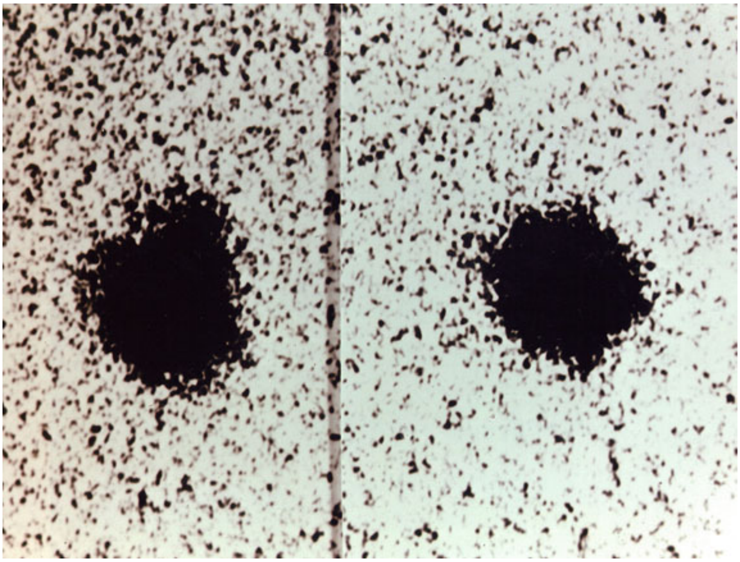


Fig. 5.1 Discovery images of Pluto's moon Charon taken in 1978. You can see the buldge of Charon on the left part (credit: U.S. Naval Observatory; Public Domain)

Fig. 5.2). Observations made it possible to determine that this point is about eight times closer to Pluto than to Charon. We can derive from this that Pluto is about eight times more massive than its companion.

New Objects Emerge and Final Death Knells

Looking at the development of estimation of Pluto's mass, it became more and more clear that such a small object could not be responsible for the perturbations of Uranus' and Neptune's orbits. So, the search for another trans-Neptunian planet started again. Although not being successful in that a large planet was found, new Pluto-like objects were discovered beyond Neptune.

In 1992, an object called 1992 QB_1 was discovered by David Jewitt and Jane Luu at the Mauna-Kea observatory on Hawaii. With its mean diameter of about 167 km it became the third largest body (after Pluto and Charon) in the trans-Neptunian zone by that time. The year 2005 was a remarkable and decisive one marking a turning point in the research of small solar system bodies. The events of that year triggered shock waves that finally should bring the old classification scheme to fall.

The American astronomers Michael Brown, Chad Trujillo and David Rabinowitz discovered Eris, an object that was initially thought to have a mass larger than that Pluto. On the same day, just a few hours before, a group around Spanish astronomer José Luis Ortiz Moreno announced the discovery of another large body in the Kuiper belt, Haumea. Later, the Brown and his colleagues published the discovery of another object, Makemake. The astonished public was, thus, informed about the discovery of three Pluto-like objects within a few days. All

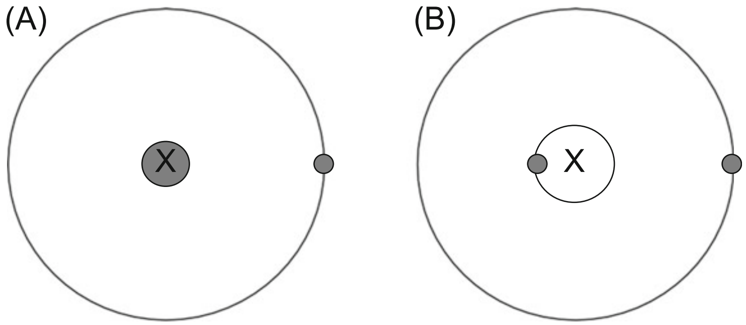


Fig. 5.2 Illustration of the barycenter (center of mass) in two scenarios: (**a**) the primary object in the center has a clearly dominant mass compared to the orbiting moon. Hence the barycenter lies close to the central objects core; (**b**) differences in masses of primary object and moon is lower. The barycenter lies outside of the primary object (similar to the Pluto-Charon system)

of them were more or less resembling Pluto's size. This in the end should sound the death knell for Pluto's further existence as a planet.

5.3.3 Rebirth as a Dwarf Planet

At least the events of 2005 led to renewed and thorough discussion about Pluto's planetary status in particular and the definition of what a planet is in general. The discussion culminated in the 2006 IAU definition of a planet. Following this still controversial definition Pluto fulfills all but the last criteria. It orbits the sun, is supposed to be in hydrostatic equilibrium but did not clear its neighborhood. Though no longer being considered a planet, Pluto maintained a special role. It became the prototype of a class called plutoids or ice dwarfs which stands for a trans-Neptunian object being large enough to have a round shape.

5.4 Pluto and Its System in Detail

Having looked at its discovery story and its demise, it is time to answer the question what we actually know about Pluto.

As we have seen previously in this chapter, it is difficult to observe and measure Pluto due to its large distance, its small size and faintness. For a long period of time, it was only possible to indirectly derive information about its characteristics. This only changed with the Hubble Space Telescope which took the first surface map of Pluto in 1994 (and again in 2003). However, the view on Pluto is not sharp enough. So it is not possible to see any detailed surface structures such as craters, mountains or canyons as we are used to observe on other "planets". The maps show regions on

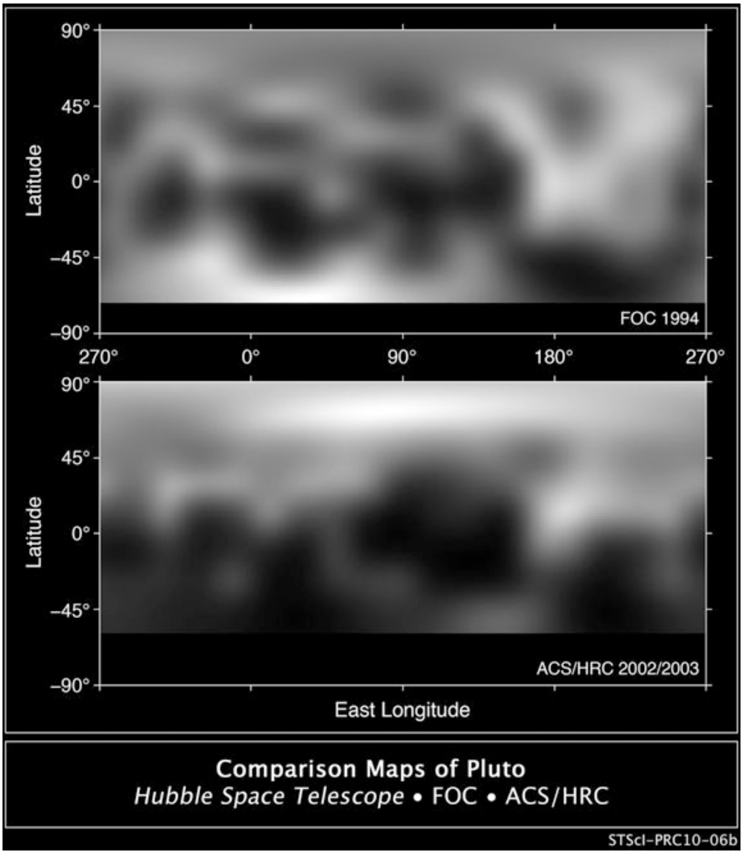


Fig. 5.3 These are two Hubble photo maps of the dwarf planet Pluto, as seen in 1994 and 2002–2003. The *white areas* are surface frost, and the *dark areas* are a carbon-rich residue caused by sunlight breaking up methane that is present on Pluto's surface. A comparison of the maps shows that Pluto's brightness has changed between 1994 and 2003 (credit: NASA, ESA, and M. Buie (Southwest Research Institute))

its surface that differ from each other. There are white, frosty regions and dark regions with carbon residue caused by the sunlight breaking up methane that is available on Pluto's surface (see Fig. 5.3).

Since Earth-bound observations are limited in their capabilities of revealing the secrets of this small icy world, space probes are the means of desire. First plans to visit Pluto arose in the early 1990s in form of the "Pluto Kuiper Express". The involved scientists urged for an early launch in order to arrive at Pluto before its atmosphere froze out again. Pluto had been in its perihelion in 1989 and was now departing to the fringe of the outer solar system again. The increasing distance to the sun thereby cooled down its atmosphere. The farther Pluto is away from the Sun, the cooler it gets and at a certain point of time, the volatiles forming its atmosphere will freeze out and fall back to its surface leaving behind no atmosphere.

Financial issues and problems with the launch vehicle led to a cancellation of this NASA project in 2000. A new project was initiated in 2003: New Horizons. The probe was successfully launched on January 19, 2006 and passes Pluto in July 2015 in a currently planned distance of 9,600 km and Charon only being 27,000 km away. Detailed global surface maps of Pluto and Charon are expected having a resolution of about 25 m per pixel. Detailed studies of Pluto's atmosphere will follow.

5.4.1 Orbit: Where Is It?

Pluto was long believed to reside on the far fringe of our solar system. It is not very long ago that we understood that there is still much more existing behind it.

After its discovery, astronomers were surprised by its orbital characteristics. They were so different from what was known about the other planets so as to cast questions about its origin and nature. Pluto revolves around the Sun in 247.68 years in a highly eccentric ($e = 0.2488$) and inclined orbit (about 17°) whereas the other planets follow near circular orbits around the Sun and move in more or less the plane, the ecliptic (see Fig. 5.4).

Due to its highly eccentric orbit, Pluto sometimes is closer to the Sun as Neptune. This was also substantially different from the other planets. No two other planets cross their orbits. Pluto reaches its perihelion, its closest point to the sun, at about 29.69 AU (last time in 1989). It can be as far as about 49.31 AU distant from the Sun at its aphelion.

So, Pluto crosses Neptune's orbit? Is it then not probable that the two will end up in a big inferno someday by colliding with each other? The answer is simple: no! There are several reasons for that. Eventually, Pluto only appears to cross Neptune's orbit depending on the perspective. The two orbits actually do not intersect. When Pluto is closest to the Sun and consequently closest to Neptune's orbit, it is also with a distance of about 8 AU farthest above Neptune's path. Furthermore, Pluto's orbit is stabilized due to its lock into a highly stable 2:3 mean motion resonance with Neptune which means that for every two orbits Pluto makes around the Sun, Neptune makes three. Each time, the two bodies return to their initial position and the cycle repeats.

What does this actually mean? Just imagine, Pluto is in its first perihelion, i.e. closest to the Sun. Then, at this point of time, Neptune is about 50° behind Pluto. During Pluto's second perihelion, Neptune will have made about 1.5 orbits around the Sun and hence be about 50° ahead of Pluto. In reality, Pluto and Neptune never come really close to each other. Their minimum distance is over 17 AU. There are also a few further factors, besides the dominant 2:3 resonance, that contribute to stabilize Pluto's orbit.

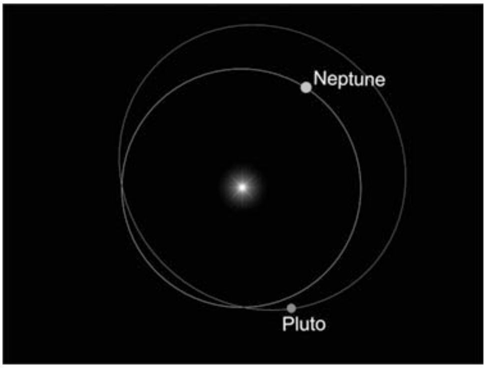


Fig. 5.4 Pluto's orbit compared to Neptune (credit: NASA/JPL)

5.4.2 Origin: Where Did It Come from?

The question remains: where did Pluto come from? Several early hypotheses were circulating. One very popular, in particular among the public, was that Pluto once had been a moon of Neptune. It was then, someday, knocked out of its orbit around Neptune by the ice giants largest current moon Triton. One argument brought forward in favor of this theory is the peculiar rotation of Pluto, its rotation axis being heavily tilted compared to other planets. Yet, this hypothesis had never convinced the majority of astronomers for a simple reason: Pluto, as we have just seen, never comes close enough to Neptune on its orbit. So, how should this knock out have happened when the two bodies are always that far apart? This required multiple large perturbations caused by objects other than Neptune. No such objects are known and even if they existed such a chain of events is considered to be more than unlikely.

Today, we understand that Pluto is merely one, though being one of the largest, of thousands of smaller bodies in the Kuiper belt. Bodies being in a 2:3 resonance with Neptune have been called plutinos, named after Pluto their largest and most prominent member. Just as a side note: you should pay attention not confuse the term "plutino" with plutoid since plutinos are bodies being in the same resonance with Neptune as Pluto whereas plutoid is the term for dwarf planets or ice dwarfs beyond Neptune's orbit.

Pluto is a Kuiper belt object (KBO) and hence part of the Kuiper belt. In this belt, as we have seen, a vast number of small icy objects is located which are considered to be remainders of the original protoplanetary disc around the Sun. Most likely, Pluto is thus a planetesimal, that failed to accrete to a full planet. Pluto is believed to have come to its current position as the other KBOs due to a sudden migration undergone by Neptune in the early stages of the solar system. During this migration Neptune is believed to have drawn near the objects of the proto-Kuiper belt. The encounter with the Neptune had three possible effects: first, an object was caught by the gas giant and set in orbit around it. This appears to be the case for Neptune's largest moon, Triton which shares many characteristics with Pluto and other KBOs.

Secondly, bodies were locked into resonances (like Pluto and the plutinos). Thirdly, bodies were expelled into chaotic orbits, some of them now forming part of what is called the scatted disc.

5.4.3 Rotation

A Plutonian day lasts about 6.39 Earth days. Quite long compared to other larger bodies in our solar system. Its largest moon, Charon, plays a key role in that. Due to its size and mass, Charon's orbit is bound to Pluto's rotation by their mutual tidal forces, i.e. the revolution period of Charon equals the rotation period of Pluto. Both bodies are mutually tidally locked.

Yet not only the length of a Plutonian day is different, its rotation is also peculiar in that it is a retrograde motion, i.e. it is rotating on its side in relation to its orbital plane.

Not many major bodies in the solar system do have such a retrograde rotation. Among them are Uranus and Venus. The axial tilt is about 122° which is huge compared to about 23° of our Earth. This causes extreme seasonal variations. At times of its solstices, about one fourth of its surface is covered by permanent daylight whereas another fourth remains in darkness.

5.4.4 Physical Characteristics

Putting aside orbital and rotational characteristics of Pluto which can be directly observed, measured or derived, many physical characteristics of the dwarf planet remain veiled. Thus, astronomers resorted to indirect measures such as using occultations of stars and other means. Many aspects, however, are still speculative and await their confirmation upon arrival of the spacecraft "New Horizons" in 2015. One thing, however, is clear. Pluto is a very strange place to be. Its surface may be more comet-like than what has been expected following the decades of its discovery.

For a very long period of time, it was not possible to resolve Pluto's planetary disc in telescopes. The disc has an angular diameter of only 0.11 arcseconds at its best. Really tiny compared to other bodies in our solar system. Mars, for example, can have an angular diameter of 25 arcseconds and Neptune 2.3 arcseconds. Even dwarf planet Ceres has a larger diameter of about 0.8 arcseconds although it is clearly smaller in size. However, this effect is owed to the fact that Ceres is much closer to us than Pluto.

So what does that actually mean? Just imagine you would try to see a pinhead from a distance of about 5 km. This is a quite difficult task. You should also not to forget the problems related to the turbulent atmosphere ground-based telescopes have to tackle. These turbulences cause stars and other objects to twinkle and drastically reduce the theoretically available angular resolution of telescopes.

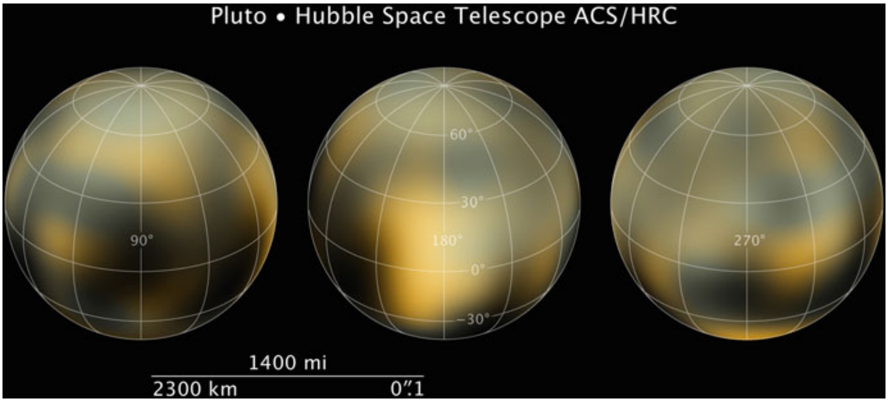


Fig. 5.5 The changing faces of Pluto as captured by the Hubble Space Telescope (credit: NASA, ESA, and M. Buie (Southwest Research Institute))

Only with advent of the Hubble Space Telescope (HST), it became possible to image Pluto's disc directly. Before this, astronomers had to rely on indirect observations. They developed some fancy techniques to obtain information, techniques that are also used for other objects in our solar system and beyond.

One way to determine surface details was by deriving brightness maps from eclipses with Charon. Each time Charon passes in front of Pluto the total brightness of both bodies varies because Charon absorbs some of the reflected light that comes from Pluto. We can then use these variations to determine the brightness distribution on the Plutonian surface: when Charon eclipses a brighter area on Pluto, there will be a bigger drop in the total brightness than in the case where a dark area is covered. These derived maps, however, are very coarse grain and do not provide much detail (Fig. 5.5).

A huge step forward was made with the Hubble Space Telescope. Its angular resolution of about 0.4 arcseconds and the absence of a turbulent atmosphere while observing made it possible to determine more detailed surface maps for the first time ever. These maps (as depicted in Fig. 5.3 and 5.5, taken in 1996 and 2002, show details of several hundred kilometers in size including polar regions, as well as bright and dark areas.

By employing spectroscopic analysis it was further possible to find out more about Pluto's composition and to derive theories about the nature and origin of the bright and dark areas that were seen on the maps. These analyses revealed that Pluto's surface is mainly composed of more than 98 % nitrogen ice with traces of methane and carbon monoxide. It is assumed that the bright areas are frost and the dark ones carbon rich residues caused by sunlight breaking up methane which is present on its surface.

Observations made with the Hubble Space Telescope suggest that Pluto has a density of about 2 g/cm^3 which would imply, taking common models and the spectroscopically measured chemical material as a basis, that it consists of a

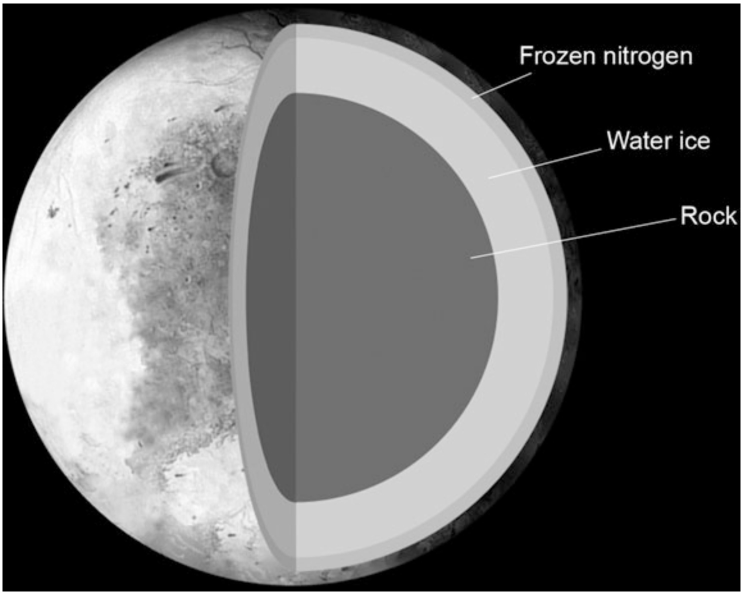


Fig. 5.6 Internal structure of Pluto (credit: Jcpag2012; CC BY-SA 4.0)

mixture of about 70 % rock and about 30 % water ice. It, thus, resembles quite well Triton, Neptune's largest moon, and thereby substantiates the suspicion of their common nature and origin as (proto-)Kuiper belt objects. Pluto is also similar reddish in color as Triton and one of the most contrastive bodies in our solar system.

Scientists believe that Pluto's internal structure is differentiated with a clear separation into a solid rocky core of about 1700 km in diameter surrounded by a mantle of water ice and covered by a layer of frozen methane and nitrogen (see Fig. 5.6). A more homogeneous, non-differentiated structure is hardly conceivable. Radioactive decay, present in all larger bodies in the solar system, would heat up the ices enough for the rock to separate from them. However, owing the presence of radioactive decay, some scientists assume that there might be a subsurface ocean of liquid water ice between 100 and 180 km thick at the boundary of core and mantle.

Independent from the existence of such an ocean, Pluto's surface is definitely one of the coldest in the solar system. In 2006, a temperature of about −230 °C was measured by scientists using the Submillimeter Array, a network of radio telescopes located in Hawaii. This came as a big surprise as this is about 10 °C warmer than was otherwise expected. The involved scientists also managed to separate the thermal emissions of Pluto and Charon for the first time. Charon, contrary to Pluto, showed the expected temperature of −220 °C.

So, how does it come that Pluto is colder than expected? The cause for this is to be found in Pluto's very thin atmosphere. Its highly eccentric orbit, varying between roughly 30 and 50 AU, has a huge influence on its atmosphere and surface temperature. On the one hand, when Pluto is close to its perihelion, its surface heats up and the nitrogen ice and methane on it sublimate, i.e. turns into gas and goes into and forms the atmosphere. On the other hand, when Pluto departs from the Sun, the farther the distance, the colder the atmosphere is. The gases in it then tend to freeze out and fall down on the surface. This is sometimes called an anti-greenhouse effect.

Usually, on many planets like our Earth, we have a greenhouse effect where the incidence of sunlight on the surface is absorbed and used to heat up the surface. In the case of Pluto, however, it is the other way round: the energy coming from the sunlight is not absorbed but used to convert the ice into gas. This effect cools down the planet. You can best understand this effect by considering an analogous process. Think about humans sweating. The evaporation of sweat from the skin (like the gas leaving the surface) has a cooling effect because the evaporated sweat carries the body's excess heat away.

Getting this insight is not a matter of course. Until the 1980s it was not even sure whether Pluto had an atmosphere at all. A tiny, icy rock like Pluto was hardly believed to be able to have an atmosphere? So the key question was whether it was massive enough to maintain an atmosphere? Its presence was only definitely detected by the Kuiper Airborne Observatory in 1988. The scientists observed occultations of stars by Pluto. During the course of such an event Pluto passes in front of a star and thereby causing a decrease in brightness of the occulted star. Two scenarios were possible. First, if Pluto had no atmosphere and moved in front of a star, the drop of brightness would happen abruptly. Secondly, in case of presence of an atmosphere, it would gradually drop.

The latter was the case with Pluto. Additionally, the rate at which the brightness drops gives hints on the atmospheric pressure present. The findings suggested about 0.15 pascal which is only about 1/700,000 that of Earth. By observing further occultations this value could be further refined. But Pluto's atmospheric pressure is still on debate: NASA suggests about 0.3 pascal, whereas the European Southern Observatory proposes 1.5 pascal.

Using spectroscopy, it was possible to determine the composition of Pluto's atmosphere. It consists of a very thin envelope of nitrogen, methane and carbon monoxide. In 2006, traces of ethane were detected in it. By far more details are expected upon arrival of the New Horizons spacecraft.

However, one thing that is already now clear is the existence of a temperature inversion due to the presence of methane, a very powerful greenhouse gas. This causes the temperature to increase by about 3 °C per kilometer altitude difference. Thus, its upper atmosphere with its −170 °C is substantially warmer than the surface with its −230 °C.

5.4.5 The Plutonian System

Up to now, five moons of Pluto are known, the most famous and largest being Charon. Charon was discovered by James Christy at the US Naval Observatory Flagstaff Station in 1978. The moons Nix and Hydra followed in 2005. Kerberos

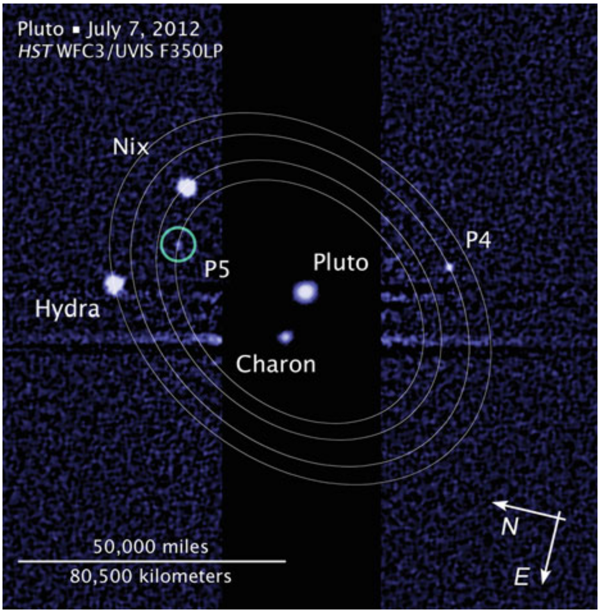


Fig. 5.7 This image, taken by NASA's Hubble Space Telescope, shows five moons orbiting the distant, icy dwarf planet Pluto. In this image, Kerberos and Styx were not given official names yet (credit: NASA, ESA, and L. Frattare (STScI))

was discovered on 2011 and the up to now last and smallest moon, Styx, was found in 2012 (see Fig. 5.7). The moons are unusually close to Pluto compared with other bodies in the solar system and their respective moons.

We have already seen that the Pluto-Charon system is very special and may be one of the very few real binary systems in our solar system. Both Pluto and Charon are rotating around their common barycenter which is true for all systems consisting of at least two bodies (see Fig. 5.2). It is also the case for the Earth-Moon system. Our planet and Moon orbit around their barycenter. However, due to the large difference in masses, the barycenter of the Earth-Moon system lies very close to the center of the Earth (scenario A of Fig. 5.2). This is different with Pluto and Charon. Charon's mass is believed to be at about 10 % of Pluto's mass which is considerably high for such a constellation. Hence, the barycenter of the system is not almost identical with Pluto's center but actually lies above the surface of the dwarf planet (see situation B of Fig. 5.2). It is their special relation that makes Pluto and Charon sometimes be called a double dwarf planet system.

In addition, both bodies are tidally locked to each other, i.e., Charon is locked to Pluto and vice versa. What does this mean? You probably have seen the effects of tidal locking before without recognizing it. Our Moon, for example, is tidally locked to Earth. The effect of such a locking is that our moon always shows the same hemisphere to us. This is also the case with Charon. It always presents the same hemisphere to Pluto. However, there is a difference to the Earth-Moon system. In addition, to Charon being locked to Pluto, Pluto is also locked to Charon. This means that not only Charon is always showing the same hemisphere but also Pluto. An observer on Charon will always be given the same view of Pluto. In

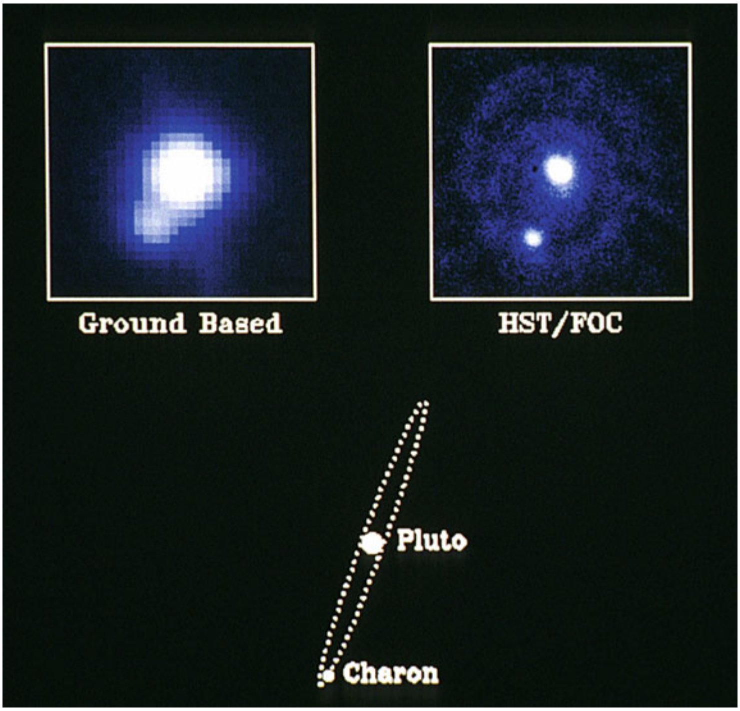


Fig. 5.8 Image showing Pluto and Charon resolved by the Hubble Space Telescope in 1990 (credit: NASA, ESA, and STScI)

effect, these two celestial bodies revolve around each other as if joined with a rod connecting two opposite points on their surfaces.

Charon

What do we actually know about Pluto's largest moon? It is very difficult to observe Charon. Only in the 1990s, the Hubble Space Telescope (HST) was able to resolve Pluto and Charon into their separate discs (see Fig. 5.8). Before this, they simply appeared on images as a smeared irregular smudge like on Charon's discovery image.

It was almost impossible to derive any useful information from those old images other than the existence of the moon and its rotation period of 6.387 days. Nowadays, modern adaptive optics makes it possible to resolve the two bodies via ground-based telescopes. Yet, the problem remains that Pluto and Charon are so distant as not to allow determining any details about their surfaces via direct imaging. Even New Horizons which is currently en route to Pluto is not able to take images with resolved surface features (see Fig. 5.9). This will, however, change until July 2015.

Furthermore, spectroscopic analysis is difficult since their average distance of 19,570 km to each other is small. However, a trick helped. As the two bodies orbit

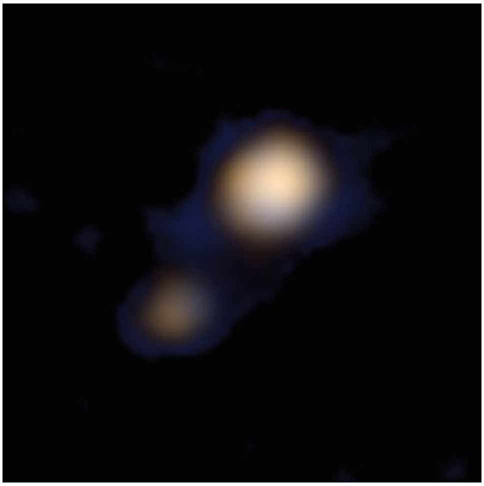


Fig. 5.9 Pluto and Charon—first color image from the New Horizons mission taken on April 9, 2015 (credit: NASA/ Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute)

each other, they will regularly eclipse each other. This allowed astronomers to first take spectra of Pluto and then of the combined system of Pluto and Charon. Then, in the next step, the spectra of Pluto were subtracted from the combined spectra. The results were the spectra of Charon which allowed deriving some information about its surface composition.

Other than Pluto, Charon is not covered by nitrogen and methane ices. Instead it is largely covered with less volatile water ice. It appears to lack an atmosphere similar to that of Pluto. Whether Charon has an atmosphere at all remains an open question. Analysis carried out by the spacecraft New Horizons will deliver further insight during its flyby of Pluto and Charon in July 2015.

The high abundance of water ice further indicates that Charon is an icy and less rocky body compared to Pluto. This finding is also supported by the moons estimated density of 1.65 ± 0.06 g/cm^3 which is considerably lower than that of Pluto (2.03 ± 0.06 g/cm^3). This suggests an internal composition of 55 ± 5 % rock and about 45 % of water ice. Pluto is supposed to consist of about 70 % rock.

Charon's internal composition is thus in line with the currently accepted theory of its origin. It is assumed that in the early phase of the solar system about 4.5 billion years ago a giant impact occurred on Pluto. A large Kuiper belt object collided with Pluto and destroyed itself and parts of Pluto's outer mantle which primarily consists of water ice. The debris of the outer mantle was blown into space and later Charon coalesced from that debris.

There exist two diverging theories about Charon's internal structure. Either Charon may be differentiated into a mantle of ice and a core of rock as is the case for Pluto or the moon may simply be formed from a homogenous mixture of rock and ice all the way through. Again, New Horizons is expected to deliver the answer to this problem.

Furthermore, photometric analysis has revealed that Charon's albedo is not equal over its whole surface. The moon appears to be brighter, having a higher albedo at the equator and darker at the poles. The reasons for this are not yet known.

Nix and Hydra

Two additional moons were found on images of the Hubble Space Telescope (HST) in June 2005. They were subsequently named Nix and Hydra. Both orbit Pluto in its equatorial plane and thus follow the example of their big brother Charon. A characteristic of the Plutonian system is that Pluto's moons orbit it on near-circular orbits. This is true for Charon. Nix and Hydra are no exception. Although Hydra's orbit is slightly more eccentric than Nix' ($e_{Hydra} = 0.0051$ vs. $e_{Nix} = 0.0030$ vs. $e_{Charon} = 0.0$). It takes Nix 24.9 days to orbit Pluto once, whereas the Hydra's orbital period is 38.2 days.

Astronomers have not been able to directly measure Nix and Hydras's size. Hope lies, again, with New Horizons. Nix diameter was estimated, taking into account its distance to Earth and assumed albedo values, to be in the range of 46–37 km and Hydra's diameter to be somewhere between 61 and 167 km. Nix is slightly fainter than Hydra, suggesting that it is somewhat smaller in size. At the time of their discovery Hydra was about 25 % brighter which may suggest that it is about 10 % larger than Nix.

The origin of these two small moons is considered to be the same event that formed Charon, the giant impact which Pluto suffered in its young phase.

Kerberos and Styx

Pluto's fourth moon was discovered on June 28, 2011 on images taken by the Hubble Space Telescope. It orbits Pluto in a distance of about 59,000 km on a circular path in about 32.1 days, following the tradition of the other moons. Kerberos is located between the larger moons Nix and Hydra. Kerberos is the second smallest moon of Pluto with an estimated diameter of 23–34 km. The diameter is again estimated, taking into account its distance to Earth and assumed albedo values in the range of 0.04 (like the darkest known Kuiper belt objects) and 0.35 (roughly resembling Charon's albedo).

The smallest known moon of Pluto is called Styx and was discovered on June 26, 2012 again on images of the Hubble Space Telescope. Its diameter is supposed to be 10–25 km. Styx orbits Pluto in a distance of 42,413 km in 20.1617 days. Its orbit hence lies between the orbits of Charon and Nix.

Because of its small size, Styx is likely to be irregular in shape. Kerberos and Styx are both thought to have formed from the debris of the aforementioned impact event, which would have led to losses of the more volatile ices, such as those of nitrogen and methane, in the composition of the impactors. This process is expected to have created a body consisting mainly of water ice.

5.5 The Remaining Dwarf Planets

Until now we had a closer look at the special case of Pluto. But what about the remaining dwarf planets? These will be the subject of the following sections.

5.5.1 Ceres: An Isolated Dwarf

The dwarf planet Ceres has an eventful history. We have already touched it when we dealt with asteroids as it was the first asteroid to be discovered. Although by the time of its discovery it had been initially classified as a planet (Fig. 5.10).

Ceres is the only dwarf planet in the inner solar system. All other dwarf planets we know about, such as Pluto, Haumea and Makemake, have their orbits in the trans-Neptunian region, i.e. beyond the orbit of the gas giant Neptune. This makes it special and stimulates theories about its origin.

Furthermore, Ceres is not only the first asteroid that was discovered, it is also the by far largest and most massive object in the main asteroid belt.

The History of Its Discovery

We have seen it before, but we will go a bit more into details here as we do not refer to the general discovery of asteroids but the deal with the special case of Ceres in this section.

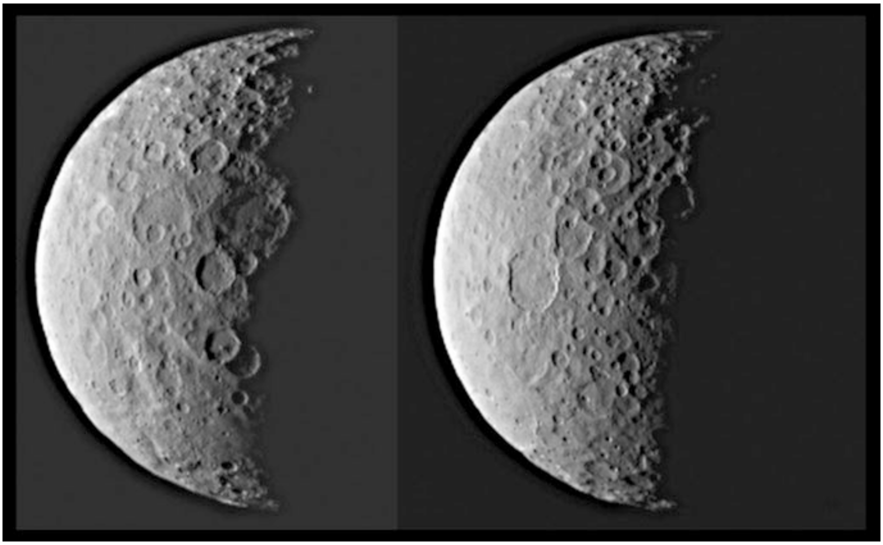


Fig. 5.10 Ceres in half shadow. NASA's Dawn spacecraft took these images of dwarf planet Ceres from a distance of about 40,000 km on Feb. 25, 2015 (credit: NASA/JPL-Caltech/UCLA/MPS/DLR/IDA)

Ceres was discovered by Giuseppe Piazzi on January 1, 1801. By that time the first international sky survey was going on carried out by the so-called “Himmelspolizey”. It was their declared goal to find the “lost” planet between the orbits of Mars and Jupiter that was suggested to exist there according to the Titus-Bode law.

Piazzi was not a member of this first pan-European adventure, but he was the first to be successful and thereby out rivalled the Himmelspolizey.

He discovered, as he wrote down later, a slowly moving star-like object which he first thought of to be a comet. He then subsequently observed it for 24 nights. The last time before he fell sick was on the night of February 11, 1801. Due to his illness he had to stop his observations.

However, he had already announced his discovery on January 24, 1801 in a letter to Elert Bode. He described his new object as a comet but also mentioned that “since its movement is so slow and rather uniform, it has occurred to me several times that it might be something better than a comet”. He was not aware that he indeed had discovered something completely new and it should take some time before the scientific community was really sure about that. The complete discovery story is like a crime story or thriller in which the suspect, Ceres, was almost caught but then was able to escape from its pursuers until it got finally nailed down.

Lots of time was lost as the notification of Piazzi’s discovery was only published in September 1801. Yet, during this period of time, Ceres’ apparent position had changed. The dominating factor to this was Earth’s own orbital motion during that time. In addition, as we know today, Ceres was by then too close to the Sun and hence almost impossible to observe. Thus, no confirming observation of Piazzi’s strange new object was possible.

Another problematic aspect was that rediscovering an object after such a long time was and still is very difficult. The methods used to predict an object’s exact position that were applied by that time were relatively imprecise and not very elaborated.

However, is it not sufficient to know the objects approximate position? Nowadays, this might work in some cases. But think back into the early nineteenth century when no photographic means and hence no blink comparators, as e.g. used by Clyde Tombaugh to discover Pluto, were available. Observations were restricted to the visual observations through a telescope. Just imagine how difficult it may be to find a faint, star-like dot of light somewhere in the sky that moves very slowly compared to the background stars.

Luckily, the young mathematician Carl Friedrich Gauss developed a new method for orbital determination, the method of least squares, which only required a few recorded positions and was nevertheless quite accurate in its predictions. Consequently, Gauss was able to predict Ceres’ position and indeed the two astronomers von Zach, who founded the Himmelspolizey, and Heinrich Olbers were able to rediscover Piazzi’s object near the predicted position on December 31, 1801, thus about 1 year after its first discovery. The object was then named after the Roman goddess of agriculture, Ceres.

A Question of Size and Classification

Ceres was found again. But was it the missing planet? Even in the largest telescopes available by that time a planetary disc could not be resolved. How large was it?

We have already discussed the problem relating to the determination of an object's size in Chap. 2 and seen that this task is more difficult than one would expect. In particular during these early days determining the size of a small object was error-prone. Hence, a huge variance existed in the postulated diameters of Ceres.

William Herschel proposed a diameter of 260 km in 1802 and thereby clearly underestimated Ceres' size. At the other end, the astronomer Johann Hieronymus Schröter determined the diameter to be 2613 km, clearly overestimating its size. Today we know quite well Ceres' size. Its diameter at the pole is about 909 ± 3 km and at the equator 975 ± 3 km which clearly makes it the largest object in the main asteroid belt. It is about three times more massive as the second placed asteroid Vesta and about two times larger than the second-largest asteroid Pallas.

In any case, independent from which diameter was considered, Ceres appeared to be much smaller than any other planet in the solar system that was known by that time, i.e., Mercury to Uranus. This even hold for th so far smallest known planet, Mercury with a diameter of roughly 4,879 km.

Since its discovery, Ceres has undergone several re-classifications. Right after its discovery, Bode thought that Ceres was the missing planet that was postulated to orbit between Mars and Jupiter. Consequently, as this opinion was shared by a majority of the scientific community, Ceres was initially classified as a planet like all the other objects that should follow which we consider to be asteroids today.

Yet, with the increasing number of discoveries, the number of planets in our solar system grew rapidly. All these new planets shared similar orbital characteristics and were sometimes really small. Not all of them could be planets. They were just so different also with respect to the orbital inclinations towards the ecliptic plane making them distinct from the terrestrial planets and gas giants. It, thus, became apparent that Ceres was merely the first of a new class of objects but not a planet.

Herschel coined the term "asteroid" for them, meaning star-like object. He reasoned this by stating that "they resemble small stars so much as hardly to be distinguished from them, even by very good telescopes." Following the new classification, Ceres was reclassified as an asteroid bearing from then on the designation (1) Ceres.

Although a difference was acknowledged between an asteroid and a planet, no formal definition of what a planet actually is was made. It lasted until 2006 before this happened. That year, the International Astronomical Union (IAU) defined planets and distinguished them from other bodies in the solar system. We have come across this definition at the beginning of this chapter.

Ceres complies with the first two criteria but fails to comply with the third one as it is not dominant enough to have cleared its neighborhood. Therefore, Ceres was re-classified into the newly created group of dwarf planets.

What Is Ceres' Mass?

The dwarf planet Ceres is not only large but also massive with approximately 9.4×10^{20} kg. This roughly corresponds to about one-third of the main asteroid belt's total mass. A trick was used needed to determine its mass, as Ceres has no moons. Usually, once we know how far away an object is, we can use the orbital periods of moons circling the object and distance of the moons from the object to determine the object's mass. For this, we measure the angular separation between the moon and the object and use basic trigonometry to convert the angular separation into distance between the object and moon. That conversion, though, first requires that the distance to the object and moon to be known. But what can we do, if we have no orbiting moons as in the case of Ceres?

Instead of moons orbiting the dwarf planet, interactions with smaller asteroids coming in close vicinity have been considered. The mass is sufficient to establish hydrostatic equilibrium, i.e., bringing it into a spherical shape. A criterion which is important for the classification as a dwarf planet and distinguishing it from almost all other asteroids and small solar system bodies in general.

Differentiated Internal: Not a Simply Body of Rock and Ice

Ceres has a quite differentiated internal structure and a complex surface. Most asteroids are simple rubble piles of rock loosely held together by gravitational forces. The dwarf planet, though, is different. It has a rocky core covered by an icy mantle (see Fig. 5.11). The mantle is about 100 km thick and contains (melted) water ice. This is quite thick and massive as it corresponds to roughly 23–28 % of Ceres' total mass. The mantle is covered by a thin crust of water ice and light minerals.

Where does this differentiation come from? This depends on the heat sources available during and after its formation. During its formation friction from the accretion of planetesimals played an important role. Thereafter, radioactive decay was the driving force. During such decay, energy in form of warmth is freed and heats up the body. This caused the water ice in the mantle to melt. However, the outer 10 km did not melt. The temperature was not high enough to compete with the ambient temperature in the solar system. Thus, the water ice remained frozen and formed a solid crust of ice. Heavier material such as silicates and minerals, i.e. more rocky material, sunk down and formed the core.

All this may also have caused a resurfacing of the dwarf planet by water volcanos and tectonics. These processes would have erased older geological structures present on Ceres. However, due to its relatively small size, Ceres cooled down quickly and made all these processes come to a halt. Today, Ceres is considered to be a geologically inactive body.

A Complex Surface for the Dwarf

Ceres' surface composition appears to be similar to that of C-type asteroids whose surfaces are dominated by dark carbonaceous compounds. Ceres is quite dark with an albedo of only 0.09, i.e., only 9 % of the incident sunlight is reflected. Although it is similar to C-type asteroids, differences exist, as we will see in the following.

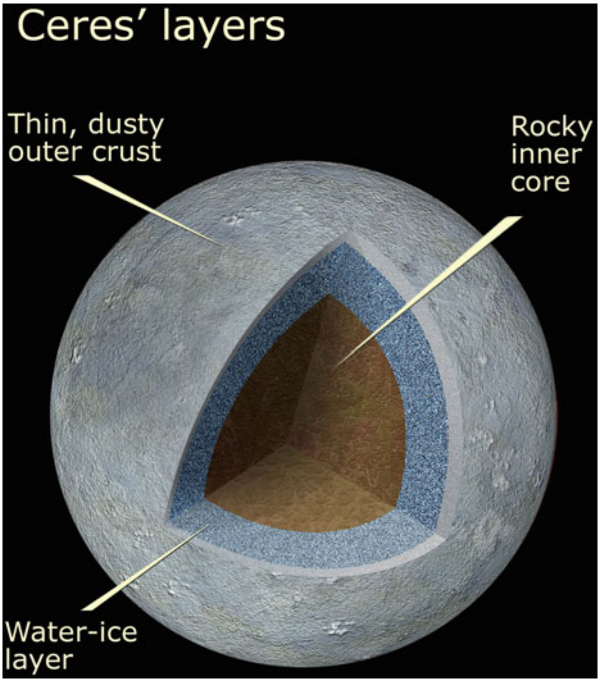


Fig. 5.11 Cutaway of view of Ceres shows the differentiated layers of the dwarf planet (credit: NASA, ESA, and A. Feild (STScI))

Radio observations showed that the surface is covered by powdery regolith which is a layer of loose, heterogeneous superficial material covering solid rock. It includes dust, soil, broken rock, and other related materials.

The maximum temperature on its surface was determined to be about 235 K or −38 °C. At these temperatures, however, water ice is unstable and may be subject to sublimation. If it sublimes, it leaves behind dark carbon rich material which may explain Ceres dark surface.

An important milestone in the observation of Ceres is the arrival of the spacecraft Dawn at March 6, 2015. Before only limited possibilities to directly observe Ceres' surface existed, e.g. by photos taken by the Hubble Space Telescope (see Fig. 5.12). As you can see from the picture, only a few surface features could be determined. The HST revealed a dark spot with a diameter of about 250 km. It was named Piazzi in order to honor Ceres' discoverer Giuseppe Piazzi. For quite some time, it had not been clear what this dark spot was. Astronomers are now quite sure that it is most likely a very large crater.

In addition to observations in visible light, near-infrared images were taken over a whole rotation period with the Keck telescopes located at the Mauna Kea Observatory at Hawaii. These showed several dark and bright features moving with the dwarf planet's rotation and thus being affixed to its surface.

Two of these features are nearly of circular shape and are potentially craters. One of them was mapped to the crater Piazzi discovered in visible light by the HST.

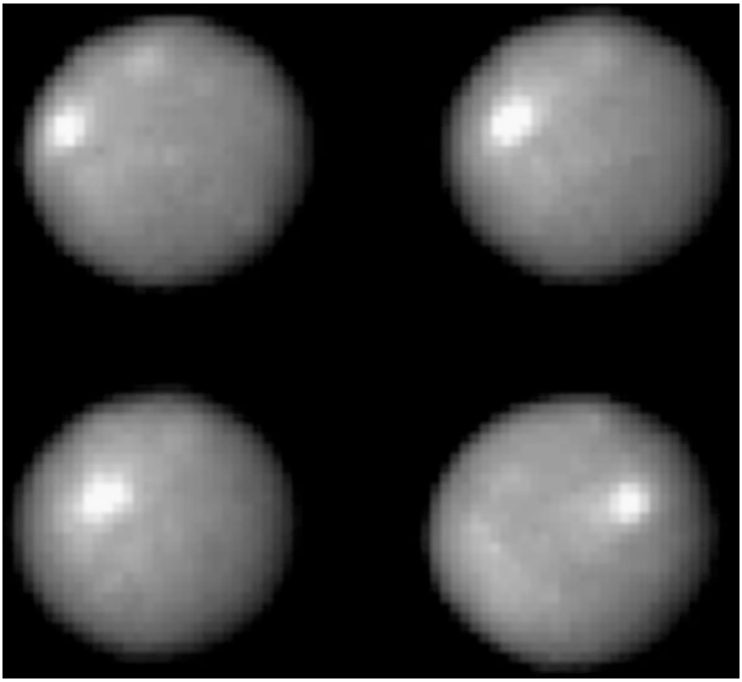


Fig. 5.12 Ceres imaged by the Hubble Space Telescope in 2003/04. Until the arrival of Dawn these wet the images with the best resolution (credit: NASA, ESA, J. Parker (Southwest Research Institute), P. Thomas (Cornell University), and L. McFadden (University of Maryland, College Park))

Already while approaching Ceres, Dawn revealed many new details of the dwarf planet and it became possible to create a more detailed map of its surface (see Fig. 5.13). A remarkable number of craters were newly discovered. Surprisingly, Dawn also detected some "mysterious" bright spots.

The spacecraft entered the orbit of Ceres on March 6, 2015 and started its even more detailed exploration of its host. Also then, two bright spots were visible having an albedo of about 0.4 (see Fig. 5.14). The brightest spot was still too small to be resolved. The dimmer second spot, however, could be more closely examined. The spot appeared to be located in the middle of an 80 km wide crater. The nature of the spots is still unknown.

Dawn was able to observe that the spots brightened during daytime and got fainter at dusk. This suggests that they are at least somehow interacting with solar radiation. Further research analysis is necessary in order to solve this mystery.

Atmosphere

Ceres has a very thin atmosphere. Its surface is partially covered by water ice. Yet, water ice is highly unstable and starts to sublimate when being exposed to direct solar radiation. Similar processes can occur in the outer parts of the solar system.

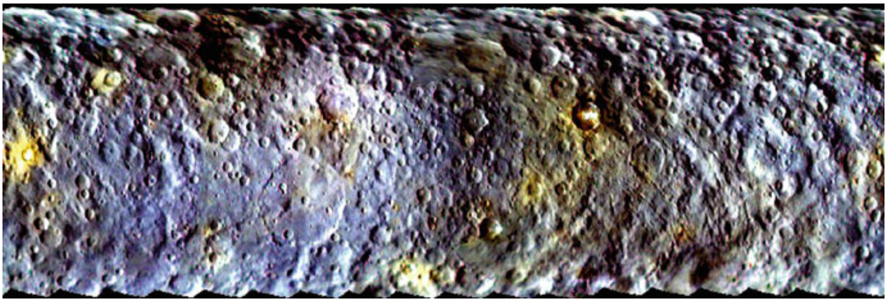


Fig. 5.13 Map-projected view of Ceres was created from images taken by NASA's Dawn spacecraft during its initial approach to the dwarf planet, prior to being captured into orbit in March 2015 (credit: NASA/JPL-Caltech/UCLA/MPS/DLR/IDA)

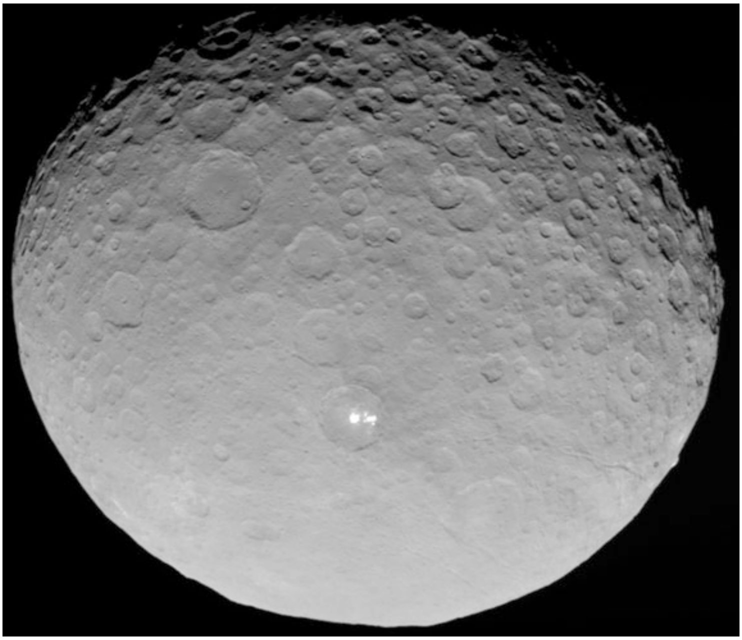


Fig. 5.14 Dawn captures two, yet unexplainable, *bright spots* on Ceres' surface (credit: NASA/JPL-Caltech/UCLA/MPS/DLR/IDA)

However, the strength of this radiation is much stronger in the main asteroid belt than in the far more distant regions of trans-Neptunian objects.

When the solar radiation hits the surface, the induced energy causes the water ice to partially sublimate. The freed gas then escapes the surface and forms part of the thin atmosphere. You should, however, not think of a stable atmosphere as we have it on our home planet Earth. Ceres' atmosphere is much more unstable and transient. It also depends on the dwarf planet's distance from the Sun. On the one

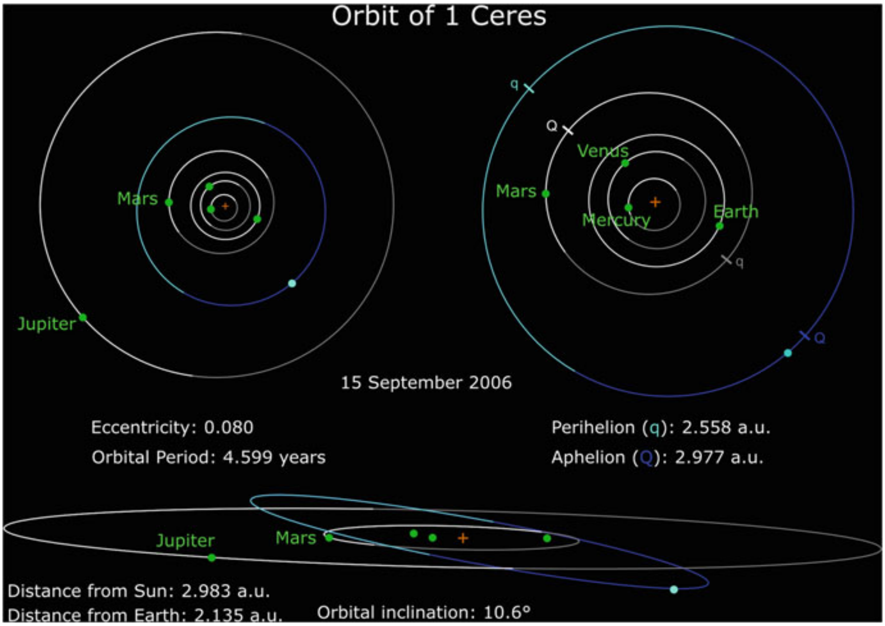


Fig. 5.15 Orbit of Ceres (credit: Orionist, CC BY-SA 3.0)

hand, while being close to its perihelion, more radiation hits Ceres and causes more ice to sublimate. On the other hand, the farther Ceres is away from the soon, the less active the sublimation process is. Unfortunately, Dawn is visiting Ceres at the time close to its aphelion. Thus, it remains questionable whether the spacecraft will be able to detect a significant atmosphere and sublimation processes.

Orbit or Home Sweet Home

After we have learnt so many details about this exotic dwarf planet, the question where it is exactly located still needs to be answered. Of course, we know that it is within the main asteroid belt and thus between the orbits of Mars and Jupiter. Its semi-major axis is at 2.767 AU and thus beyond the Kirkwood gap at 2.5 AU which defines the transition between the inner and the middle belt. Ceres orbit is nearly circular ($e = 0.08$) and brings it at as close as 2.5577 AU to the Sun at its perihelion and as far away as 2.9773 AU (see Fig. 5.15). The orbit is moderately inclined by 10.6° towards the ecliptic plane and it takes about 4.6 years to revolve around the Sun.

Origin

As we have seen, Ceres is as somehow atypical object in the main belt. Thus, a valid question you may ask is: where does it come from? No definite answer can be given so far. Several theories of its origin exist.

One theory proposes that Ceres is one of the few surviving protoplanets left over from the formation of our solar system. Most of them have been destroyed by collisions, formed the terrestrial planets or were ejected from the solar system by Jupiter's gravitational influence. Ceres' differentiated surface and in particular internal structure suits this assumption well.

A second theory, suggested by William B. McKinnon in 2008, postulates that Ceres originally formed in a different region of the solar system, in the proto-Kuiper belt and then migrated to its current location due to effects of the phase of planetary migration. If we take solar radiation, space weathering and induced sublimation processes into account, the surface is similar to many KBOs.

5.5.2 Eris: Pluto's Angel of Death

Eris, bearing the official minor planet designation (136199) Eris, is probably the most massive dwarf planet being about 25 % more massive than Pluto and might also be the largest trans-Neptunian object with an estimated diameter of 2326 ± 12 km. Eris is most famous for sounding the final death knell for Pluto's status as a planet (see Fig. 5.16).

Discovery

Eris was discovered by the American astronomer Michael Brown and his colleagues Chad Trujillo and David Rabinowitz on January 5, 2005 from images taken on October 21, 2003 (see Fig. 5.17). Why this long delay?

The team had systematically scanned the sky for larger trans-Neptunian objects already for several years and had discovered already quite a few of them, such as (50000) Quaoar, (90482) Orcus and (90377) Sedna. They employed automatic image-searching software which analyzed all the images they had taken. This sped up the process of discovering objects drastically and allowed to cover a larger region of the sky with their survey.

However, they had excluded all objects moving less than 1.5 arcseconds per hour in order to reduce the number of so-called false positives, i.e. objects appeared to be TNOs but actually were something else.

When they discovered Sedna in 2003, Brown and his colleagues found out that it had only moved with about 1.75 arcseconds per hour. This was much slower than they had expected these objects to be. Thus, the decided to re-analyze all the images they had collected over the years with a lower speed limit.

Indeed, this additional work of going through all the images again, partially also manually, was worth the effort. They discovered the object 2003 UB_{313} which should later on become Eris. This object was moving very slowly against the background stars and had thus been ignored during the first run of analysis.

The discovery was announced on July 29, 2005 at the same day as the discovery of Makemake and only 2 days after Haumea.

Further observations were carried out to determine Eris' preliminary orbit. Subsequent analysis also revealed that Eris has a moon, Dysnomia.

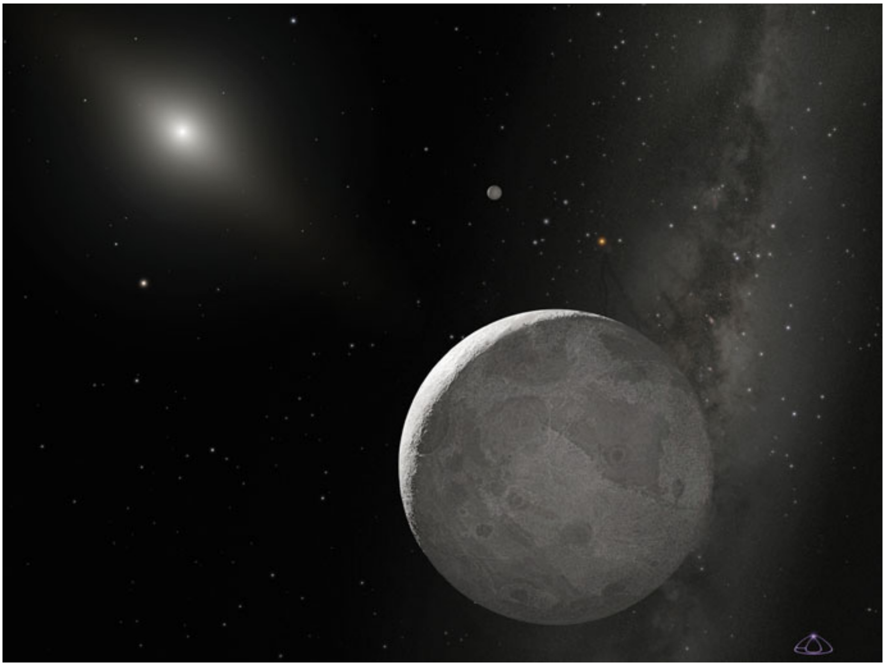


Fig. 5.16 Artist's conception of dwarf planet Eris with its moon Dysnomia (credit: NASA, ESA, and Adolph Schaller (for STScI))

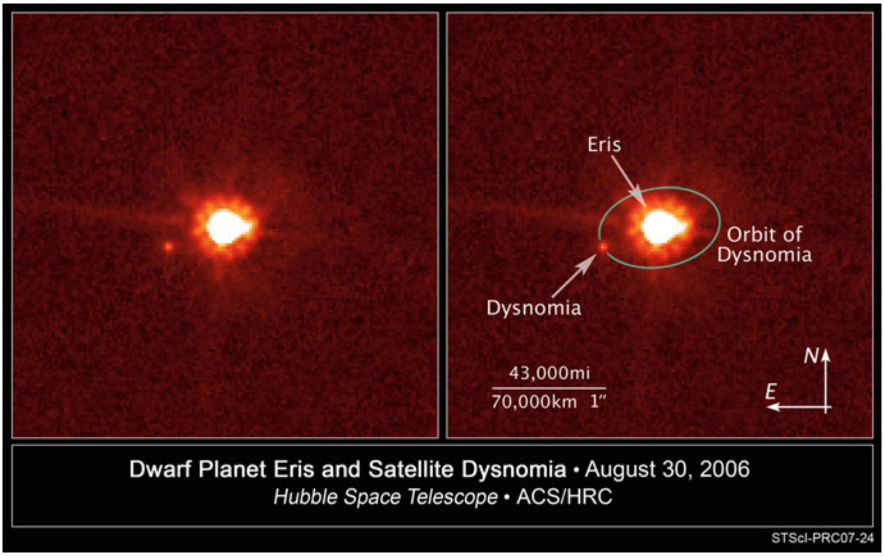


Fig. 5.17 The *left part* shows the dwarf planet Eris (*center*) and its satellite Dysnomia taken with NASA's Hubble Space Telescope on Aug. 30, 2006. The *right part* shows a labeled version of the image of dwarf planet Eris and its satellite Dysnomia taken with NASA's Hubble Space Telescope on Aug. 30, 2006. Dysnomia's projected orbit around Eris is superimposed on the image (credit: NASA, ESA, and M. Brown (California Institute of Technology))

Currently, Eris has an apparent brightness of 18.7 mag bringing it into the reach of good, high-quality amateur telescopes. If it is relatively bright why was it not discovered earlier? Again the same reasons as for Haumea and Makemake apply. We simply looked at the wrong regions of the sky too close to the ecliptic plane.

Classification: The Fate of Pluto

Eris was officially recognized as a dwarf planet by the International Astronomical Union (IAU) in 2008. It is also described as a plutoid, i.e. an ice dwarf whose orbit is beyond that of Neptune. It is also considered to be a scattered disk object (SDO).

Initially, Eris was thought to be larger than Pluto which had by that time still the status of being the ninth planet in our solar system. This tempted NASA to hastily announce the discovery of the tenth planet. Doubts arose within the scientific community and led to a re-evaluation of Pluto's planet status. The IAU, consequently, revised the definition of what a planet is in 2006. Thus, it can be said without any doubt that the discovery of Eris sounded the final death knell for Pluto as a planet. Doubts had existed for a very long time, as we have also seen earlier in this chapter.

Eris is named after a Greek goddess bearing the same name. However, sometimes when searching for material about Eris, a different name besides its provisional designation 2003 UB_{313} pops up: Xena. Why that? Why another name?

At the time of its discovery, it was not clear which status the newly discovered object will be assigned: a minor planet or a planet. Both groups of objects have a different nomenclature. Thus, Eris had also been known as Xena for some time.

Orbit

Eris has a highly eccentric and inclined orbit ($e = 0.44$, $i = 44°$) with its semi-major axis at 67.7 AU. At its perihelion Eris is about 37.91 AU from the Sun whereas during the aphelion this extends to 97.65 AU (see Fig. 5.18). The orbital period is 558 years. Its perihelion brings Eris well within the orbit of Pluto but not close enough to Neptune to be gravitationally influenced by the gas giant.

Eris' orbital characteristics suggest that it is a scattered disk object (SDO) rather than a Kuiper belt object (KBO). It could also be a trans-Neptunian object (TNO) that had originally been a KBO but was then subsequently scattered into its current more distant orbit due to gravitational interactions with Neptune. In addition, as we will see, the limited information we have about Eris' surface characteristics supports the theory of Eris being a SDO.

The dwarf planet last reached its perihelion between 1698 and 1699 and was at its aphelion in 1977. This means, however, that it still is in its most distant part of its orbit. Currently, Eris is the most distant known object in our solar system if we neglect long-period comets and manmade space probes such as Voyager 1 and 2 as well as Pioneer 10. In theory, the detached object Sedna, which we came along while discussing trans-Neptunian objects, is farther away from the Sun having a

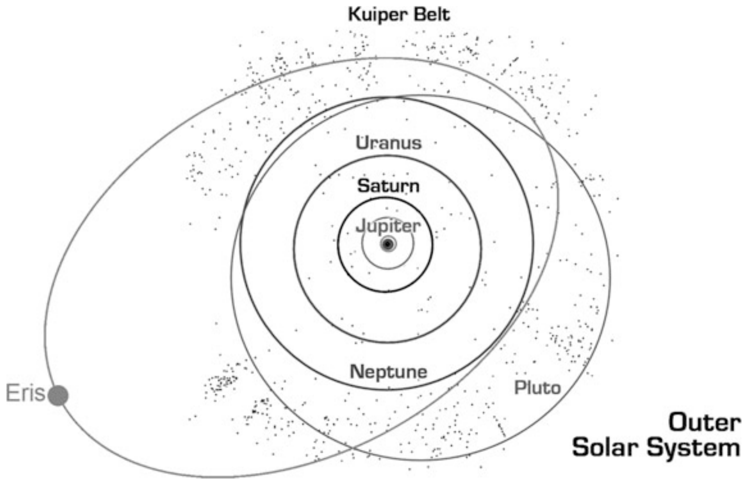


Fig. 5.18 The dwarf planet Eris' orbit in the context of the outer solar system (credit: NASA)

semi-major axis of about 524 AU. Yet, currently, Sedna is moving along a part of its orbit that is closer to the Sun. Sedna will come to perihelion around 2075–2076.

A Matter of Size

Since its discovery Eris' diameter has been gradually refined and reduced. In 2003, considering images of the Hubble Space Telescope, Eris' diameter was estimated to be 2397 ± 100 km. For this, the classical approach of using the distance to Eris, its albedo and its brightness was used. This is comparable in size with Pluto having a diameter of 2368 ± 20 km. At that time, Eris appeared to be slightly larger than Pluto considering the best case of tolerances. Something that also became apparent is Eris' extremely high albedo of 0.96. No other larger body in our solar system has a higher albedo.

In 2007, the Spitzer Space Telescope was used to refine the dwarf planet's diameter. The measurements indicated a diameter of $2600 + 400–200$ km.

In 2010, a rare event occurred, a stellar occultation. Never before had an occultation for such a distant object in our solar system been observed. The results casted serious doubts on previous measurements, in particular those of the Spitzer Space Telescope. Data obtained from the occultation during which Eris passed in front of a star, indicated a diameter of 2362 ± 12 km.

This is very similar to Pluto. It is very difficult to say which object is larger, Pluto or Eris. Pluto's atmosphere further complicates this assessment which interferes with measurements of its solid surface. So, the exciting question remains: which object will make the race for the largest TNO we know?

Mass Also Matters

Fortunately, the discovery of Dysnomia helped determining Eris' mass much more precisely. We have already seen how we can determine an object's mass when considering an orbiting moon. Dysnomia's orbital period was found to be

15.774 days and having a mean distance of 37,430 km from the dwarf planet. Using this information allowed to estimate the mass of Eris to be about $(1.67 \pm 0.02) \times 10^{22}$ kg. If we compare this to Pluto's mass of 1.25×10^{22} kg, we can see that Eris is about a fourth more massive as Pluto while having approximately the same size. Hence, Eris' density must be higher.

If we take the diameter determined during the stellar occultation into account, we end up with a density of 2.52 ± 0.05 g/cm^3. This is substantially larger than Pluto's 1.75 g/cm^3.

So, we now have a rough estimate of Eris' density. What does this tell us? As both, Pluto and Eris, appear to have a similar size but their density varies, they must be of different composition. Eris most probably is composed of much more rocky and denser material compared to Pluto.

Some astronomers suggest a composition of about 70 % rock and 30 % ices. Such an assumption is, however, unreliable and questionable as it highly depends on Eris' actual size and density which we do not know for sure yet.

Surface, Composition, and Atmosphere

Again, it is not possible to observe any surface features directly from Earth independent from the telescopes being ground-based or in space. Hence, indirect measurements, such as spectral analysis are employed, to derive some features of Eris' surface composition.

On January 25, 2005 spectra were obtained at the 8 m Gamma Gemini North Telescope on Hawaii. Analysis of the infrared spectrum revealed the presence of methane on the dwarf planet's surface. Further analysis indicated abundances of carbon monoxide and nitrogen. This is similar to Pluto.

The surface, due to its highly eccentric orbit, varies between 30 and 56 K. Sublimation processes may be possible at these temperatures as methane is highly volatile. This, however, raises another question. How can we explain the observed abundances of methane if it is so volatile and sublimation may occur?

Two possibilities are conceivable. First, Eris has always resided in such large distances to the Sun where it is cold enough not to trigger sublimation. Secondly, there may be an internal source for methane which replenishes the gas escaping the surface.

No further details are known about Eris' surface. We can only observe that its color is different from Pluto (and also Neptune's moon Triton which is thought to be a captured TNO). Whereas Pluto and Triton show a reddish color, Eris appears to be simply grey like many other SDOs.

Eris may have an atmosphere similar to the one of Pluto. When the dwarf approaches its perihelion, sublimation processes will be enforced causing gases to escape from the surface and forming a thin atmosphere. Then, when Eris is departing from its perihelion, the surface temperature will continuously decrease until it reaches its minimum at the aphelion. The falling temperatures will enable re-sublimation, i.e. the gases in the atmosphere will freeze out and fall back onto the surface.

Fig. 5.19 The W. M. Keck Observatory, on top Mauna Kea, Hawaii (credit: SiOwl, CC BY 3.0)

Dysnomia: A Companion for Eris

Eris' moon, Dysnomia, was discovered on September 10, 2005 by Michael Brown and his colleagues using the 10 m Keck II telescope at the Mauna Kea (see Fig. 5.19). The additionally used a relatively new technology, the laser guide star adaptive optic instrument which tries to minimize the effects of Earth's turbulent atmosphere.

They found that Dysnomia orbits Eris on a near circular path ($e = 0.013$) at a mean distance of 37,430 km within 15 days, 18 h and 31.7 min. Different values of its diameter exist ranging between 100 and 490 km.

Dysnomia is most likely the result of a collision between Eris and a larger object.

5.5.3 *Haumea and Its Family*

The small solar system body Haumea is classified as a dwarf planet and belongs to the group of plutoids, i.e. the ice dwarfs that have their orbits beyond Neptune. Its official designation as a minor planet is (136108) Haumea.

Discovery and Classification

Haumea was discovered in 2004 by the American astronomer Michael Brown and his colleagues using the Palomar Observatory in California (see Fig. 5.20). It was also discovered independently from them by José Luis Ortiz Moreno and his colleagues at the Sierra Nevada Observatory in Spain.

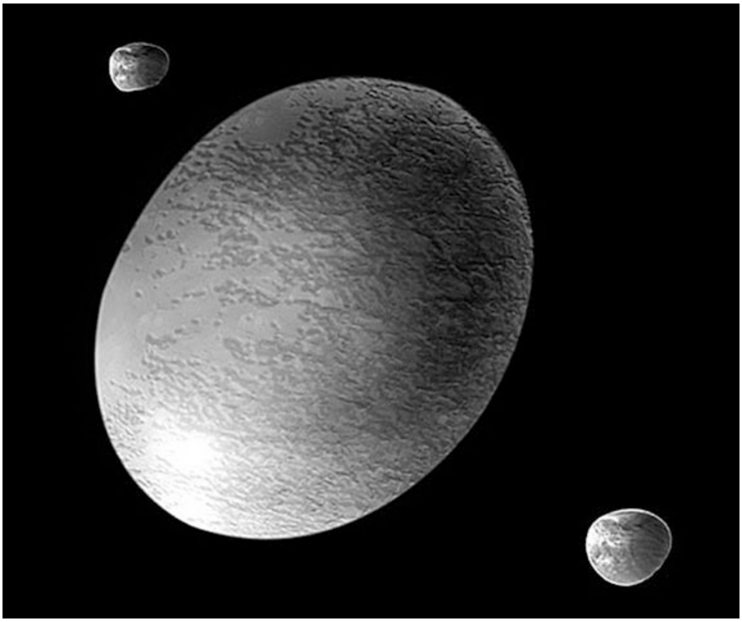


Fig. 5.20 An artist's concept of dwarf planet Haumea and it's moons, Hi'aka and Namaka (credit: NASA)

There is a heavy and emotional debate going on over the discovery of Haumea with reciprocal allegations from both sides. How discovered it first? Was everything correct in the procedure? All this is still debated and led to an uncommon approach chosen by the International Astronomical Union (IAU) when announcing the discovery of the dwarf planet on September 17, 2008. It is standard to also name the discoverer when publishing the notification. In this case, the IAU, however, refrained from doing this.

Yet, in this announcement, Haumea was officially recognized as a dwarf planet as it is massive enough to reach hydrostatic equilibrium and orbits the Sun. Hence, the first two criteria for a planet as defined by the IAU in 2006 are fulfilled. However, Haumea fails to comply with the third one. It is not dominant enough to have cleared its neighborhood. Therefore, it cannot be considered to be a planet but a dwarf planet.

However, a debate was going on whether it really had reached hydrostatic equilibrium as observations suggested that Haumea was not of spheroid. It more resembled an ellipsoid. Astronomers, however, suggested that such as distortion is most probably the result of its extremely fast rotation period of 3.9 h. This is one of the fastest that can be found in the solar system for similar sized objects.

Haumea is named after the matron goddess of the island of Hawai'i where the Mauna Kea Observatory is located.

Orbit and Rotation

Haumea had been originally classified as a classical Kuiper belt object (KBO), as it has a typical orbit of a cubewano. Its orbital period is 284 years and its semi-major axis lies at about 43.342 AU from the Sun. Haumea reaches its closest point to the Sun, the perihelion, at 35 AU and its orbit takes Haumea as far away from our central star as about 51 AU.

The dwarf planet is in a weak, non-stabilizing 7:12 resonance with Neptune, i.e. for every 12 orbits of Neptune around the Sun, Haumea makes seven. The gravitational effects induced by this resonance gradually modify Haumea's initial orbit over the course of a billion years. Hence, compared to other cubewanos, its orbit cannot be considered to be absolutely stable.

One aspect that is striking is Haumea's visual magnitude of 17.3 mag. This makes it the third brightest object in the Kuiper belt after the dwarf planets Pluto and Makemake and renders it rather an easy target for good, larger amateur telescopes.

Why was it then not found before if it is such a relatively "easy" target? The answer to this is related to its orbit which strongly inclined by 28° towards the ecliptic plane. Most objects in our solar system that had been known before, except for comets and some other "strange" bodies, orbit within or near the ecliptic plane. Therefore, earlier sky surveys mainly concentrated on the ecliptic plane and nearby regions. Haumea was clearly outside the scope of these observations. Only later, after the ecliptic was well covered, surveys extended to regions farther away. This is also when Haumea and other KBO were detected.

After its discovery, the next step was to find out more about this object. Photometric analysis was used to determine and investigate its light curve. This curve, which describes Haumea's varying light intensity as a function of time, revealed large fluctuations in brightness over a period of 3.9 h which suggested its rotational period be at the same value. As previously pointed out, this makes it the currently known fastest rotating object that is in hydrostatic equilibrium. Yet, this rapid movement has some serious implications on its shape as we will see in the following.

Physical Characteristics

Usually the size of an object can be determined by measuring its angular diameter, taking into account the distance to it and using some basic trigonometry. However, the problem with Haumea as for most other small solar system bodies is, as their classification suggest, that they are simply too small to measure their angular diameter. For this, we would need to be able to resolve their "planetary" discs, which is hardly possible even with the currently largest available telescopes.

Therefore, we need indirect measures such as also explained in the context of asteroids which, e.g., take into account the object's brightness, distance and albedo. The idea behind is strikingly simple. Bodies appear bright to us either when they are large or because they are highly reflective, i.e. have a high albedo. If we then know the albedo, we can roughly estimate the object's size.

Though being simple in principle, there is a major problem: for most objects we do not know their albedos. We often have to guess it or derive it from similar

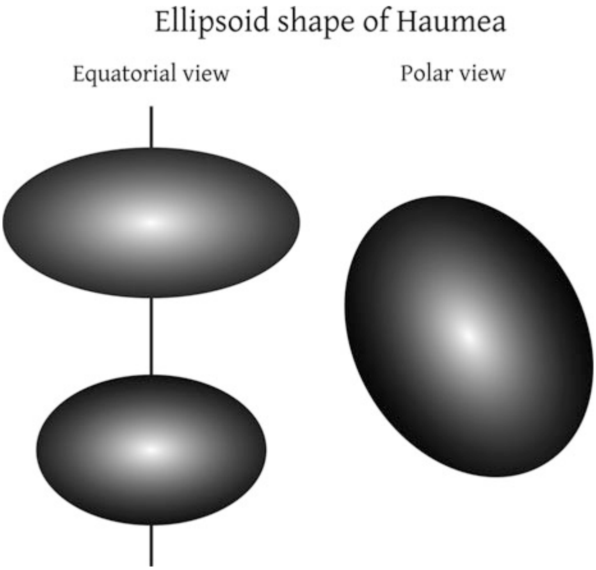


Fig. 5.21 The calculated ellipsoid shape of Haumea (credit: Kwamikagami, CC BY-SA 3.0)

objects. This can lead to highly unreliable and varying results. Fortunately, Haumea is large enough to measure its thermal emissions which can give an approximate value of its albedo and finally allow the determination of its size.

Haumea's size, however, turned out to be a complex issue being related to its rapid rotation which causes some distortion and makes the dwarf planet to look more like an ellipsoid than a sphere. More detailed analysis suggests triaxial ellipsoid with approximate dimensions of $2000 \times 1500 \times 1000$ km and an evenly distributed albedo of 0.71 (see Fig. 5.21).

The Spitzer Space Telescope determined Haumea's diameter to be roughly $1150 + 250–100$ km and having an albedo of 0.84. We can already see from these results how difficult it is to determine the dwarf's actual size which in the end still remains partially unknown to the present day.

Haumea's fast rotation is not only responsible for its shape but also imposes strong constraints on its composition. Numerical analysis and simulations seem to suggest a density in the range of 2.6–3.3 g/cm^3. If we compare it to other objects we can find that our Moon has a density of 3.3 g/cm^3 and Pluto, e.g., 2.0 g/cm^3.

How can we arrive at such an assumption? We can have a look at a similar body rotating as fast as Haumea and determine how it will look like at different densities. If we consider Haumea to have a lower density, such as the one of Pluto, the dwarf planet's rapid rotation would result in a more elongated shape than observed fluctuations in the brightness allow.

It is better to use a more descriptive example to show this effect. Imagine a body with a low density, such as a balloon filled with water. Now we let it rotate rapidly. The balloon will stretch and reach an elongated shape. If we now fill the balloon with material of higher density, such as a yogurt, the balloon will be elongated to a lesser extent.

Haumea's density gives indications on its structure. It probably comprises silicate minerals such as olivine and pyroxene which resembles many rocky objects in the solar system.

Scientists assume that Haumea is basically a rock covered with a thin layer of ice. A thick layer of ice is not very likely to exist anymore due to a heavy collision that happened during its evolution and may have formed the Haumea family as we will see in a few moments.

Surface

As for many other KBOs, not much is known about Haumea's surface yet. Direct observations are practically not possible due to its large distance and small size. However, indirect measurements allow some indications. Spectral analysis shows strong crystalline water ice features. This is strange, as the development of crystalline water ice requires temperatures below 110 K whereas Haumea's surface temperature is below 50 K. This is more or less the temperature region where amorphous ice dominates.

Furthermore, crystalline water ice is highly unstable when exposed to solar radiation, i.e. space weathering. It will revert back to amorphous ice in an order of 10 million years. KBOs, however, are considered to be much older. The only explanation for the fact that crystalline water ice can still be detected on Haumea's surface is that some kind of resurfacing must have happened within the last 10 million years. The reasons that triggered these processes and what kind of resurfacing processes were involved remains yet unknown.

Moons: Haumea Is Not Alone

Haumea is accompanied by two moons both discovered in 2005 by Michael Brown and his team (see Fig. 5.20). They used the telescopes at the Keck Observatory at the Mauna Kea on Hawaii.

The first moon, (136108) Haumea I Hi'iaka is the largest and brightest of the two and lies a bit further outward. It has a diameter of roughly 310 km and orbits Haumea on a near circular path every 49 days.

The second moon, (136108) Haumea II Namaka is about only one-tenth the mass of Hi'iaka and orbits Haumea on a highly elliptical orbit with 18 days. On its orbit it can come closer to Hi'iaka which in turn perturbs Namaka's orbit by gravitational interactions.

Family and Origin

Haumea is member of the so-called Haumea family, a collisional family whose members share similar physical and orbital characteristics. It is the first family identified among trans-Neptunian objects (TNO).

The family potentially formed by the impact of a larger object on Haumea. Haumea is the largest member of the family which in addition consists of its moons and the Kuiper belt objects (55636) 2002 TX_{300}, (24835) 1995 SM_{55}, (19308) 1996

TO_{66}, (120178) 2003 OP_{32}, (145453) 2005 RR_{43}, (86047) 1999 OY_3, (416400) 2003 UZ_{117}, (308193) 2005 CB_{79}, 2003 SQ_{317} and (386723) 2009 YE_7.

Collisional formation of the family requires a progenitor some 1660 km in diameter, with a density of about 2.0 g/cm^3, similar to Pluto and Eris. During the formational collision, Haumea lost roughly 20 % of its mass, mostly ice, and became much denser up to the currently assumed values.

The presence of the Haumea family indicates an origin in the scattered disk. Two reasons for this can be considered. Today's Kuiper belt is too scarcely populated. The likelihood of a collision with an impactor of the required size in order to form the family considered over the age of the solar system is only 0.1 %. It is very difficult to believe that this happened.

In addition, the family cannot have formed in the proto-Kuiper belt which was much denser. However, when Neptune during planetary migration ploughed through it scattering and perturbing many of the proto-Kuiper belt bodies, the family would have been disrupted. Consequently, the scattered disk remains the most likely source of Haumea and its family.

5.5.4 Makemake

Makemake is another dwarf planet having the official minor planet designation (136472) Makemake. It is perhaps the largest classical KBO having a diameter of roughly two-thirds of Pluto's. You should remember that Pluto is not a member of the classical population ("cubewano") but a plutino.

Discovery and Classification

Makemake was discovered on March 31, 2005 by Michael Brown and his team at the Palomar Observatory in California (USA). The discovery was announced on July 29, 2005. It was then officially recognized as a dwarf planet by the IAU in July 2008.

Makemake is a classical KBO, a cubewano, and a member of the dynamically hot population. This might have had implications on its relatively late discovery. Though its brightness is only about a fifth of Pluto, it is still one of the brightest KBOs. Makemake, however, was not discovered until well after many much fainter KBOs. The reason for this is similar to the case of Haumea. Its relatively high inclination of 29° towards the ecliptic plane made it out of scope for earlier sky surveys.

The name Makemake is derived from a creator deity. It is the creator of humanity and the god of fertility in the mythology of the Rapa Nui, the native people of the Easter Island. This name was chosen, as the date of discovery was close to Easter 2005.

Orbit

Makemake's orbit is similar to that of Haumea. Although it is slightly shifted farther outwards. The highly inclined and moderately eccentric orbit ($i = 29°$,

$e = 0.16$) has its semi-major axis at 45.660 AU and an orbital period of 310 years. Makemake reaches its perihelion at 38.543 AU and its aphelion at 52.778 AU.

Makemake's Size

As with the other KBOs it is particularly difficult to determine the size of Makemake. Similar approaches as for Haumea and Pluto were applied. The lack of moons makes it more difficult to obtain good estimates of its diameter.

Results of infrared measurements of the Spitzer Space Telescope and the Herschel Space Telescope were combined with similarities of spectrum with Pluto. This delivered an estimated diameter between 1360 and 1480 km.

A more precise estimate was obtained by observing a stellar occultation in 2011. During such an event the light intensity received from the star is measured. While Makemake passes the star, the intensity will drop. Measuring the intensity of the drop and its duration gives a rough estimate of Makemake's diameter. The longer the drop lasts, the larger Makemake is.

The measurements obtained from this event suggest that Makemake is slightly larger than Haumea and has a diameter of about (1502 ± 45) km $\times$ (1430 ± 9) km.

This makes Makemake the third largest of the known trans-Neptunian objects after Pluto and Eris.

Spectra and Surface

Licandro et al were able to measure Makemake's spectra in the visible light and the near-Infrared in 2006. The results suggest that its surface is astonishingly similar to Pluto. It appears to be red in visible spectrum. It is significantly redder than Eris.

The near-Infrared spectra revealed the existence of methane (CH_4) as is the case on Pluto's and Eris' surface. Methane is probably present on the surface in form of large grains at least 1 cm in size. Abundances of ethane (C_2H_6) and tholins were also confirmed. The later ones are supposed to be created by photolysis of methane induced by solar radiation, in particular solar UV radiation.

So, how does this work? What happens during photolysis? During photolysis, sometimes also called photo dissociation, a chemical compound, such as methane, is broken down by photons. A photon having a specific energy affects the chemical bounds of methane (CH_4) when it hits the molecule. The result of this dissociation can be a number of short-lived, fast reacting radicals such as CH_2, CH, and methyl radicals (CH_3). The self-reaction of methyl radicals then leads to the formation of ethane (C_2H_6).

$$2\,CH_4 + 2\,h\nu \rightarrow 2\,CH_3 + 2\,H \rightarrow C_2H_6 + 2\,H$$

The tholins on Makemake's surface are considered to be responsible for its reddish color.

Further photometric measurements were made in the far-infrared (24–70 μm) and the submillimeter range (70–500 μm) by the Spitzer Space Telescope and the Herschel Space Telescope. These suggest that the dwarf planet's surface is not homogeneous. On the contrary, the majority of its surface is probably covered by nitrogen and methane ice. However, dark spots moving with the surface while rotating exist that make up about 3–7 % of the total surface. The nature of these dark spots is still unknown.

Atmosphere

Looking at the similarities between Makemake and at least Pluto, it appeared to be plausible that Makemake also exhibits a similar atmosphere as Pluto. To verify this assumption, the stellar occultation of April 23, 2011 was used. However, Makemake abruptly blocked the starlight while passing in front of it. This indicates that no atmosphere is present. If the latter was the case, the drop of brightness would not be so sudden but more slowly.

This observation, however, does not preclude Makemake from having an atmosphere. It may have a transient atmosphere, similar to Pluto, close to its perihelion created by the sublimation of methane.

Chapter 6
Exploration of Small Solar System Bodies

So far we have learnt a lot about the different types of Small Solar System Bodies and dwarf planets. Yet, the question remains. How do we get all the information from these small, distant bodies? These small objects are very difficult to observe. Using optical telescopes, even the largest ones we currently have on Earth, it is hardly possible to see something else than a faint star-like dot. Only for very few objects we have been able to resolve their discs. Yet, we had no chance to observe any surface details directly. This is in particular true for the more distant objects in the outer regions of our solar system. Just imagine how challenging it was to resolve Pluto and obtain very coarse surface maps. With regard to this, we should not forget that Pluto is among the largest objects there in that region. Other objects are much smaller and can be even more distant. We have seen that the dwarf planet Eris might be larger than Pluto. However, it is also clearly more distant to us. This raises the bar again for observing it.

So what possibilities do we have to observe these bodies? First, we have to look at the different types of objects that are classified in the group of Small Solar System Bodies. If we, e.g., consider comets close to their perihelia and Near Earth Asteroids (NEA), which can come pretty close to Earth, other techniques are available to observe and analyze them. For these objects, some methods are applicable that require small distances, such as radar observations. In contrast, the trans-Neptunian objects and centaurs are by far more difficult to observe, as they are very small and faint.

In the following, we will briefly go through the basic techniques that can be applied for this purpose and also see how we can indirectly determine some exemplary characteristics of small, distant objects (e.g. mass, diameter). This chapter will be concluded with a more detailed discussion of spacecraft in past and present mission to small solar system bodies.

M. Moltenbrey, *Dawn of Small Worlds*, Astronomers' Universe,
DOI 10.1007/978-3-319-23003-0_6

6.1 Light Curves

Light curves are a powerful tool to obtain information about an object that can otherwise not be resolved. They are determined by photometric analysis. Photometric methods measure the flux or intensity of an astronomical object's electromagnetic radiation. It helps to determine luminosity of an object and allows studying light variations of object such as variable stars, active galactic nuclei, supernovae and small solar system bodies.

6.1.1 Light Curves and Rotation Periods

What are light curves? A light curve is graph of light intensity of a celestial object or region, as a function of time. A light curve can be used to estimate the rotation period of small solar system bodies and moons. It is often not possible to actually resolve a small object in our Solar System even with the largest and most powerful telescopes. Thus, no direct information about the object's surface etc can be obtained. To circumvent this problem, light curves are determined by measuring the total light emitted or reflected by an object as function of time. The light variations may be due to the object's shape. Just think about a peanut-like object, such as an asteroid, orbiting the Sun. Depending on which part is facing the Sun, a smaller or a larger one, different amounts of sunlight will be reflected.

Figure 6.1 shows the light curve of the asteroid Eros as obtained by the NEAR spacecraft. The image shows Eros' brightness variations over the course of one Eros day which is roughly 5.3 h. If the light curve covered a longer period of time you could see the curve pattern repeat. By measuring the time between two repeating features of the curve (e.g. peaks), we can determine Eros' rotation period. The curve may also give some indications on surface properties.

6.1.2 Light Curves and Occultations

Light curves can also be used to determine an object's size and derive information about its atmosphere even in cases where it cannot be resolved in telescopes. For this purposes, special events are considered, so-called occultations during which the small body we want to analyze passes in front of another object. Quite common are stellar occultation where the occulted object is a distant star. How does it work?

The basic principle is shown in Fig. 6.2 in which an object (dark grey) passes in front of another one (light grey). Let us assume that the two objects are an asteroid and a star. The drawings are of course not at scale.

In the first step, the light of the star is not obstructed. So, we can measure its full brightness at (1). Then, the asteroid begins to pass in front of the star (2). The

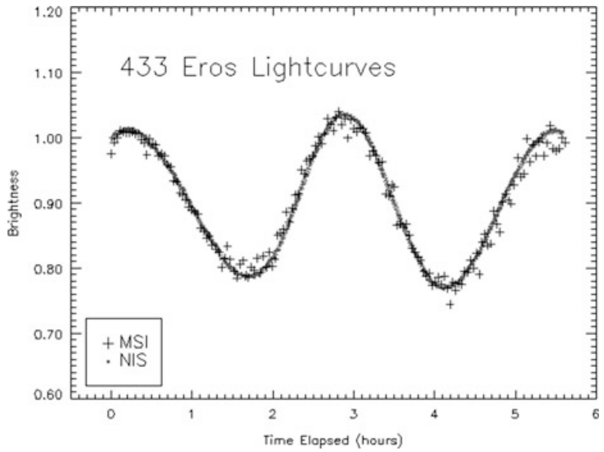


Fig. 6.1 Light curve of asteroid Eros as obtained by the NEAR spacecraft (credit: NASA)

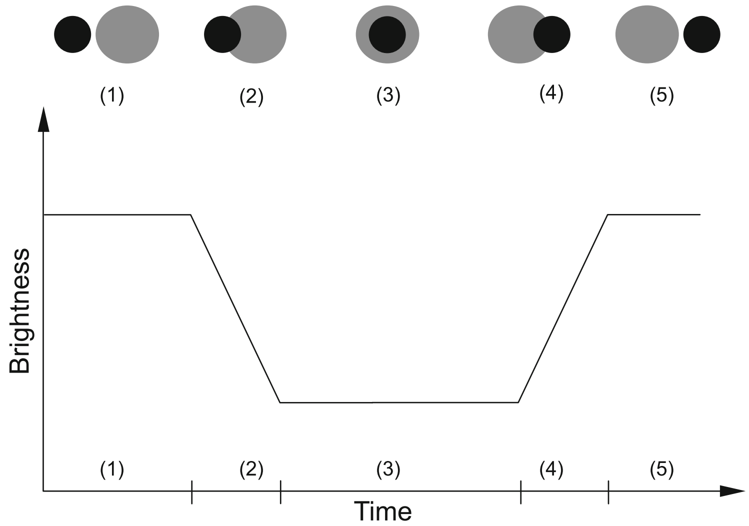


Fig. 6.2 Basic principle of using stellar occultations for analyzing the involved objects

asteroid absorbs some of the starlight. Consequently, the star's brightness decreases. At step (3), the asteroid is completely in front of the star. The star's brightness is reduced to a maximum. In the following steps, the starlight returns to normal while the asteroid recedes.

If we now measured the time between the start and the end of the occultation and take into account the asteroids orbital velocity, we can determine its size.

This method, however, is even more powerful. We have seen that astronomers were able to detect ring systems around such small and distant objects such as the

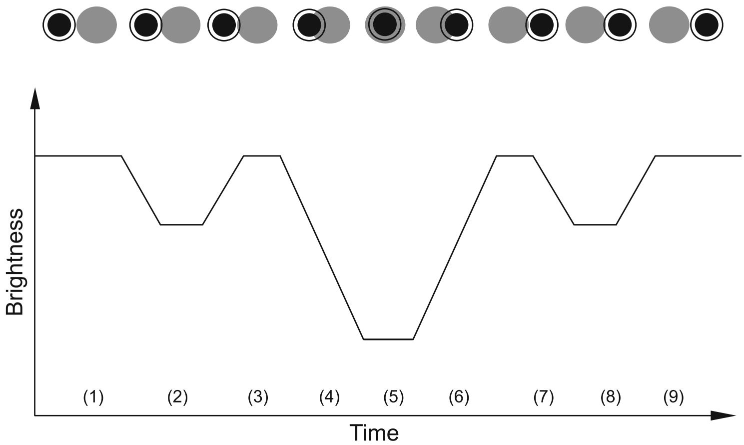


Fig. 6.3 Basic principle of how to detect a ring system around a body using stellar occultations

centaurs Chiron and Chariklo. They also used stellar occultations for this. The basic idea is similar to the one just outlined.

If we have a closer look at Fig. 6.3, we can see that our previous example has been extend by the asteroid having a ring around it. In the case of a present ring system, the resulting light curve looks different. We now have two additional smaller drops in brightness. They occur when the ring moves in front of the star (steps (2), (3) and (7), (8)). You can imagine that your instruments have to be very sensitive to detect such small drops. Simply assume the case of Chariklo where the detected rings or only a few kilometers wide.

6.2 Size and Mass Determination

We may now be able to determine the body's size via the help of occultations. But what can be done if there are not occultations or we do not have the exact details of the object's orbit.

6.2.1 Determining the Size

Other methods exist that can help us to determine an object's size. The first one may only give a rough estimate but is often sufficient for first considerations. For this, we assume an evenly distributed albedo over the asteroid's whole surface. We can then deduce its mean diameter, as we know that an object of a specific size should reflect a certain amount of light when having a specific albedo. This approach is, however, much depending on the assumed albedo. The assumed albedos are often rough estimates, e.g. taking into consideration similar objects for which the albedo is better known. The diameter can, therefore, vary considerably.

The second method is more accurate. We also need the distance to the object once more. However, we also need the object's angular diameter for this. This means we must be able to resolve it and measure the diameter. This is hardly possible for most of the small bodies. Approaching spacecraft, however, can use this method. How does its work? If we know the distance d to the object and its angular diameter, we can use simple trigonometry to obtain the physical diameter by:

$$D = \frac{2\pi \cdot d \cdot A}{360}$$

A simple method though in case of our small solar system bodies not always useful.

6.2.2 *Mass Determination*

Determining the mass is a bit trickier and requires a companion of the object we want to analyze. This companion could be a moon or any other object that interacts gravitationally with our subject.

The idea behind this type of measurement is not difficult to understand and depicted in Fig. 6.4.

We need to measure a few parameters such as the distance d of the object from Earth, the orbital period of the moon P circulating object and the moon's distance D from object. How, do we get D? To determine the moon's distance to its primary, we can use trigonometry again by applying the distance d to Earth and the measured angular distance a between the moon and its primary.

Then, it is possible to derive the common mass of the double system by employing Newton's law of motion and gravity by

$$m + M = \left(\frac{4\pi^2}{G}\right) \cdot \left(\frac{D^3}{P^2}\right)$$

Usually, the moon's mass is negligible as it is by far small than that of the object under consideration. Hence, with good approximation the total mass resembles the object's mass.

6.2.3 *Spectroscopy*

Spectral analysis can be used to determine characteristics of a small solar system body, especially those spectra taken in the visible and near Infrared light. Spectroscopy can be used to derive many properties of distant objects, such as their

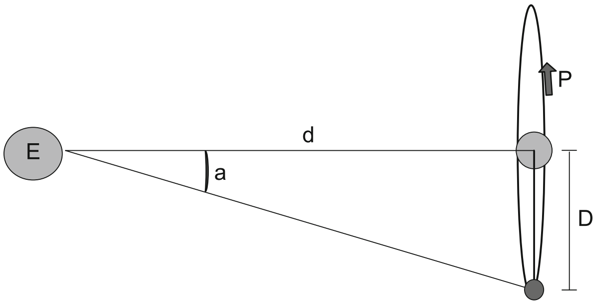


Fig. 6.4 Determining the mass of the system comprising of a moon and a primary body

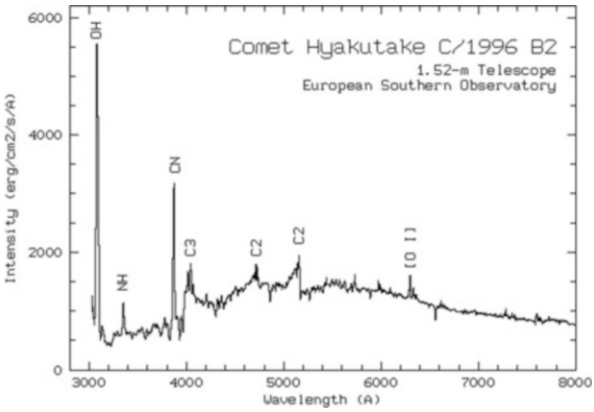


Fig. 6.5 Complete visual spectrum of the comet C/1996 B2 Hyakutake (credit: ESO)

chemical composition, temperature, density, mass, distance, luminosity, and relative motion using Doppler shift measurements.

Dust and molecules on the surface of small bodies can be analyzed by spectroscopy, as they causes absorption lines. When light is reflected off of a small body, its spectrum is changed. This is because the sunlight incident on the surface materials absorbs part of the spectrum and reflects part due to its molecular nature, resulting in light and dark bands within the spectrum reflected off the small body.

By determining which wavelengths of the spectrum were absorbed, and how strongly each band was absorbed relative to other bands, we can get an indication of what mixture of materials are on the small body's surface.

Figure 6.5 shows the complete visual spectrum of the comet C/1996 B2 Hyakutake.

6.2.4 Radar Observations

Another possibility to observe Small Solar System Bodies, in particular asteroids and comets that are in the inner solar system, is the use of radar. It allows observing nearby astronomical objects by reflecting microwaves off target objects and analyzing the reflections. It is a relatively new technique that has only been carried out for about 60 years.

Radar astronomy is substantially different from radio astronomy in that the latter is a passive observation and the former an active one. Radar systems have been used for a wide range of solar system studies and have become more and more popular with asteroids and comets in the recent years. The number of observed objects though remains small. Only 18 comets have been studied by radar, including 73P/Schwassmann-Wachmann and its breakup. By far more asteroids have been observed so far: 539 Near-Earth asteroids and 138 main belt asteroids.

The Basic Principle

The basic principle of radar astronomy is the same as for most radar applications on Earth, too. Radar pulses are transmitted to a target object. These pulses bounce there and are received again. Now, we measure the time it takes for the signal to come back. As soon as we know the time, we can determine how far the object, e.g. an asteroid is away, as we know the velocity of propagation of the radar waves.

Yet, this is not the only information we can derive from radar observations. If we further consider the Doppler shift we can even determine how fast the object is moving and in which direction. We understand under the Doppler shift the change in frequency of a wave, e.g. a radar wave, for an observer moving relative to its source. A daily life example is an ambulance on the street that sounds its siren while you observe it. The sound of the siren changes while the ambulance approaches, passes, and recedes from you. Compared to the emitted frequency, the received frequency is higher during the approach, identical at the instant of passing by, and lower during the recession. This gives the typical sound experience you probably already have heard several times.

We can apply this consideration to astronomical purposes. An asteroid moving towards us compresses the echo to higher frequencies and when it is receding it stretches the echo to lower frequencies.

What Can We Observe?

The resolution we can obtain with radar observation is highly dependent on the sender/receiver combination and the antenna that is used. If we think about an asteroid at a distance of about 8 million kilometers, the resolution is as high as to allow details of 10 m to be seen. The measurement of velocities is even more accurate and may be determined in such a situation with a precision of about one-tenth of a millimeter per second.

This is highly valuable. We have seen in the previous chapters that using optical telescopes, even the largest ones on Earth, asteroids only appear as points of light showing no further details. For comets it is even worse since their nuclei are concealed by the coma through which we cannot see from Earth.

In contrast radar observations even allow studies of surface features of these small solar system bodies. Nowadays it is even possible to generate more or less accurate 3D models of an asteroid or comet if we have a sequence go radar observation of the respective object (see Fig. 6.6).

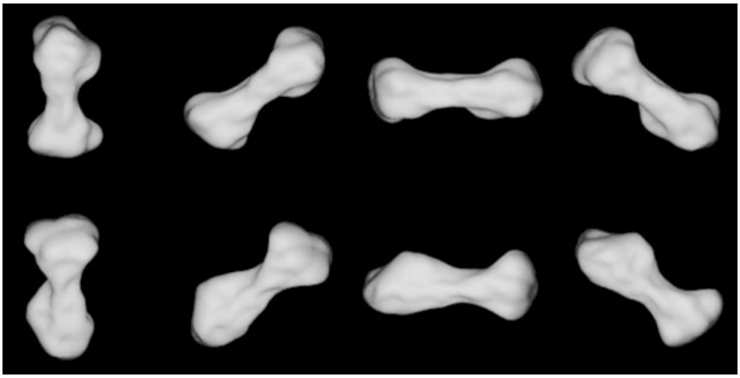


Fig. 6.6 These images show several views from a radar-based computer model of asteroid 216 Kleopatra (credit: NASA/JPL)

Surface Features: How Does It Work?
Determining surface features is, as you can imagine, more difficult than simply measuring the body's distance, velocity and direction it is moving to. Let us assume we are observing an asteroid. When the asteroid reflects the radar pulse we sent toy it, the echo of the pulse will blur over time. Why that? This is dependent on the structure of the asteroid as it is not a spherical body. The part of the overall signal reflecting off the nearest peak of the asteroid has to travel a shorter distance than the part that bounces off the farthest peak. The former one will thus return first.

We can now cut the echo into fine slices, e.g. each cut is made after a certain time such as 10^{-7} s. Through this we obtain a set of cross sections through the asteroid. The more powerful the echo of a slice is, the more of the asteroid is in the slice. The general idea of this approach is not much different from a CAT scan in medicine where a set of cross sections of the human body or some parts is obtained.

In addition, the asteroids rotation and Doppler shift exhibited but the asteroids movement the helps to determine the surface features. Algorithms can then be applied to obtain the aforementioned 3D.

6.3 Exploration by Spacecraft

The most fascinating and fruitful way of exploring the Small Solar System Bodies is of course by sending spacecraft to do the measurements and taking the photo on location. A spacecraft can get very close to the target object or even land on it to do in situ measurements. The results we can obtain like this would be impossible to get from Earth. Sometimes it is even possible to return samples from these tiny, distant objects. Two spacecraft already have returned samples. The Japanese spacecraft Hayabusa flew to the asteroid Itokawa, landed, took samples and returned them to

Earth. NASA's Stardust mission collected interstellar and cometary dust and brought them back to Earth.

How could we achieve this without spacecraft? Of course, there are also many drawbacks. The mission planning takes a very long time and such missions are very expensive. Several missions had been proposed and were then finally cancelled either in the planning phase or at a later point of time due to budgetary constraints. Furthermore, the more complex the mission and the more elaborated, fancy technology is involved, the higher the chances are that problems arise. Many missions that were or have been sent to small bodies in our solar system have experienced more or less severe technical problems. Some could be solved, some not.

In the following, we will have a brief look at various completed and still ongoing missions to small bodies in our solar system.

6.3.1 Rosetta

Rosetta is a currently ongoing mission to comet 67P/Churyumov-Gerasimenko. The spacecraft is the first manmade object that entered orbit of a comet and brought a lander on the surface of the comet's core. Although only preliminary results exist yet, these look more than promising and give a fascinating insight into the nature of a comet.

Mission Background

The first ideas of the Rosetta mission date back into the mid 1980s. When Halley's comet was on its way into the inner solar system heading rapidly towards its perihelion a whole armada of spacecraft, including ESA's Giotto and the Soviet Vega 1 and 2 probes, was ready to observe this icy snowball. This armada gained lots of new insights.

However, already then it became apparent that further, more specialized missions would be necessary in order to get better and more detailed results and answer all the questions that came up during the armada's operation.

In the 1980s three dominant space explorers existed: NASA, ESA and the Soviet Union. China, Japan, India and other nations that today play an increasingly important role in spaceflight, were negligible by then. However, the Soviet Union dropped out because of the political and economical turbulences that led in the end to the fall of the Soviet Union.

NASA and ESA, however, decided to cooperate. A nice side effect of the planned cooperation was saving money. NASA planned a mission, which they named Comet Rendezvous Asteroid Flyby (CRAF) and ESA a sample return mission—Comet Nucleus Sample Return (CNSR). Both missions were supposed to use the Mariner Mark II spacecraft design and thereby both agencies thought they could reduce their individual costs drastically.

Unfortunately, NASA cancelled their CRAF mission due to budgetary constraints in 1992. Shortly after, in 1993, ESA realized that their sample return mission was too ambitious for the foreseen budget. Nevertheless, the officials wanted to stick to a mission to a comet. Following this, the mission was redesigned. Similarities to NASA's CRAF mission were undeniably apparent. The new European mission included an asteroid flyby and a rendezvous with a comet. Both were the original mission targets of CRAF. In addition to this, the ESA's mission planners decided to incorporate a landing unit which should be brought onto the surface of the cometary nucleus.

The probe is named after the Rosetta Stone, a stele of Egyptian origin featuring a decree in three scripts. The lander is named after the Philae obelisk, which bears a bilingual Greek and Egyptian hieroglyphic inscription. A comparison of its hieroglyphs with those on the Rosetta Stone catalyzed the deciphering of the Egyptian writing system. Similarly, it is hoped that these spacecraft will result in better understanding of comets and the early Solar System.

Problems, Delays, and a New Target

The mission planning further proceeded and the space probe was built. Rosetta's original launch date was scheduled for January 13, 2003. Then, the mission was delayed due to unexpected reasons. The probe required a strong rocket that was capable of transporting the spacecraft's huge mass into space. ESA's newly developed Ariane 5 rocket was considered to be the first choice. No other European rocket that was powerful enough existed.

Ariane 5 made fast progress but then during its maiden flight on June 6, 1996 a dramatic incident occurred—the rocket exploded right after its launch together with its payload. The search for errors and fixing them also delayed the Rosetta mission.

The comet 46P/Wirtanen had been chosen as the spacecraft's target. However, due to the delays imposed on the mission by the Ariane 5 problems, the launch window for that comet closed. A new target had to be found.

67P/Churyumov-Gerasimenko (Chury) was chosen, a relatively young comet discovered in 1969. Scientists assumed that it had not been totally outgassed so that reasonably high activity could be expected on its journey to the Sun.

Chury has an orbital period of only 6.45 years and is assumed to have originated from the Kuiper belt or the scattered disk. It has undergone quite strong perturbations of its orbit due to the gravitational influence of Jupiter. Numerical studies revealed that Chury had a very close encounter with the gas giant in February 1959. Before this encounter its semi-major axis was at approximately 2.7 AU. Jupiter reduced it to about 1.3 AU during this encounter.

A new date for the start of the mission was set to February 26, 2004. However, bad weather conditions hindered the launch from taking place that day. Fortunately, soon after, weather improved and an Ariane 5 rocket with Rosetta onboard could finally lift off on March 2, 2004 at 07:17 UTC (see Fig. 6.7).

Brief Overview of Technology and Instruments

Rosetta was finally on its way to Chury. Yet, what does the probe look like? Which instruments are onboard and what is supposed to be achieved with these?

The technical details of Rosetta are manifold. A detailed discussion of these would clearly go beyond the scope of this book. However, for those interested in

Fig. 6.7 Rosetta successfully lifted off from Europe's Spaceport in Kourou, French Guiana, at 07:17 GMT on 2 March 2004 (credit: ESA/CNES/ARIANESPACE-Service Optique CSG)

more details, the reference section at the end of this book will provide some material.

Rosetta's bus is a 2.8 × 2.1 × 2.0 m central frame and aluminum honeycomb platform and weighs approximately 2.9 tons. This weight includes the Philae lander with about 100 kg and the scientific instruments with about 165 kg. The scientific payload had been mounted on top of the spacecraft.

The electrical power necessary for operating the spacecraft comes from two solar arrays. Both arrays together cover an area of about 64 m^2. The usage of solar arrays for a mission going beyond the orbit of Mars was a clear first in spaceflight. No other probe had tried this before. Most of the others that headed for the outer solar system (such as the Voyager and Pioneer probes) had been powered by nuclear batteries.

The solar arrays deliver energy dependent on the solar radiation that falls on them. At Chury's perihelion about 1500 W are expected, during Rosetta's hibernation phase, we will come to that soon, about 400 W and finally when the probe's full operations begin approximately 850 W.

Rosetta's scientific instruments can be ordered according to their foreseen purpose. A huge bunch of instruments are dedicated to the exploration of Chury's nucleus and coma.

Among these instruments is ALICE, an ultraviolet (UV) imaging spectrograph which is supposed to search for and quantify the noble gases, such as neon, argon and helium, in the nucleus. Using the determined abundances temperature estimates over the lifetime of the comet are possible, i.e. which temperatures had it been exhibited to since its formation.

A second instrument is named OSIRIS which stands for "Optical, Spectroscopic, and Infrared Remote Imaging System". It is equipped with two cameras. The first one has focal length of 140 mm and covers a wide field of view (about $12° \times 12°$). The second one with a 700 mm lens is able of taking images of a narrower field of view ($2.2° \times 2.2°$). Both cameras are equipped with a 4 megapixels sensor. OSIRIS' main purpose is to obtain detailed surface maps of Chury. When Rosetta is at its closest approach to the comet (~1 km), the image resolution is about 2 cm per pixel (see Fig. 6.8). It was also used to determine a landing site for Philae.

The "Visible and Infrared Thermal Imaging Spectrometer" (VIRTIS) is designed for taking images of Chury's nucleus in the infrared and also to search for infrared spectra of molecules in its coma.

MIRO, the "Microwave Spectrometer for the Rosetta Orbiter"), is used for detecting the abundances and temperatures of volatile substances like water, ammonia and carbon dioxide via their microwave emissions.

Another instrument for observing the nucleus is CONSERT ("Comet Nucleus Sounding Experiment by Radiowave Transmission"). The CONSERT experiment will provide information about the deep interior of the comet using radar. The radar will perform tomography of the nucleus by measuring electromagnetic wave propagation between the Philae lander and the Rosetta orbiter through the comet nucleus. This allows it to determine the comet's internal structure and deduce information on its composition.

The last instrument dedicated to the exploration of nucleus and core is RSI ("Radio Science Investigation"). It makes use of the probe's communication system for physical investigation of the nucleus and the inner coma of the comet.

The four following instruments and related experiments deal with the exploration of gas and dust particles of Chury. ROSINA, the "Rosetta Orbiter Spectrometer for Ion and Neutral Analysis", comprises two mass spectrometers which shall detect the composition of Chury's faint and almost non-existent atmosphere.

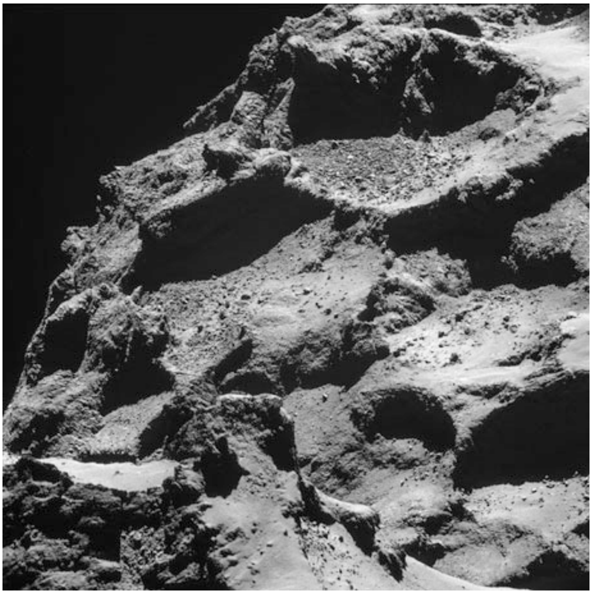


Fig. 6.8 Image of comet 67P/Churyumov–Gerasimenko at less than 10 km from its surface (credit: ESA/Rosetta/NAVCAM, CC BY-SA 3.0 IGO)

The "Micro-Imaging Dust Analysis System" (MIDAS) is a high-resolution atomic force microscope which investigates several physical aspects of the dust particles which are trapped on a silicon plate.

Of particular relevance may be the composition of dust particles. This will be measured by a mass spectrometer included in COSIMA (Cometary Secondary Ion Mass Analyser).

The last instrument is GIADA, the "Grain Impact Analyser and Dust Accumulator". GIADA analyzes the dust environment of the cometary coma by measuring the optical cross section, momentum, speed and mass of each grain entering inside the instrument.

Rosetta's Trajectory or How Do We Get to the Comet?

The Ariane 5 rocket launched and brought its payload (Rosetta) into space. But how does Rosetta get to the comet? There is currently no rocket available that would be powerful enough to send a spacecraft weighing about 3 tons directly to a comet. Even if such a rocket existed, it would not be feasible to use such a transporter for economic and energetic reasons. A vast amount of fuel would be required for this.

Instead, a more economic but also slower approach was chosen for Rosetta: a series of so-called swing-by maneuvers. In general, swing-by maneuvers, sometimes also called gravity assist, work according to the following principle. A relatively lightweight spacecraft, such as Rosetta, passes a much more massive object, usually a planet. The gravitational influence of the planet then changes the spacecraft's flight trajectory. During this alteration also the velocity of the vehicle may be reduced or increased. As a side note we should mention that of course, the angular momentum has to be preserved and thus also a force is exerted onto the

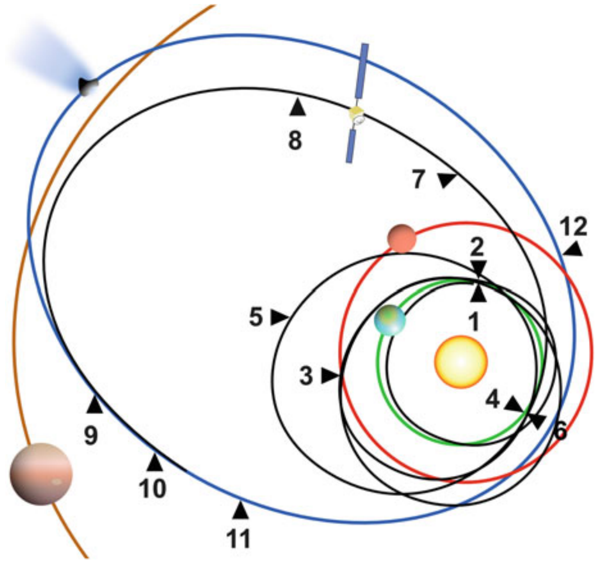


Fig. 6.9 Rosetta's trajectory on its way to comet 67P/Churyumov–Gerasimenko (credit: Kulandru mor, CC0 1.0)

planet. However, due to the spacecraft's very low mass compared to the mass of the planet, the effect is negligible.

Figure 6.9 depicts the series of swing-bys used to bring Rosetta to Chury.

Position (1) in Fig. 6.9 shows the launch when the Ariane 5 rocket brings Rosetta into a near-Earth orbit. About 1 year later on March 4, 2005, Rosetta's first swing-by maneuver takes place. Rosetta approaches Earth as close as 1900 km distance (see position (2)). The probe's trajectory is modified and it moves away from Earth and finally crosses the orbit of Mars (position (3)). A very close swing-by at Mars occurs on February 28, 2007. During this passage, Rosetta is only about 250 km away from Mars' surface and slows down by this maneuver.

On November 13, 2007, Rosetta encounters Earth again at a distance of 5296 km (position (4)).

Visit to Steins and Back to Earth Again

The last swing-by at Earth sends Rosetta directly to asteroid (2867) Šteins (see Fig. 6.10) which it passes on September 5, 2008 (position (5)). The passage of the 4.6 km diameter asteroid is extremely short lasting only about 7 min and occurring at a distance of about 800 km. Nevertheless, there was enough time to obtain some results and take many photos. This short flyby was also used to test and calibrate the scientific instruments on board of Rosetta.

Rosetta was able to determine Šteins' light curve and derive a rotational period of about 6 h. Images show that Šteins has a very irregular shape almost looking like a brilliant in the sky. No moons could be detected.

After the short trip to the asteroid, Rosetta returned to Earth for the next swing-by at November 13, 2009 (position (6)).

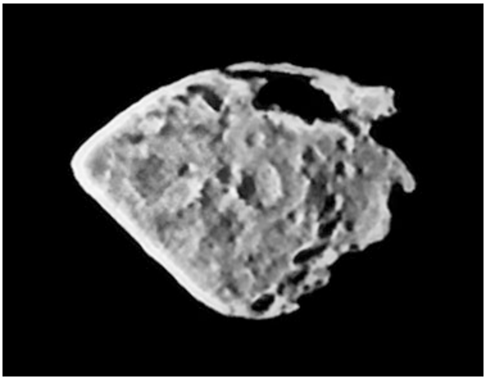


Fig. 6.10 Asteroid 2867 Šteins as imaged by ESA's Rosetta spacecraft using the OSIRIS camera on 5 September 2008 (credit: ESA)

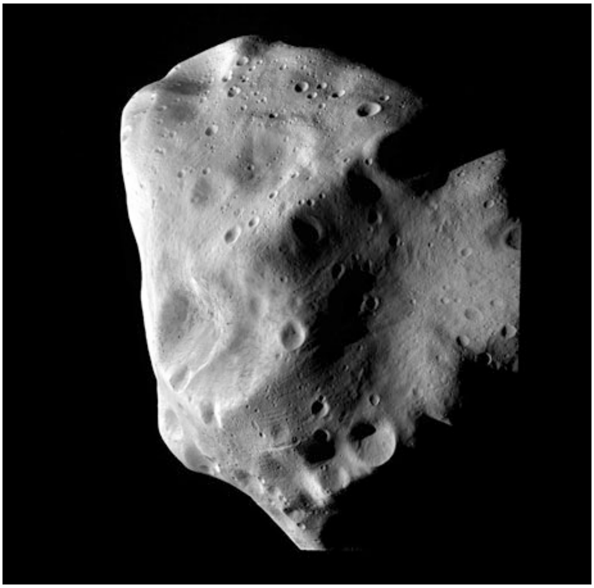


Fig. 6.11 Asteroid Lutetia at closest approach of Rosetta (credit: ESA 2010 MPS for OSIRIS Team MPS/UPD/LAM/IAA/RSSD/INTA/UPM/DASP/IDA)

Visit to (21) Lutetia

Next stop on its journey was the asteroid (21) Lutetia (position (7)) which Rosetta passed on July 10, 2010 in a distance of 3162 km. Figure 6.11 depicts the asteroid. Lutetia is about 100 km in diameter and hence much larger than the previously visited asteroid Šteins.

Lutetia was discovered by the German-French astronomer Hermann Goldschmidt (1802–1866) from the balcony of his apartment in Paris on November 15, 1852.

The flyby provided a huge number of images which had resolutions of up to 60 m per pixel. Rosetta photographed approximately 50 % of Lutetia's surface and mostly the northern hemisphere as this was in a better position when Rosetta flew by.

Lutetia is an inner asteroid belt object having a semi-major axis of 2.4 AU. It is hence located slightly closer to the Sun as the Kirkwood gap at 2.5 AU and lies almost completely within the ecliptic plane. Its orbital period is 3.8 years.

The images obtained from this irregularly shaped asteroid reveal about 350 craters ranging from 600 m to about 55 km in diameter. Most of the detected craters lie in the northern hemisphere.

Hibernation: Sleep Well Rosetta

In July 2011 Rosetta was put into hibernation, the so-called Deep Space Hibernation. The sleep was planned to last for about 31 months from July 2011 to January 2014 (positions (8) and (9)). This was done in order to save energy resources and thereby to extend the lifetime of the spacecraft. Just before it was set asleep, Rosetta was able to have a first and for a long time last glimpse at its destination. It briefly observed Chury from a distance of about 1 AU in March 2011.

Time to Wake Up and Arrival at the Comet

After the 31 months had passed, the Rosetta control center sent a wake up signal to the spacecraft on January 20, 2014 (position (9)). The scientists excitedly awaited the response? Would it wake up after such a long time? The feeling of relief was great, when hoped for confirmation signal of Rosetta was received in the control center.

In May 2014, the spacecraft was slowed down and the distance to Chury was drastically reduced from about 2 million kilometers at the beginning of the month to about half a million kilometers.

Location (10) marks the position when Rosetta entered Chury's orbit. On August 6, 2014 Rosetta overtook Chury and stopped approximately 100 km in front of it when it then entered a hyperbolic orbit around the comet. Already there the gravitational influence exhibited by the comet's nucleus was clearly measurable by the probe.

In mid-September this orbit was changed into an elliptical one with a mean distance of 30 km to the cometary nucleus. Until October 10, 2014 the distance was further decreased to about 10 km.

Since August 2014, Rosetta took detailed images of the comet's nucleus whose surface layout had been completely unknown prior to the arrival of the spacecraft. Besides gaining insight into the structure of the nucleus and its surface, the mapping was also essential for determining a proper and secure landing site for Philae.

On August 25, 2014 ESA announced five potential landing sites (see Fig. 6.12) of which site "J" was finally chosen on September 15, 2014.

It is Time to Land

On November 12, 2014 at 8:35 UTC (position (11)) Philae detached from Rosetta and started its descent to the nucleus. Several images were taken by Rosetta to document Philae's departure (see Fig. 6.13).

The lander finally touched down at 15:33 UTC. Two harpoons had been supposed to fire upon landing and thereby anchoring the lander on the surface. Gravity is so low that without additional means the lander would bounce back into space.

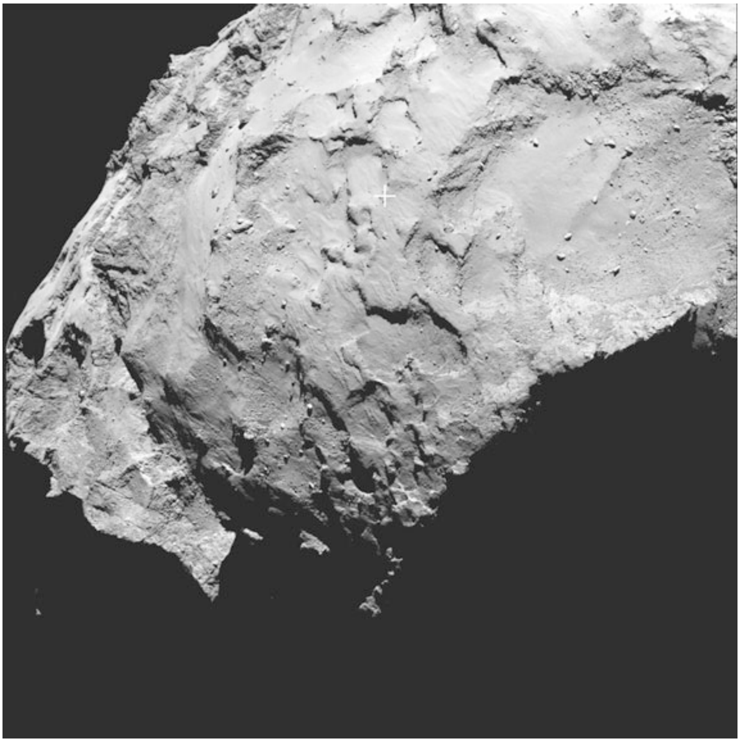


Fig. 6.12 Philae's primary landing site on the comet (credits: ESA/Rosetta/MPS for OSIRIS Team MPS/UPD/LAM/IAA/SSO/INTA/UPM/DASP/IDA)

Fig. 6.13 This image shows the Philae lander at 10:23 GMT on 12 November. The image shows details of the lander, including the deployment of the three legs and of the antennas (credit: ESA/Rosetta/MPS for OSIRIS Team MPS/UPD/LAM/IAA/SSO/INTA/UPM/DASP/IDA)

However, something went wrong. The harpoons did not fire for unknown reasons. Philae bounced twice from the surface before it finally came to rest on 17:33 UTC.

Unfortunately, its final landing site, which remains unknown until now, was not a favorable one. The final location has been determined within an accuracy of a few hundred meters by data obtained by CONSERT. Images taken by the lander itself suggest that Philae was standing somewhere in a shadowed region, perhaps a crater or gap. This meant, however, that contrary to what had been planned, the lander would not receive enough solar radiation to keep its batteries working. It was solely depending on solar energy. The lander had two batteries. The primary battery had a capacity that was sufficient to operate the lander and its experiments for about 60 h. The secondary battery had originally been planned to be recharged by solar energy. Hence, without recharging the lander would cease operation after a few hours.

Yet, this was not the only problem. The photos also revealed that Philae was not standing even with the ground but appeared to lean in a 30° angle against a rocky structure with two of its four legs up in the air. The scientists hoped that it would still be possible to carry out the core experiments.

Philae Delivers and Is Repositioned

Indeed, some results could be obtained. The landing site had a different composition. The ground was assumed to be soft and fluffy. Yet, it turned out to hold a large amount of water ice under a thin layer of dust.

Furthermore, organic molecules, carbon and hydrogen in Chury's atmosphere could be detected. The mission had foreseen that Philae would drill into the ground and retrieve some sample from below the surface. However, the lander was not able to drill into the surface, most likely due to hard ice.

After most of the core experiments had been carried out and the battery life tended towards its end, the scientists tried to slightly change Philae's orientation and thereby putting its solar arrays into a more favorable position. Apparently, this succeeded. The newly gained energy, however, was not enough to allow further operation. Consequently, the electrical power available dwindled rapidly and all instruments were forced to shut down. Contact was finally lost on November 15, 2014 at 00:36 UTC.

Will It Wake Up Again?

Yet, this might not be Philae's end. It had been designed to have the possibility to go into such a sleep mode. If the comet gets closer to the Sun, the lander will be able to capture more energy. If it had not been damaged during its sleep, it may be able to reactivate it as soon as the secondary battery has been sufficiently recharged. Several attempts so far, the last one being in March 2015, have failed however. Time will show whether the little brave lander will wake up again and continue its experiments.

First Results of Rosetta

One of the first findings of Rosetta was an apparent magnetic field oscillating at 40–50 mHz. However, Philae showed that Chury did not have a magnetic field. So, how do these contradictory results match? It took some time until the scientists came up with a solution to this problem. The original detected field was indeed not a magnetic field of the comet but was the result of interactions with the solar wind.

Rosetta further revealed data about Chury’s composition. In particular, water vapor escaping from its surface was analyzed. What was the ratio of deuterium and hydrogen? This question is relevant since previous theories suggested, as we have seen in Chap. 3, that most of Earth’s water originated from comets. The findings of Rosetta confirmed earlier results that had been determined with the comets Hale-Bopp and Hyakutake. The ratio was substantially different compared to Earth’s water. Consequently, comets cannot be hold accountable for the majority of the water available on Earth. Other sources must have existed.

Chury’s shape had been originally, after first images taken from a large distance, described to resemble a baked potato. Closer photos make it, however, more look like a rubber duck. The cometary nucleus consists of potentially two solid bodies loosely held together.

Chury’s mass has been determined to be 1013 kg $\pm$ 10 % and thus about three times higher than originally assumed. The findings of Philae have also been confirmed in that the nucleus is covered by an about 20 cm thick layer of dust having beneath this layer hard ice or at least a mixture of ice and dust. Furthermore, the porosity appears to increase towards the center of the comet.

By far more results are expected with increasing cometary activity while Chury is further approaching the Sun.

6.3.2 Dawn or the Story of Two Small Solar System Bodies

Dawn is the first spacecraft of NASA’s Discovery Program. The spacecraft’s purpose is to study two potential protoplanets in the main asteroid belt: the asteroid (4) Vesta and the dwarf planet Ceres. There are several mission firsts. Dawn is the first spacecraft to orbit two extraterrestrial bodies. In addition, it is also the first one to visit both Vesta and Ceres. No other spacecraft has visited a dwarf planet so far. Dawn with its arrival at Ceres on March 6, 2015 is the first one. Currently, the spacecraft New Horizons is on its way to dwarf planet Pluto but will only receive it in July 2015. Besides these aspects, Dawn is also the first mission to use ion propulsion as the only means of impulsion (Fig. 6.14).

Mission Background

The Dawn mission has turbulent history with several ups and downs accompanied by technical problems. The project was cancelled in December 2003 but soon after, in February 2004 reinstated.

In November 2005, work on the spacecraft was stopped again by the Jet Propulsion Laboratory (JPL), the organization responsible for the mission, due to budgetary problems. It had become apparent that the costs of the mission and particularly for building the spacecraft would bust the agreed budget.

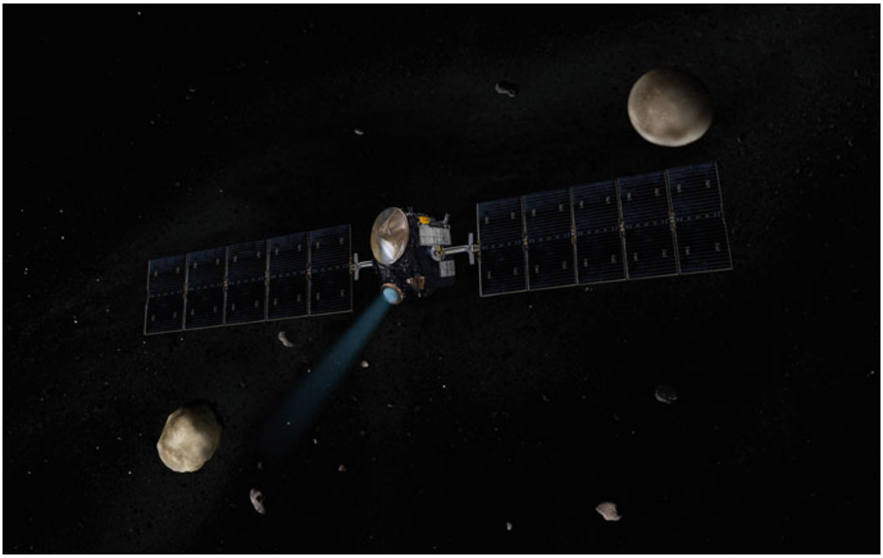


Fig. 6.14 Artist's concept of the Dawn spacecraft (credit: NASA/JPL-Caltech)

The JPL informed the public that a final decision whether the mission would be continued or stopped was to be expected at the beginning of 2006. In March 2006, it was announced that the mission was finally cancelled in spite of the spacecraft had almost been completely manufactured.

Subsequently, the manufacturer of Dawn, the Orbital Sciences Corp., appealed the decision and offered to build the spacecraft ay cost, foregoing any profit in order to gain experience in a new market field.

Consequently, the Dawn mission was reinstated on March 22, 2006 and launch preparations continued. Some considered these permanent ups and downs as a bad omen for the mission. Indeed, although the administrative and budgetary problems seemed overcome, the mission should turn out to be far from being free of any further problems. The problems simply changed their nature to technical ones as we will see soon.

Further Delays and Start of the Mission

The original mission plan foresaw that Dawn should be launched as a payload of a Delta II rocket in July 2007. However, the preparations had to be stopped on June 11, 2007. A tool fell down and damaged one of the solar panels. This was seen as critical since the launch window that made a journey to both targets possible was only open until October 2007. A serious damage and its subsequent repair might have endangered a timely launch. Fortunately, the problem could be fixed easily causing almost no delay. The mission team breathed a sigh of relief. Perhaps too early. The mission was indeed delayed but for other reasons.

One reason was to avoid potential conflicts with the Phoenix mission to Mars which was supposed almost at the same date. The launch of Dawn was hence

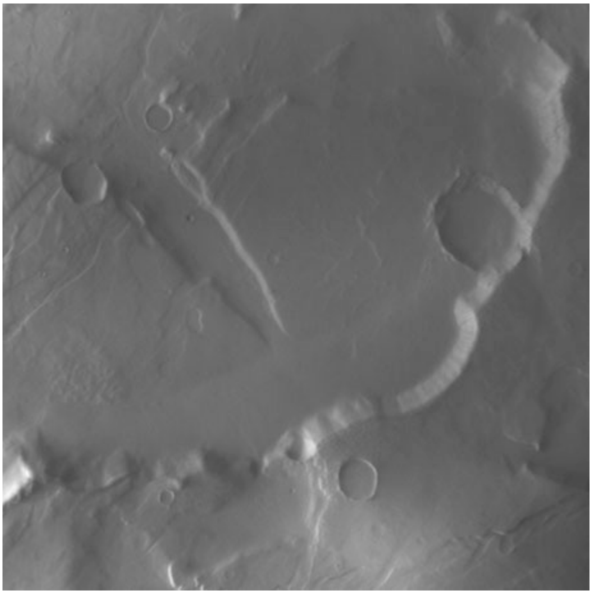


Fig. 6.15 Near-infrared image from the framing camera on NASA's Dawn taken near the point of closest approach to Mars on Feb. 17, 2009, during Dawn's gravity assist flyby (credit: NASA)

rescheduled for September 26, 2007. Phoenix was successfully launched on August 4, 2007. Everything was ready for Dawn which already awaited the start of its adventures on the launch pod. Yet, luck was not with the mission again. Bad weather conditions on the scheduled date made a launch impossible. The day after, on September 27, everything was fine and Dawn was launched.

From Earth to Vesta

The Delta II rocket successfully brought Dawn into space. The first 81 days after the start were used by the mission team to check Dawn's functionality and the scientific instruments for errors.

Then, after these extensive checks, one of Dawn's ion thrusters ignited and the spacecraft set out course to Vesta on December 17, 2007. A swing-by maneuver was planned at Mars in order to adapt Dawn's trajectory and gain the necessary speed to reach its first destination, Vesta. The spacecraft passed Mars with a distance of only 549 km on February 11, 2009 (see Fig. 6.15). On very same day, the spacecraft placed itself in safe mode, resulting in some data acquisition loss. The spacecraft was reported to be back in full operation 2 days later, with no impact on the subsequent mission identified. The root cause of the event was reported to be a software programming error.

Approaching Vesta

After leaving Mars, Dawn traveled in an elongated outward spiral trajectory continuously towards the asteroid Vesta. On May 3, 2011 the spacecraft managed to take a first photo of its destination from a distance of 1.2 million kilometers to the asteroid. The final approach phase began.

On June 12, 2011 Dawn's speed relative to Vesta was slowed down in order to prepare entering an orbit of Vesta about 1 month later. Indeed everything worked out fine and Vesta captured Dawn into orbit on July 16, 2011.

Dawn's initial orbital altitude was set to 2750 km above the surface of the asteroid. This was close enough to obtain first detailed images of Vesta's surface and to spectrally analysis its surface composition. We will come back to that in a moment.

Yet, this was not close enough. The mission team lowered Dawn's orbit to about 600 km above the asteroid and started mapping Vesta's surface by using stereo images.

The Journey to Ceres: A Whole Clutch of Problems

Due to the huge amount of information that had already been gathered by the spacecraft, the Dawn mission team decided to extend the originally planned stay to August 26, 2012. However, technical problems caused another delay, so Dawn could only set sails for Ceres on September 5, 2012.

The technical problems did not stop after leaving Vesta but continued. Some of the seriously endangered the whole mission. On September 11, 2014, for example, Dawn's ion thruster suddenly ceased to fire which triggered the spacecraft to go into safe mode. Again, some scientific data was lost. The cause for the problems could not be identified. Thus, the team decided to exit the safe mode 2 weeks later on September 25, 2014. Astonishingly, the ion thruster went back to normal operation without any signs for the previous problems.

The main communication antenna caused further headaches with the mission team. It turned out that unexpectedly problems aiming the antenna towards Earth arose. This endangered the communication with the control. To overcome the problem preliminarily and to ensure contact with the spacecraft another antenna with weaker capacity was temporarily used.

The engineers feverishly search for a solution to this antenna problem. Using the weaker antenna temporarily enabled communication with the spacecraft. However, with increasing distances this would get more difficult. In addition, due to its weaker capacity, the bandwidth for sending scientific data back to Earth was very limited. Much data was in danger to get lost.

The mission engineers decided to reset the system which turned out to be a good solution. The main antenna worked normal again. So far, no further problems with neither the thrusters nor the antenna have occurred until the current day.

Approaching Ceres: First Encounter with a Dwarf Planet

On December 1, 2014 Dawn was able to photograph the extended disk of Ceres for the first time. The resolution further improved with each single day Dawn came closer to Ceres. Since the end of January 2015, the images taken by Dawn clearly the resolution of images taken by the Hubble Space Telescope (HST) which had delivered the best results so far in 2003 and 2004 (see Fig. 6.16).

Dawn entered into the orbit of Ceres on March 6, 2015. Upon arrival, the spacecraft took the first full topographic map of the dwarf planet's surface. On April 23, Dawn has reached an orbital altitude of 13,500 km but continues to spiral

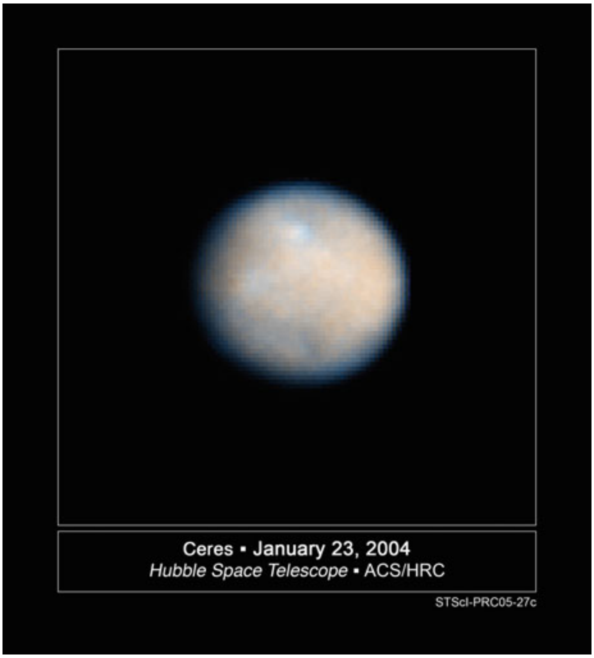


Fig. 6.16 This is a NASA Hubble Space Telescope color image of Ceres, the largest object in the asteroid belt. The observations were made in visible and ultraviolet light between December 2003 and January 2004 (credit: NASA, ESA, J. Parker (Southwest Research Institute), P. Thomas (Cornell University), L. McFadden (University of Maryland, College Park), and M. Mutchler and Z. Levay (STScI))

downwards to a planned survey orbit of about 4430 km. It will then use its framing camera to create a surface map of the whole dwarf planet in a very high, never seen before resolution.

Spacecraft Design and Technology

Dawn is a relatively large space probe with a main bus of about 2.36 m with the solar arrays being retracted. When the solar arrays are fully extended, the size increases to 19.7 m. The solar arrays in the latter state cover about 36.4 m^2. Compared to other probes, such as Rosetta for example, Dawn is lightweight with its mass of 1108 kg.

Dawn is equipped with several scientific instruments. The first one is the so-called "Framing Camera" (FC) which basically consists of two identical cameras due to redundancy issues. Each of them has an aperture of 20 mm and a focal length of 150 mm. They are used for surveying and mapping the surfaces. Most of the images used in this book were taken by the Framing Camera.

The second instrument is the "Visible and IR Spectrometer" which is a modification and improved version of the visible and infrared thermal imaging system used in Rosetta.

The "Gamma Ray and Neutron Detector" (GRAND) comprises 21 sensors with a very wide field of view. It is used for measuring the abundances of major rock-forming elements such as oxygen, magnesium, aluminum, silicon, calcium, titanium and iron. By using GRAND, it is possible to determine Vesta's and Ceres' detailed chemical composition and internal structures.

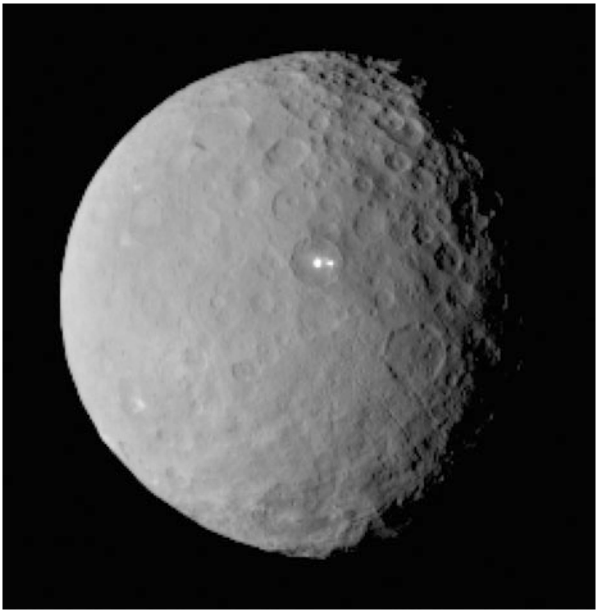


Fig. 6.17 Dwarf planet Ceres taken by Dawn on February 19 from a distance of nearly 46,000 km. It shows that the brightest spot on Ceres has a dimmer companion, which apparently lies in the same basin (credit: NASA/JPL-Caltech/UCLA/MPS/DLR/IDA)

We have said it before, Dawn is the first spacecraft to completely rely on ion propulsion for moving in space. It is propelled by three xenon ion thrusters. Instead of a spacecraft being propelled with standard chemicals, the noble gas xenon is given an electrical charge, or ionized. It is then electrically accelerated to a speed of about 30 km/s. When xenon ions are emitted at such high speed as exhaust from a spacecraft, they push the spacecraft in the opposite direction due to the conservation of momentum.

First Scientific Results

Dawn has obtained so much data that it will take some time to proper analysis it. With regard to Vesta, Dawn was able to verify the asteroid's differentiated interior, as we also discussed it in Chap. 2. Data from Dawn suggests that Vesta's core has a diameter of ably 220 km. Vesta appears to be the only remaining protoplanet left over from the time of the formation of the solar system. Dawn was also able to finally reveal the origin of the anomalous dark spots and troughs on Vesta's surface. Only the detailed images taken by the spacecraft allowed to determine that these structures apparently are the consequence of impacts, in particular the giant impacts that created the two large craters near Vesta's South Pole, Rheasilvia and Veneneia.

For Ceres, it is still too early to have proper results. However, while approaching Ceres, Dawn discovered two bright spots on its surface that appeared to rotate with the dwarf planet (see Fig. 6.17). The nature of these spots is still unknown. The brightest spot was still too small to be resolved. The dimmer second spot, however, could be more closely examined. The spot appeared to be located in the middle of a 80 km wide crater. The nature of the spots is still unknown.

Dawn was able to observe that the spots brightened during daytime and got fainter at dusk. This suggests that they are at least somehow interacting with solar radiation. Further research analysis is necessary in order to solve this mystery.

6.3.3 New Horizons: A Journey to the Outer Solar System

The New Horizons mission, the first spacecraft in NASA's ambitious New Frontiers program, is special in a few respects. Once started to explore the last unvisited planet in our solar system, Pluto, it is now on course to visit a dwarf planet, Pluto, and continue its journey to yet unknown destinations. Only one thing is clear 1 day it will end up in the infiniteness of our Universe.

Mission Background

In the early 1990s first ideas for a mission to Pluto, this strange and exotic world at the fringe of our planetary system, came up. By that time, Pluto had still been the last unexplored planet in our solar system. Due to its small size and huge distance to Earth it had been hardly possible to obtain meaningful results. It was almost not possible to resolve Pluto's disc in even the most powerful telescopes available. Most information was gathered by spectral analysis and photometry. Basic surface structures were derived from this data. However, nobody knew how it actually really looked like on this small world. Probably the only way to obtain detailed results and to unveil its secrets was by sending a spacecraft to it.

The scientists were in a hurry to set up a mission as soon as possible. Pluto had reached its perihelion in 1989 and was now moving farther away from the Sun. Many things were unknown with regard to Pluto. The scientists wished to able to also observe Pluto's thin and somehow mysterious atmosphere. We have seen it before, Pluto's atmosphere is different from our Earth's atmosphere or any of the atmospheres of the major planets in our solar system. Pluto's atmosphere is formed by sublimation of ices, such as methane. On the one hand, when the dwarf planet is approaching its perihelion, temperatures on its surface increase and at some point of time sublimation can take place. On the other hand, after its perihelion Pluto departs from the Sun and heads for the distant regions of its orbit. With increasing distance, temperatures fall and the atmosphere finally freezes out. Pluto is left without a measurable atmosphere. Therefore, the scientific community was eager to reach Pluto before it lost his atmosphere. Waiting for its next perihelion was not an option. This will only be in the year 2247.

NASA called for concepts of a Pluto mission in the early 1990s. Two proposals were short-listed: Pluto Fast Flyby and the Pluto Kuiper Express. However, both of them failed for technical and financial problems. The Pluto Kuiper Express was eventually cancelled in 2000 mostly for budgetary reasons. A new round of call of concepts was announced. Several mission proposals were handed in. After a 3 month period of evaluation the New Horizons concept was accepted on June 8, 2001. The proposing team was asked to provide a more detailed preliminary design study for a flyby mission to Pluto. The mission was finally approved in November 2001.

Planet or Not Planet: A Dispute

Back then, when the mission was officially started, Pluto still had its planet status. During the course of the New Horizons mission, this status was stripped off from Pluto. The former planet was demoted to a dwarf planet. Some project members, above all the mission's principal investigator Alan Stern disagreed with the new classification. Some of the team members still describe Pluto as a planet. Doubts on the three criteria defined by the International Astronomical Union (IAU) in 2006 were raised. Criteria (I) and (II), namely that the candidate body had to orbit the Sun and be in hydrostatic equilibrium, were basically accepted. However, the third prerequisite was considered to be problematic. This required that the planet candidate was dominant enough to have cleared its neighborhood, i.e., removed other objects from its orbit. The mission team and also other scientists raised their voices that most planets would fail here. They argued that other objects were present in the planetary orbits. In particular they pointed to Trojan asteroids that are co-orbiting all planets, except Mercury and Saturn. The debate is still going on. However, the vast majority of the astronomical community has accepted the new IAU classification scheme in the meantime.

Scientific Goals: What Is Expected from the Mission?

The New Horizons spacecraft will starting from 2015 enter mostly unknown and unexplored territory. The overall scientific goal is therefore, to understand better the formation of the Pluto system, the Kuiper belt, the Scattered Disk and to gain more insight into the evolution of the early solar system. The probe is supposed to study Pluto's and Charon's atmosphere, surface, interior and environment of Pluto's moons. After the flyby of the Pluto–Charon system other Kuiper belt objects (KBOs) will be explored.

The main objectives hence are the following. Firstly, New Horizons will map the surface composition of Pluto and Charon. Today, only very limited information is available giving only rough ideas of what is actually going on there. Secondly, the probe will try to characterize the geology and morphology of Pluto, Charon and selected KBOs. Thirdly, not much is known about Pluto's atmosphere. Hence, New Horizons will analyze the dwarf planet's atmosphere and search for an atmosphere around Pluto which is predicted by some theoretical models. Fourthly, instruments will map the surface temperature on Pluto and Charon.

The fifth objective is relatively new. New Horizons will search for rings around Pluto and other yet unknown moons. It was only recently, that rings were detected with two small solar system bodies: the centaurs Chiron and Chariklo. Originally being KBOs or scattered disk objects, this might hint that other objects have similar structures. Pluto is one of the most massive KBOs. Therefore, the probability might be quite high.

Sixthly, after the flyby of Pluto and Charon, New Horizons will start to investigate a few other KBOs.

Fig. 6.18 Artist's concept of the New Horizons spacecraft during its planned encounter with Pluto and its moon, Charon (credit: Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute (JHUAPL/SwRI))

Spacecraft Design and Technology

To achieve all these goals, the design of New Horizons was specially chosen and specific instruments that fit this purpose were selected or developed.

The body of New Horizons forms a triangle of dimensions 0.7 m $\times$ 2.1 m $\times$ 2.7 m. The main structure column is a 7075 aluminum alloy tube between the launch vehicle adapter ring at the rear end and the 2.1 m radio dish antenna affixed to the front side (see Fig. 6.18).

In general, the structure is larger than actually would be necessary. There is quite a bit of empty space in between. Why should you do that? Is the volume not a critical mission aspect? Indeed it is, the more voluminous an object is there more expensive and difficult it is to send it into space. However, with New Horizons the structure is especially designed to act as a shielding, reducing electronics errors caused by radiation from its own radioisotope thermoelectric generator (RTG)—its engine so to speak. The spacecraft comprises 16 thrusters.

The Problem of Data Communication

The scientific instruments onboard New Horizons, we will come to them in a minute, first need to record their measurement data to a solid-state buffer and are then only much later on transmitted to Earth. The buffer consists of a primary disk and a backup disk each having a capacity of 8 GB. The delay between gathering the data and its transmission is construction-conditioned. All of the spacecraft's instruments are body mounted and cannot be moved. This means, however, that when one of the cameras wants to take pictures, the whole probe has to be moved in order to point the camera lens towards the target. Most likely the antenna used for data transmission will then no longer be pointed towards Earth and hence communicated data would be lost. Therefore, the results are first written into the buffer and as soon as a data communication can be established again transmitted to Earth.

You may ask why this inflexible design? It is true that previous probes often had rotatable instrumentation platforms which could be moved in a way such that the instruments could be pointed to almost any arbitrary point in space without losing the communication to Earth.

Though being a very flexible approach, the likelihood of a jam or other technical problems due to rotation mechanism were quite high. The New Horizons approach provides for an improved reliability and saves costs and weight, as no additional mechanisms are required.

Scientific Instruments

There are seven instruments onboard the spacecraft. All of them are used to achieve the mission objectives. We will briefly go through them in order to better understand the mission.

The "Long Range Reconnaissance Imager" (LORRI) is designed to take images with a long focal length. The camera's aperture is 208.3 mm and provides a resolution of roughly 1 arcsecond. It is especially designed for high resolution images.

The second instrument is the "Solar Wind At Pluto" (SWAP) device which will measure charged particles escaping from Pluto's atmosphere and are then torn away by the solar wind. It will provide insight whether Pluto has a magnetosphere and allow analysis of the structures of the solar wind.

PEPSSI, the "Pluto Energetic Particle Spectrometer Science Investigator", basically is a ion- and electron spectrometer that searches for neutral atoms escaping from Pluto's atmosphere and are subsequently charged by interactions with the solar wind.

Alice is an ultraviolet (UV) imaging spectrometer which makes observations in the far and extreme UV range (50–180 nm). This instruments helps to determine Pluto's atmospheric composition by detecting emissions from there in the "airglow" mode. A second mode is available, the so-called "occultation mode" which will create artificial stellar occultations. To do this, the instrument is pointed at the Sun or another bright star through Pluto's atmosphere. By this, it will become possible to further study the dwarf planet's atmosphere.

Another optical instrument is Ralph, a telescope with 6 cm aperture which will be used for taking photos. Its ultimate purpose is to make colored maps of Pluto's and Charon's surface.

The SDC (Student Dust Counter) is sometimes also called the Venetia Burney SDC to honor the girl who proposed the name Pluto for Clyde Tombaugh's discovery in 1930. It was built by students at the University of Colorado Boulder and is designed to measure dust particles continuously throughout the probe's trajectory.

The "Radio Science Experiment" (REX) is a radio wave experiment carried out by the spacecraft's main antenna. After the flyby of Pluto signals will be sent to New Horizons from the antennas of the Deep Space Network distributed all of the Earth. During the transit of the signal through Pluto's atmosphere, the signal will be altered and this can be detected by the probe. This allows to derive characteristics of the composition of the atmosphere.

Mission Timeline

We have seen a lot on the mission's background and technology so far. However, how does the mission look like? What has happened so far?

The schedule for the mission prior to its launch was quite strict. A launch window only existed from January 11, 2006 to February 14, 2006. However, only until February 2 it was possible to use a swing-by maneuver at Jupiter. After that date only a direct route to Pluto was possible which would mean an increased travel time of several years. The mission team tried its best to get the spacecraft up into space before this date.

The launch was initially scheduled for January 17, 2006 but had to be postponed due to bad weather conditions. The next day (18.1) was planned but could not be hold either due to a blackout at the control center. The launch, hence, was rescheduled to the following day. The cloudy sky asked for several attempts until New Horizons was finally launched mounted on a Atlas V rocket at 19:00 UTC.

Transit to Jupiter

After the Atlas V rocket set the probe out into the space, several trajectory correction measures were made, as planned, in order to bring New Horizons on course towards Jupiter for its swing-by.

On April 7, 2006 at 10:00 UTC New Horizons passed the orbit of Mars and headed towards the main asteroid belt. No flybys of asteroids had been planned. The mission team wanted to save energy for the exploration of Pluto and the Kuiper belt. Nevertheless, they started to look for an asteroid that might come close to the probe by chance. Indeed they found out that the spacecraft would pass close to the asteroid (132524) APL while cruising through the main belt. (132524) APL is a tiny asteroid with a diameter of about 3–5 km.

On June 13, 2006 the flyby occurred and Ralph was used to photograph the asteroid and to make measurements. The light curve of the asteroid could be determined and photos were taken. Why Ralph and why not LORRI? The latter one provides higher resolution images? LORRI could not be used, as the spacecraft was still too close to the Sun. The sensitive sensors of LORRI, particularly designed for the regions in the outer solar system, could have been destroyed by the strong sunlight. All in all a successful test for New Horzions' instruments.

Encounter with a Giant
The flight continued and the spacecraft left the main belt. On September 4, 2006, LORRI was used for the first time and a image of Jupiter, New Horizons' next stop, was taken from a distance of about 291 million kilometers or roughly 2 AU. A bit later between September 21 and 24, 2006, the first image of Pluto could be taken from a distance of 4.2 billion kilometers to 28.07 AU. The images showed nothing more than a faint, bright spot hardly distinguishable from the background stars.

The first real scientific investigation of Jupiter and its moons began in January 2007 and lasted until the end of June 2007. The spacecraft passed Jupiter closest at a distance of 2.3 million kilometers on February 28, 2007. This was not far beyond Callisto's orbit. Callisto is the outermost of the four Galilean moons of Jupiter.

The mission team had decided to focus their research on the Jovian moons. The gas giant might have been too bright for the sensitive instruments. In particular, Io, the inner most of the Galilean moons attracted the team's attention as it is a very active moon. Its large volcanoes shoot out tons of material into Jupiter's magnetosphere and beyond. Volcanic activity on Io has considerable other causes than volcanism on Earth. At our home planet it is radioactive decay that internally heats up the Earth. At Io no such decay is present. Moreover, it is the tidal forces exhibited on the moon by Jupiter and the orbital resonances of Europa and Ganyemde. The moon is kneaded all over its orbit. In analogy, think of a tennis ball which you are hitting with your racket. Every time you hit it, the ball is deformed and energy is induced. The ball will get warmer. This is exactly what is happening at Io.

New Horizons observed several eruptions. However, the eruption Tavshtar volcano on February 28, 2007 was most impressive (see Fig. 6.19). The plume rose up to an altitude of about 330 km.

The swing-by maneuver at Jupiter increased the probe's velocity by 3890 m/s and redirected its trajectory towards Pluto by increasing the spacecraft's orbital inclination to 2.5° towards the ecliptic plane.

Goodbye Jupiter, Hello Pluto
The spacecraft continued its journey towards Pluto and spent most of its time in hibernation mode in order to save energy. On June 8, 2008 it crossed the orbit of Saturn. Yet, the ring planet was too far away to allow any reasonable observations.

Next was Uranus whose orbit was crossed at March 18, 2011. Also Uranus was about 3.8 billion kilometers away on its orbit. That was too far away. So no observations were carried out. Between August 25 and 26, 2014, New Horizons arrived at Neptune's orbit. This was exactly 25 years after Voyager 2 had passed the outermost gas giant. Although being too far away, some images of Neptune were taken on July 10, 2014.

As of July 2012, New Horizons had started gathering scientific days with SWAP, PEPSSI and SDC. At the beginning of July 2013, LORRI was able to take further images of Pluto and Charon showing them for the first time as separate bodies. At the end of the same month LORRI was able to take another 12 images of Charon

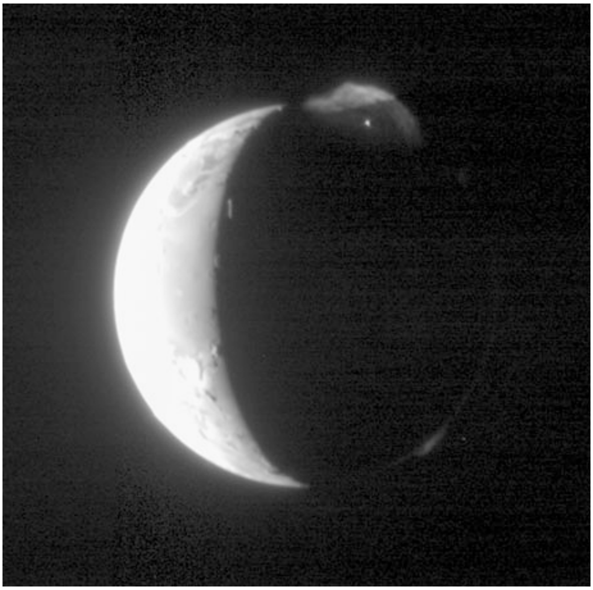


Fig. 6.19 Tvashtar volcano on Io taken by the LORRI camera of the New Horizons probe on February 28, 2007. A 290-km high plume from the volcano Tvashtar is visible (credit: NASA/Johns Hopkins University Applied Physics Laboratory/Southwest Research Institute)

revolving around Pluto. This image sequence covers almost a full rotation of the two bodies.

Finally, a critical moment was reached on December 6, 2014. It was time to fully wake up New Horizons from its hibernation mode. A wake up signal was sent. But would the spacecraft reply? Would it wake up fully functional? It did. A confirmation signal was received from it after a few hours. The next stage on its mission was ready to begin.

Approaching Pluto

The approach for the flyby began on January 4, 2015. On February 12, 2015, NASA released new images of Pluto taken during the last week of January. The images taken from a distance of 203 million kilometers already revealed more details than the previous ones. A image taken on April 15, 2015 shows Pluto in a higher resolution. The discs of Pluto and Charon are clearly recognizable and Pluto's disc may also reveal a polar cap (see Fig. 6.20). Since May 15 the resolution of images taken of Pluto is better than the previously best images taken by the Hubble Space Telescope. However, this situation will only last until 2 weeks after the flyby. After this the resolution will be worse again.

On July 15, 2015 New Horizons will reach Pluto and pass the planet at a distance of approximately 9600 km and Charon at 27,000 km. During the flyby, the spacecraft is supposed to capture global male in high resolution (about 25 m per pixel) of Pluto's surface.

The flyby will not last very long. After it the journey continues further into the Kuiper belt to visit other, not yet determined Kuiper belt objects.

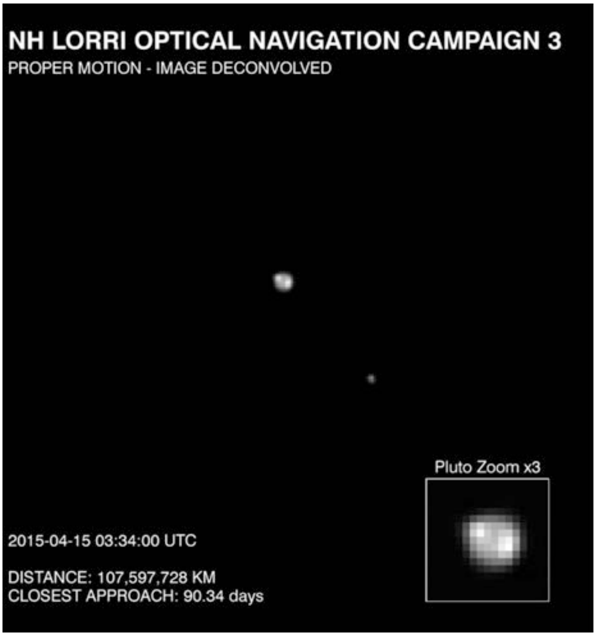


Fig. 6.20 This image of Pluto and it largest moon, Charon, was taken by NASA's New Horizons spacecraft on April 15, 2015. There might be a polar cap visible on Pluto (credit: NASA/JHU-APL/SwRI)

The Kuiper Belt Mission

The KBOs New Horizons is supposed to visit have not been selected yet. This will only be done on August 2015. However, three potential targets, provisionally named PT1 (2014 MU_{69}), PT2 (2014 OS_{393}) and PT3 (2014 PN_{70}) have been preselected on October 15, 2014. Each of them belongs to the dynamically cold population of the classical Kuiper belt objects, the cubewanos. Their diameters are supposed to lie in the range of 30–55 km—too small to be reasonably observable from Earth. Currently preferred is PT1 closely followed by PT3 which is probably slightly larger than the former one.

6.3.4 NEAR Shoemaker

The NEAR Earth Asteroid Rendezvous mission was on of the first missions to particularly visit an asteroid. Its destination was the Near Earth Asteroid (433) Eros. The mission was launched in 1996 and renamed to "NEAR Shoemaker" in 2000 in order to honor the astronomer Eugene Shoemaker after his death.

Spacecraft Design and Technology

The NEAR spacecraft has the shape of an octagonal prism which is approximately 1.7 m wide on each side. Four solar panels are fixed to the main body in a windmill arrangement. It also comprises a fixed 1.5 m high gain radio antenna and a magnetometer at the feed of the antenna. Furthermore, it comprises a X-ray solar monitor on one end and other instruments on the other end.

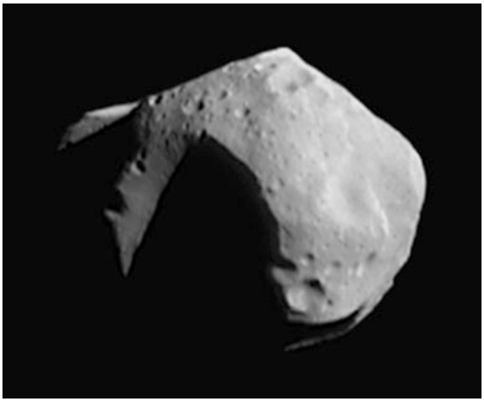


Fig. 6.21 This is a mosaic of the main belt asteroid 253 Mathilde made up of four images returned by the NEAR spacecraft during its flyby. The images were taken on 27 June 1997 from a distance of 2400 km (credit: NASA)

Mission Overview

NEAR was launched on February 17, 1996 on a Delta II rocket. During the first few weeks of cruise, a series of component functional tests checked the health of the spacecraft.

In March 1996 it managed to take some images of the recently discovered comet Hyakutake. However the distance to the comet was quite large.

On June 27, 1997, NEAR's trajectory brought it close to the $50 \times 53 \times 57$ km C-type main belt asteroid (253) Mathilde. The flyby took place at a distance from Earth of 2.2 AU (330 million kilometers). Closest approach distance was 1200 km and occurred at 12:56 UTC. NEAR was able to take many images and covered in all about 60 % of Mathilde's surface (see Fig. 6.21).

On January 23, 1998 a swing-by maneuver took place at Earth. NEAR's closest approach to Earth was 540 km. The maneuver altered its orbital inclination from 0.5 to 10.2° towards the ecliptic plane and reduced its aphelion from 2.17 to 1.77 AU which almost corresponds to Eros' aphelion being at 1.783 AU from the Sun.

The spacecraft was able to detect Eros for the first time on 5 November 1998, approximately 200 days prior to its planned closest approach. Following this early observation, clusters of images were obtained weekly for optical navigation and for initial shape and rotation determination.

Beginning on Jan. 9, 1999, a series of four rendezvous maneuvers with the main thruster were scheduled—spaced 7 days apart—to slow NEAR down in order to achieve a relative velocity between the spacecraft and Eros of 5 m/s.

However, the first of the scheduled rendezvous burns on December 20, 1998 aborted due to a software problem. The contact to the spacecraft put itself in safe mode and was lost immediately after this and could not be re-established for over 24 h.

The original mission plan foresaw for the burns to be followed by an orbit insertion burn on 10 January 1999, but the abort of the first burn and loss of communication made this impossible.

Consequently, after the communication had been re-established, the mission hastily worked out a new plan in which NEAR flew by Eros on December

Fig. 6.22 Eros as images by the NEAR spacecraft on February 29, 2000, from an orbital altitude of about 200 km (credit: NASA/JPL/JHUAPL)

23, 1998 at a speed of 0.965 km/s and a distance of 3827 km to Eros. During this flyby several images of Eros were taken and data was collected by the near IR spectrograph.

Orbit insertion occurred then on February 14, 2000 about 1 year later as originally planned. But fortunately it was possible. The original orbital altitude was at 350 km from the surface but was continuously reduced to about 50 km (see Fig 6.22).

Starting from January 2001, a series of very closes passes were made bringing the probe as close as 5–6 km above the surface. Detailed images were taken during these passes. On January 25, 2001, NEAR approached the surface by about 2–3 km.

After these passes and being about 1 year in orbit of the Eros, the mission team decided to begin a controlled descent. On February 12, 2001, NEAR landed on the surface of Eros. To the big surprise of the mission team, the landing was extremely soft and the probe remained fully functional. The mission team hence decided to continue and extend the mission. For that purpose, the reprogrammed parts of the software onboard the spacecraft so that it was possible to carry out further analysis on the surface of Eros. It was particularly interesting to collect data of the surface composition as there was only a few centimeters of distance between NEAR and the soil. About 2 weeks later, the contact to NEAR was lost signaling the end of the mission.

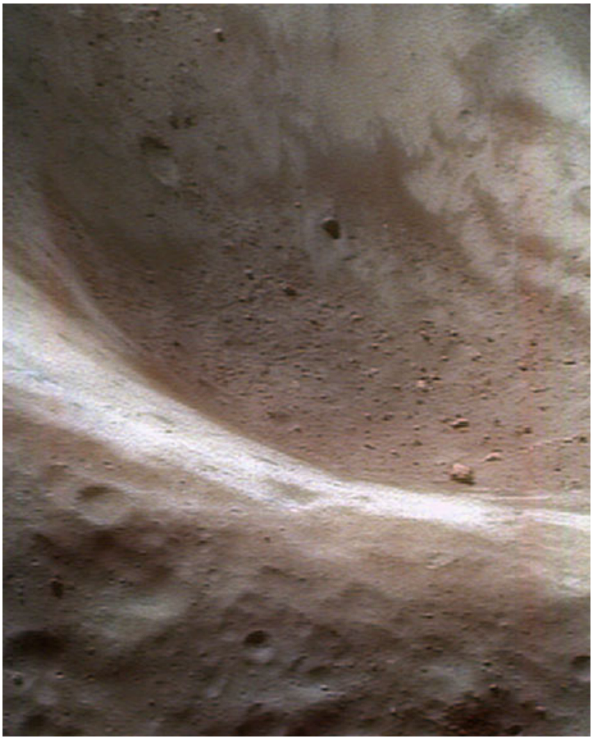


Fig. 6.23 False-color view of a large, 5.3-km diameter crater that is close to Eros' slimmest part (credit: NASA/JPL/JHUAPL)

Results

The NEAR Shoemaker spacecraft delivered a huge amount of data and holds the record for some mission firsts. It was the first asteroid orbiter and the first spacecraft to land on an asteroid. The probe provided detailed maps of Mathilde's and Eros' surfaces. In particular, data returned from Eros was unique in its details owing of course the 1 year orbiting time of the NEAR.

Several craters, rifts and troughs were detected on Eros. They are considered to be remnants of a heavy bombardment. In addition, all over Eros' surface debris is scattered, the rocks vary in size. The way most of the debris is scattered resembles patterns of a heavy impact that occurred about 1 billion years ago. This particular impact may also be accountable for an area with small craters with diameters less than 500 m that covers about 40 % of Eros' surface.

NEAR also found that the majority of the asteroid's surface is covered by regolith which is a layer of loose, heterogeneous superficial material covering solid rock. It includes dust, soil, broken rock, and other related materials.

Figure 6.23 shows a false-color view of a large, 5.3-km diameter crater that is close to Eros' slimmest part which is about half the asteroid's diameter there. NEAR took a series of color pictures intended to measure the properties of regolith inside the asteroid's craters. The image was taken from an altitude of 50 km. The redder hues represent rock and regolith that have been altered chemically by exposure to the solar wind and small impacts. Bluer hues represent fresher, less-altered rock and regolith, such as the bright patches that have been less affected by space weathering. Interestingly, most of the large boulders have been just as affected as the regolith. This suggests either that the rocks are relatively old, or that they are "dirty" from an adhering film of regolith particles. In general, Eros is an S-type asteroid.

Eros' mass could also be determined to be about 7.2×10^{15} kg by taking into account the gravitational influences exhibited by the asteroid on the NEAR spacecraft. Furthermore, the dimensions of peanut-shaped Eros were measured as $33 \times 13 \times 13$ km.

In addition, during the course of the mission, Eros' orbital parameters were refined. Eros was found to be a Amor type asteroid having an orbital period of 1.76 years and an inclination of 10.83° towards the ecliptic plane. It comes as close as 1.133 AU to the Sun (perihelion). It orbits take Eros out to about 1.783 AU at its aphelion.

6.3.5 Deep Space 1

Deep Space 1 was a spacecraft launched on October 24, 1998. The main purpose of this mission was the testing of new technologies, such as a xenon ion engine. The initial target was the asteroid (9969) Braille, a small Mars-crossing asteroid ($2.1 \times 1 \times 1$ km) discovered in 1992 by Eleanor F. Helin and Kenneth J. Lawrence at the Palomar Observatory. Its aphelion is at about 3.35 AU, its perihelion at 1.33 AU and its semi-major axis at 2.34 AU.

Flyby of (9969) Braille

However, during its flight the spacecraft encountered several problems. The probably severest and most frustrating one occurred during the flyby of Braille which rendered it only a partial success. Deep Space 1 was supposed to pass the asteroid at a distance of only 240 m. Yet, it various technical problems including a software crash and issues with the tracking system, caused a different trajectory which only allowed a passage at about 28 km distance from the asteroid. Unfortunately, this was not the only problem. During the closest approach, the spacecraft's automatic camera pointed away from the asteroid. Only about 15 min after the flyby images could be taken. Yet, the spacecraft had already increased its distance to the asteroid to about 13,500 km. The larger distance in combination with Braille's low albedo rendered the fuzzy images not very meaningful.

A New Target, New Hope

However, hope came up again, when the mission team discovered that after the flyby of Braille some fuel was left. The team decided to extend the Deep Space 1's mission. A new target was searched and found in 19P/Borrelly, a short-period Jupiter family comet discovered by Alphonse Borrelly during a routine search for comets at Marseilles, France on December 28, 1904. It is a moderately sized comet ($8 \times 4 \times 4$ km) which has an orbital period of 6.8 years and a semi-major axis at 3.59 AU, its aphelion at 5.83 AU and its perihelion at 1.35 AU.

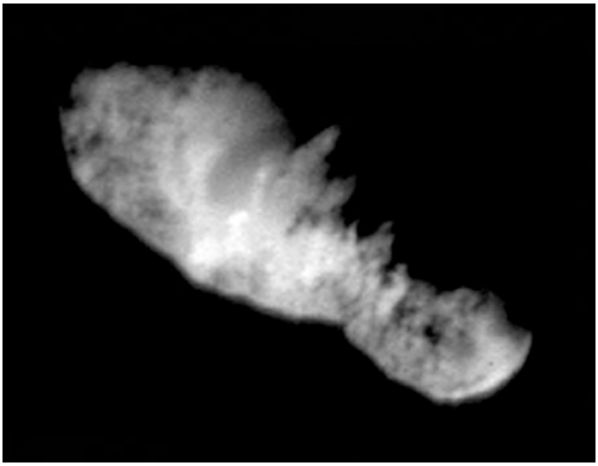


Fig. 6.24 High resolution view of the icy, rocky nucleus of comet Borrelly (credit: NASA/JPL)

The decision was made and the spacecraft set course towards the comet. On September 21, 2001 Deep Space 1 performed a flyby of Borrelly. This flyby was the complete opposite of the previous one of Braille. It was a great success. Detailed images of the cometary nucleus could be taken. These images had a much higher resolution as any previous close photos of a comet's nucleus. The by then record holder had been the spacecraft Giotto with its close-ups of Halley's comet in 1986 (see Fig. 6.24).

And It Runs and Runs

Voices had been raised that a flyby of comet Borrelly could severely damage the spacecraft. It was argued that dust particles and debris of the comet could collide with the Deep Space 1 which had not been equipped with a debris shield. Yet, the voices silenced and the probe survived the flyby without any notable collisions. Still some fuel was left, so the mission was extended again. This time, however, it was decided not to send it to another celestial object but to use the remaining time for further testings of the hard- and software. The testings were extensive and insightful. However, Deep Space 1's engines were shutdown on December 18, 2001. This meant the official end of the mission. However, the probe's communication systems are still active today so that theoretically, the spacecraft could be reactivated when needed.

New Technologies on the Testbed

Many new technologies were tested with Deep Space 1. We will just go through some of them as they are relevant for further missions. Probably one of the most important subsystems is the so-called "Autonav" which became with modifications an essential navigation technology in many other spacecraft.

Asteroids in the inner solar system move in relation with other bodies at a notable and predictable speed. This aspect can be used for determining a spacecraft's relative position by tracking such asteroids across the star background. If at least two asteroids are tracked, the spacecraft's position can be triangulated. With the help of at least two such positions it is further possible to determine the probe's trajectory. This allows for auto navigation which has several benefits compared to previous mechanisms that involved usage of the Deep Space Network (DSN) and a permanent open communication channel to Earth via which many skilled navigators influenced and determined the spacecraft's flight. Furthermore, capacities at the DSN could be reduced and thereby considerable cost saving achieved.

A new type of solar panel was used with Deep Space 1, the Solar Concentrator Array with Refractive Linear Element Technology (SCARLET) which generates about 2.5 kW at a distance of 1 AU and weigh much less and are smaller than by then conventional solar arrays.

The last subsystem we will consider is the "Remote Agent", a remote intelligent self-repair software. It was the first time that an artificial intelligence control system was used in a spacecraft. This system did not need any human intervention but was capable of analyzing the spacecraft's status and various other conditions and react thereupon autonomously.

6.3.6 Hayabusa

Hayabusa, meaning falcon in Japanese, was a spacecraft developed by the Japan Aerospace Exploration Agency (JAXA). Its goal was it to return a sample of material from the small Near Earth Asteroid (NEA) (25143) Itokawa.

Hyabusa was the first return sample mission for which a spacecraft deliberately landed on the surface of an asteroid. We have already seen that the NEAR Shoemaker spacecraft had also landed on the asteroid Eros before. However, NEAR Shoemaker was originally not designed as a lander. It was more or less simply dropped on the surface and by some chance continued to operate fully functional for about a couple of weeks. Hayabusa on the contrary was designed as real landing mission. In general, Hayabusa was supposed to have a short touchdown at the surface, collect some material and then leave again. However, the mission developed differently. The spacecraft landed on Itokawa and remained there for about 30 min before it left again. Thus, it can be regarded as a real lander in all respects.

Prior to Launch: The Mission Background

First ideas for a mission to an asteroid arose in 1986 and 1987. Scientists investigated the feasibility of a sample return mission. This was something completely new. No unmanned spacecraft before had tried to land on a body of our solar system and return from it. Asteroids and comets, the only known small solar system bodies by that time, were considered to be difficult and challenging targets. These objects are small and only have limited gravity to hold a spacecraft on the ground. In addition, their surfaces are unknown. Hence, it is very difficult to plan a landing site beforehand. Basically, you can only do that while being on location.

In the years between 1987 and 1994 the Japanese Institute of Space and Astronautical Science (ISAS) and the American NASA worked together. The results of the cooperation were manifested in the NEAR spacecraft as an asteroid rendezvous mission and a sample return mission named Stardust.

After that, ISAS proceeded with their own planning and selected in 1995 the asteroid (4660) Nereus as a first target. Nereus is a tiny asteroid with a diameter of about 330 m. It was discovered by Eleanor F. Helin on February 28, 1982, approximately 1 month after a near pass by the Earth. Nereus is potentially a very important asteroid. It is an Apollo and Mars-crosser, with an orbit that frequently comes very close to Earth, and because of this it is exceptionally accessible to spacecraft.

The planning went on smoothly and a launch date was set to July 2002. However, thereafter, problems with Japan's M-5 rocket occurred and caused a delay in the preparations. Unfortunately, while waiting for the problems to be solved, the launch window for Nereus closed. Hence a new target object had to be found. The asteroid of choice became 1998 SF_{36} which was soon thereafter named after the Japanese rocket pioneer Hideo Itokawa (1912–1999).

The problems could be solved and the M-5 rocket successfully launched on May 9, 2003 and brought the spacecraft Hyabusa into space.

The Journey to Itokawa

Hayabusa's journey to the asteroid and later on its stay there were marked with technical problems. This is not uncommon to such new and ambitious challenges as the first asteroid sample return mission.

At the end of 2003, the spacecraft experienced significantly strong solar winds triggered by a large solar flare. These damaged the solar panels that supplied the probe with energy. The reduction of available electrical power also reduced the efficiency of Hayabusa's ion engines. This in turn delayed the arrival at Itokawa by about 3 months from June to September 2005. The mission team was disappointed. Considering orbital mechanics, it was clear that the spacecraft still had to leave Itokawa at the forecasted date. This, however, meant that the stay was shortened drastically and thus also the number of landings was limited to two or in best case scenarios three landings.

On May 19, 2004, Hayabusa made a swing-by maneuver at Earth coming as close as 3700 km to our home planet. The X-axis reaction wheel failed on July 31, 2005. These three wheels (X-, Y-, Z-axis) are important for stabilizing and orienting the spacecraft.

A first image of Itokawa was released on August 14, 2005. It was taken by the Startracker camera and only showed a moving light spot which was believed to be the asteroid. Further images were taken between August 22 and 24, 2005.

On August 28, 2005, the spacecraft switched from its ion engines that had propelled the vehicle so far to bi-propellant thrusters that had been specifically designed for orbital maneuvering.

Approaching and Orbiting Itokawa

The target got closer and closer and the images taken revealed an ever increasing number of details. From September 4, 2005 on Hayabusa was ablate to confirm the asteroids elongated shape. Then, about 1 week later, the spacecraft spotted individual hills on the surface. Finally, on September 15, 2005, Hayabusa reached a position about 20 km away from Itokawa, the so-called gate position that was considered to be the entry point into an orbit around the asteroid. The Bi-propellant thrusters then ignited for a last time in order to reduce the Hayabusa's relative velocity to that of the asteroid. After successful completion, Hayabusa finally had entered the orbit around Itokawa.

On October 4, 2005, JAXA announced that the spacecraft had reached its home posit at a distance of about 7 km from the asteroid. It was also announced that the second reaction wheel (Y-axis) had failed. The probe had thus to be pointed by its thrusters. From now on, the spacecraft took very detailed images of the asteroids surface and the scientists tried to find landing sites with these high resolution images.

On November 4, a descent was tried but had to be aborted due to the detection of an anomalous signal. Later on, it was suggested that the optical navigation system of Hayabusa did not work properly due to the very complex shape of Itokawa.

Another descent was done on November 12. The probe came as close as 50 m to the surface and then ascended again. Hayabusa's mini lander MINERVA was released. Due to an error, it could not land but was ejected into space and was thus lost.

A week later on November 19, Hayabusa finally touched down on the surface of the asteroid. Communication was lost during the descent, the probe put itself into a safe mode, and it was not clear whether really samples were collected. However, after the communication had been reestablished again, Hayabusa was ordered to ascend again and return to its orbit.

Therefore, another landing was performed on November 25. This time everything worked fine and samples were taken. In the weeks after, communication was lost and it took several weeks to reestablish it and bring the spacecraft under control again. Due to this the spacecraft could not leave the asteroid as planned on December 14, 2005.

A new launch window had to be determined and was found to be in spring 2007. Thus, the spacecraft stayed much longer in orbit than planned but finally started its return flight on April 25, 2007. In May 2008, the return capsule that was supposed to bring the asteroid samples back to Earth was checked for the last time.

Several further technical issues occurred but the spacecraft was able to return to Earth. On June 13, 2010 at 10:51 UTC the return capsule was released and landed as planned in the Australian Woomera Prohibited Area. The capsule was localized about an hour later and recovered the next day. It was sealed and the long awaited samples were ready for examination.

Results

Hayabusa had indeed collected samples of asteroid surface material. This material was then studied. According to Japanese scientists, the composition of Hayabusa's samples was more similar to meteorites than to known rocks from Earth. Their size is mostly less than 10 μm. In addition, the material matches chemical maps of Itokawa from Hayabusa's remote sensing instruments. The researchers found concentrations of olivine and pyroxene in the Hayabusa samples.

In addition to these samples, the images taken by Hayabusa were particularly precious. The resolution was very high revealing details of less than 1 m in size. It is peculiar that almost no craters are visible on Itokawa's surface. Craters have been the dominating surface features of many other asteroids including those that have been visited by spacecraft, Eros and Braille.

Some parts of Itokawa's surface are covered by regolith and rocks of different sizes. Other parts show nothing like this and look like clean-swept stone. The asteroid's mean density was found ti be 2.3 ± 0.3 g/cm^3. This is less than would be expected from solid rock. Scientists thus believe that Itokawa is not one solid body but, as many other asteroids also, a simple rubble pile loosely being held together by gravitational forces.

6.3.7 Stardust

Stardust, as its name suggests, was a mission primarily designed to collect stardust, i.e., interstellar dust or dust particles from the coma of a comet. This dust is very interesting as it is believed that the dust is an indicator for the conditions in the early solar system and the composition of the interstellar medium. Hence, taking into account the limitations of space probes and their analysis capabilities, it was the primary objectives of the Stardust mission to return cometary dust as well as interstellar, sometimes also called cosmic dust.

So, what is interstellar dust? Organic or more complex compounds can be formed in stars and subsequently ejected into space, e.g., in the form of stellar wind or stellar flares, like it is the case with our Sun. This material may then when it entered the interstellar space, i.e., outside of the influence sphere of the star, condensate to dust clouds which travel through the space.

In 1993 the spacecraft Ulysses had detected that interstellar dust is streaming through our solar system. The Galileo spacecraft confirmed this finding 1 year later in 1994. However, it is hardly possible to obtain any information about the dust's composition from such a great distance. The scientists need the actual particles to make further analysis.

For this purpose, the spacecraft Stardust was developed. It was the first US mission dedicated solely to the exploration of a comet and the first unmanned mission to return extraterrestrial material to Earth.

Which Target?

It was thoroughly debated which comet should be selected. Finally, comet 81P/Wild, also named Wild 2, was chosen. This comet is named after the Swiss astronomer Paul Wild (1925–2014) who discovered it on January 6, 1978. For most of its existence (about 4.5 billion years), Wild 2 had a much less eccentric, near circular orbit around the Sun. This suddenly changed in 1974 with a very close encounter with the gas giant Jupiter which it passed at only 0.006 AU distance. Jupiter's strong gravitation perturbed Wild's orbit bringing it into the inner solar system. Its perihelion was reduced from about 5 to 1.59 AU and the aphelion of its orbit is currently located at 5.3 AU. This results in a significantly new shorter orbital period of about 6 years compared to the 43 years it had before the encounter. Further close encounters with Jupiter are not excluded.

The scientists expected that most of the original material which the comet formed in the early phase of the solar system, would still be present. Its observation would thus give insight into this early stage.

Mission Objectives

Returning samples of cometary dust and also interstellar dust in itself is a challenging task. Scientists hoped to gain more insight into the evolution of the early solar system. The main objectives of the mission therefore were the following. First, a flyby of the comet had to be achieved. This had to be done with a sufficiently low velocity. Otherwise captured particles might be destroyed. As special collection medium, an aerogel, was designed for this. Secondly, the probe was supposed to intercept a significant amount of interstellar dust particles using the same collection medium.

Thirdly, of course, the collected samples had to be returned safely to Earth.

How to Collect Sensitive Particles?

It was a crucial aspect in the mission design to find a way to collect the tiny dust particles in a non-destructive way. The involved scientists decided to use a low-density aerogel for this purpose. The used material is extremely porous and has a sponge-like character. It is aggregated in about 90 blocks which have together approximately the size of a tennis racket (see Fig. 6.25).

How does this material work? When the spacecraft encounters dust particles these will hit the surface of the aerogel softly and will be slowly decelerated while intruding into the gel without heating up. The latter aspect is of significant importance. Thermal heating would destroy the chemical composition of the dust particles. Chemical compounds could break up or new ones could be formed. It would no longer be possible to determine the original composition. Just imagine a spacecraft or a meteorite entering Earth's atmosphere. It will heat up, components such as metals can melt and so on. This has to be avoided under all circumstances.

Launch, Storm and Annefrank

Stardust was launched on February 7, 1999 abroad a Delta II rocket and was subsequently released into a heliocentric orbit with a period of 2 years. On November 9, 2000 got into a heavy solar storm, the strongest one since the beginning of solar observations. The solar wind was about 100,000 times stronger than as usual. The spacecraft was put into safe-mode as a precaution. Right after passing the storm, the systems of the spacecraft were rebooted. Fortunately, everything appeared to fine and Stardust returned to normal operation. Surviving such a

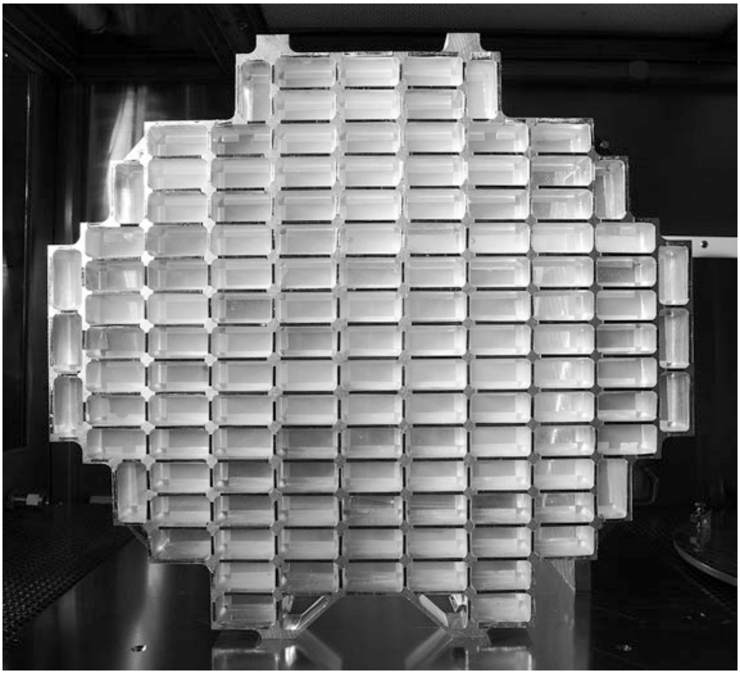


Fig. 6.25 Stardust Dust Collector with aerogel (credit: NASA/JPL)

heavy storm without damages is not always the guaranteed, e.g. The solar panels of the Japanese probe Hayabusa was seriously damaged in a solar storm.

In 2001, a swing-by maneuver was made at Earth. Stardust gained additional speed during this maneuver and its orbit was extended to a period of 2.5 years.

In 2002, the spacecraft passed asteroid (5535) Annefrank at a distance of only 3079 km. Annefrank is a typical main belt asteroid having its perihelion at 2.07 AU and its aphelion at 2.35 AU on its near circular orbit. It was discovered by the German astronomer Karl Wilhelm Reinmuth (1892–1979) on March 23, 1942. In 1995, it was named after Anne Frank, the Dutch-Jewish diarist who died in a concentration camp.

Stardust was able to determine the asteroids size as $6.6 \times 5.0 \times 3.4$ km. This is about twice as big as previously thought. The asteroid is shaped like a triangular prism, with several visible impact craters. A preliminary analysis of the Stardust imagery suggests that Annefrank may be a contact binary, although other possible explanations exist for its observed shape.

Encounter with Wild 2

After its flyby of the asteroid Annefrank, Stardust continued its journey towards comet Wild 2 which it reached on January 2, 2004. The closest approach was at 237 km, more distant than originally planned. The mission team increased the

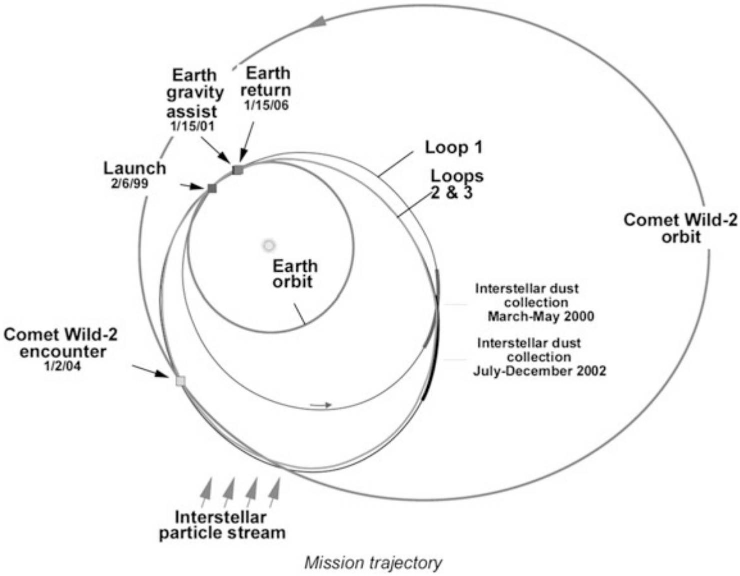


Fig. 6.26 Mission trajectory of the Stardust spacecraft enroute to comet Wild 2 (credit: JPL/NASA)

distance of 150 km because of security considerations. They were afraid that the spacecraft could be damaged by cometary debris or outbursts of cometary material.

During the flyby the sample collector plate, the tennis-racket sized container of the aerogel, was deployed. Dust grains of the cometary coma were collected by the aerogel. In addition, high-resolution images taken by Stardust provided a high-resolution view on Wild's nucleus. The flyby did not last very long. After it, the probe followed its heliocentric orbit again.

At several points of Stardust's orbit interstellar dust particles were collected (see Fig. 6.26).

Bringing the Samples Home

On January 15, 2006, the point was arrived at about 111,000 km from Earth where the sample return capsule was separated from Stardust and started its trip back home to Earth.

Shortly after the separation, Stardust ignited its engine to increase altitude and return to its heliocentric orbit. It was still fully functional and was thus put into hibernation mode to save energy resources. The mission team chose this approach, as like this they still would be able to reactivate the spacecraft when needed.

The return capsule followed its predetermined path and landed at 11:12 UTC on the ground of a military base in Utah. It was soon after discovered by helicopters and recovered. The capsule was immediately sealed and transported to the foreseen laboratory for first analysis.

What to Do Next?: An Extended Mission

Yet, what should be done with the “sleeping” Stardust spacecraft? Several ideas came up. One of them proposed sending the probe to another comet—9P/Tempel, also called Temple 1, a periodic Jupiter-family comet discovered by the German astronomer Wilhelm Tempel (1821–1889) in 1867. It is a 7.6 km × 4.9 km sized comet orbiting the Sun in about 5.5 years.

Temple 1 was not an unknown. It had already been visited by NASA’s space probe Deep Impact in 2005. During Deep Impact’s flyby a heavy impactor had been detached from the spacecraft and sent to the comet nucleus. There the impactor crashed on the surface causing an about 20 m deep crater. It had been planned that Deep Impact photographed the impact and measured relevant data. However, due to a unpredicted huge dust cloud, this failed to some respect.

This is where Stardust came into play. Why not send the spacecraft to Temple 1 and let it do the job, Deep Impact should have done? The idea convinced and NASA approved the extended mission on July 3, 2007. The new mission was then labelled “New Exploration of Temple 1” or just Stardust-NExT.

Towards Temple 1...

Stardust received a wake up call and new trajectory data. On January 14, 2009 a swing-by maneuver was made at Earth at a distance of 9157 km. About 2 years later, on February 15, 2011, Stardust-NExT passed Temple 1 at a distance of only 181 km. About 72 images were taken during the brief encounter and documented changes on the comet’s surface since Deep Impact. The images also revealed parts of the surface that had never been seen before. The scientists were particularly happy that also Deep Impact’s crater was still visible and could be photographed by Stardust (see Fig. 6.27).

Only a month later, on March 24, 2011, Stardust-NExT’s mission was officially ended and its systems shut down.

Results

The encounters with the two comets provided a vast amount of data. Already at first visual inspection of the secured aerogel revealed that the collecting had been successful. About 150 particles were found there. Furthermore, scientists were able to distinguish those impacts in the aerogel that had been caused by interstellar dust particles from those created by cometary dust particles. The latter ones were larger.

First analysis of showed that the composition of the cometary particles was very similar to that of meteorites. Also a range of organic compounds was found such as glycine, amorphous silicates, in addition to crystalline silicates such as olivine and pyroxene. This confirmed a consistency with the mixing of solar system and interstellar matter as had been indicated before by ground-based observations.

In April 2011, evidence was found that liquid water had existed in the comet’s nucleus. This was a very surprising finding as before it had always been assumed that comets would never get warm enough to melt icy materials.

The samples further indicated that the outer region of the early solar system had not been isolated. On the contrary, high temperature inner solar system material formed and was subsequently transferred to the Kuiper belt.

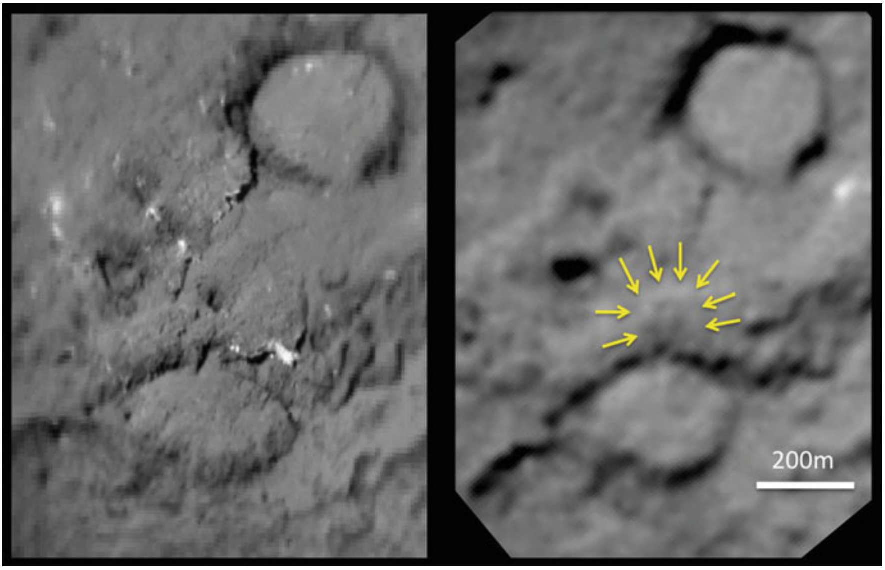


Fig. 6.27 This pair of images shows the before-and-after comparison of the part of comet Tempel 1 that was hit by the impactor from NASA's Deep Impact spacecraft (credit: NASA/JPL-Caltech/University of Maryland/Cornell)

6.3.8 Deep Impact and EPOXI

Deep Impact was, as its name already suggest, a mission, launched in 2005, during which an impact was supposed to be created on a comet. Therefore, the spacecraft was basically designed to be divided in two parts. The first one, the actual spacecraft, being responsible for the flyby of the comet and photographing with several cameras and measuring with an IR spectrometer the effects of the second part, the impactor that would at a given point of time be detached from the first part. The impactor was supposed to hit on the surface of the comet. The impactor part comprised two elements: the actual impactor which was a block of pure copper with a mass of 113 kg and additional instrumentation. The total mass of the second part was 372 kg.

You may ask why copper was used for the impactor element? Copper is not very common with comets. It has hardly been detected on any comet and anyhow if then only in very small traces. Just imagine what would happen if a body consisting of an element common on comets would hit the cometary nucleus. The chemical composition of the surface region would alter. The abundances of chemical compounds before the impact and after it would most likely differ. Reliable measurements could not be obtained. By using copper, the scientists assumed, the results would not be falsified.

Mission Objectives

The idea of creating an impact on a comet may remind some of you of some blockbuster movies. The common scenario of these movies often is that a comet will collide with Earth and thereby endanger life on our home planet. Finally, mankind is saved by destroying the comet, e.g. by attacking it with missiles or other impactor. So, was this scenario a blueprint of the Deep Impact mission? Was it a test run for saving our planet from a possible comet impact?

For sure not! The mission was of mere scientific nature. Although the previously described catastrophe scenario was often referred to in the media coverage on the Deep Impact mission.

What were the scientific goals then? The findings were supposed to answer some fundamental questions about comets. What is the composition of a comet nucleus? What can we learn about a comet's surface properties? Density? Porosity? Stability? As we have seen in Chap. 3, several theoretical models exist about the internal structure of comet. Which one is correct? Can we rule out at least one? How is the surface of the nucleus composed? Are there different layers of material? What do they look like? These are only a few questions, the Deep Impact mission was supposed to answer or at least give some indications.

Mission Background and Prior to Launch

Deep Impact was not the first mission of its kind proposed. In 1996, a similar comet impact mission was proposed to NASA. However, the selection committee at NASA was skeptical about the outcome of such a mission and therefore, the proposal was not approved.

In 1999, a refined proposal of that 1996 mission bearing the name "Deep Impact" was handed in and finally approved by NASA. Comet 9P/Tempel also called Temple 1, a short period Jupiter-family comet discovered by Wilhelm Tempel in 1867, was selected as a primary target. It is a 7.6 km $\times$ 4.9 km sized comet orbiting the Sun in about 5.5 years.

The mission preparations started soon after and swiftly progress was made. The launch was scheduled for late 2004. However, technical problems delayed the mission, so that the actual launch only took place on January 12, 2005. The original mission plan had foreseen several swing-by maneuver. The delay made them impossible. Hence, the spacecraft was sent on a direct flight to comet Temple.

The Impact

The approach phase extended from 60 days before encounter (May 5, 2005) until 5 days before encounter. On June 14 and June 22, 2005, Deep Impact observed two outbursts of activity from the comet, the latter being six times larger than the former. Impact phase began nominally on June 29, 2005, 5 days before impact. The impactor successfully separated from the flyby spacecraft on July 3 UTC. The first images from the instrumented Impactor were seen 2 h after separation.

The impact actually occurred on July 4, 2005. The brightness of the comet increased by about six times for a brief period of time directly after the impact. The impactor, also being equipped with cameras, was able to return images to Deep Impact as late as 3 s before the impact. Its impact speed was about 10.3 km/s and the impact energy was roughly equivalent to approximately 4.7 tons of dynamite.

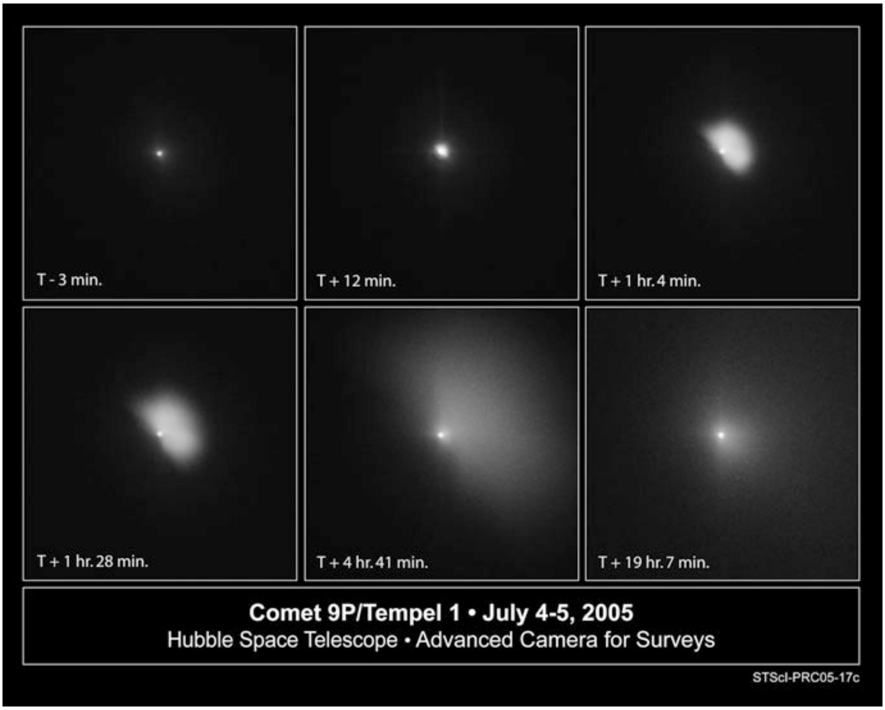


Fig. 6.28 This series of Hubble Space Telescope images captures the ejection of a bright plume of dust following the July 4 collision between impactor released by the Deep Impact spacecraft and comet 9P/Tempel 1 (credit: NASA, ESA, P. Feldman (Johns Hopkins University), and H. Weaver (Johns Hopkins University Applied Physics Lab))

About 5 million kilograms of water and 24 million kilograms go dust were released from the Tempel's nucleus. A giant gas and dust cloud was the result which unfortunately hindered Deep Impact from carrying out all the planned measurements. The spacecraft was not able to analyze the newly created about 20 m deep crater in detail. The successor probe, Stardust, helped out there a few years later.

During the impact phase, Deep Impact was not alone. A whole armada of spacecraft were observing the event. Among them were the Hubble Space Telescope (see Fig. 6.28), the X-ray telescopes Chandra and XMM-Newton, the Spitzer Space Telescope and the probe Rosetta currently being on its own way to an encounter with a comet.

Results

A big surprise was that by far more dust particles were excavated from the cometary nucleus than water. The dust particles were also finer than expected. This finding seemed to rule out only one model of cometary structure. The comet could not be very porous where the comets were merely loose aggregates of material.

In addition, several compounds could be detected by spectral analysis of the impact such as clays, carbonates, sodium, and crystalline silicates. These findings, however, caused some headaches with the involved scientists. Clays and carbonates usually require liquid water to form. This was contrary to the common assumption that comets would never get warm enough to melt icy materials.

Where did the liquid water come from? Several hypotheses were postulated. Perhaps an earlier body existed that was large enough and comprised liquid water. The comet could have been formed from that parent body in some way. Another explanation was the existence of radioactive decay during some period of the comet's lifetime. However, the nature of this is still unknown.

A New Mission

It took some time to send all the gathered data back to Earth. It was only on August 3, 2005 that the download was complete. This day is thus considered to be the official end of the primary mission.

Yet, Deep Impact had survived the whole mission and remained fully functional. Should the spacecraft be abandoned? It was decided by officials that the mission should be extended. A call for proposals was announced.

Two of them came on the short list. The first one, the "Deep Impact eXtended Investigation of comets" (DIXI), proposed to send Deep Impact to another comet. The second proposal, the "Extrasolar Planet Observation and Characterization" (EPOCh) mission, suggested to observe extrasolar planet.

In 2007, NASA announced that both projects were selected and combined into one mission: "Extrasolar Planet Observation/eXtended Investigation of comets" (EPOXI).

Comet 103P/Hartley was selected as a target after the previous chosen one, 85P/Boethin, could not be found anymore. Boethin was too small and faint for its orbit to be calculated accurately. It did not return to the inner solar system as expected.

On November 4, 2011, Deep Impact EPOXI passed Hartley in a distance of about 700 km. During flyby, the spacecraft took several images and observed various jets coming from the nucleus.

Additionally, the spacecraft used the High Resolution Instrument onboard, a telescope, to perform photometric observations of previously discovered transiting extrasolar planets from January to August 2008.

At the end of 2011, a new course was set towards the asteroid (163249) 2002 GT with an expected arrival date in 2020.

Deep Impact observed Comet Garradd (C/2009 P1) from February 20 to April 8, 2012. The comet was 1.75–2.11 AU from the Sun and 1.87–1.30 AU from the spacecraft. It was found that the outgassing from the comet varies with a period of 10.4 h, which is presumed to be due to the rotation of its nucleus. The dry ice content of the comet was measured and found to be about 10 % of its water ice content by number of molecules.

Another comet was briefly observed by Deep Impact EPOXI. The spacecraft had a short look at the comet C/2012 S1 (ISON) in February 2013. Soon after, technical problems came up and on August 8, 2013, contact was lost. On September 20, 2013, NASA officially ended the Deep Impact mission.

6.3.9 The Halley Armada

The Halley Armada is the popular name used for a group of five spacecraft that were sent to Halley's comet during its 1986 journey into the inner solar system. It was the first time in science history that spacecraft were able to obtain scientific measurements and data from a comet on location in space. No manmade vehicle had come so close to a comet as these five. The spacecraft involved were ESA's Giotto, the Soviet Vega 1 and 2, and the Japanese Suisei and Sakagake.

Scientists all over the world wanted to use the unique opportunity to study this famous and long-known comet in detail. All that had been known about comets before, had been derived by spectral analysis, visual ground-based observations and other Earth-bound means. Nobody had seen a comet nucleus before. Only theoretical models about the structure and composition of a nucleus existed. However, no one knew for sure as cometary nuclei are usually not directly visible from Earth. They are too small. When the comet travels through the inner solar system the size might not be of such an issue. Yet, the direct view on the nucleus is then concealed by the cometary coma and tails. When the comet is outside the inner solar system, it is usually to faint to observe. The best solution, therefore, to get some more insight is to actually send a spacecraft there for taking photos and doing measurements from a close distance.

The Vega Probes

The Vega probes, Vega 1 and 2, were the result of a cooperation between the Soviet Union and several western scientists. The name is an acronym and the "Ve" stands for Venus and the "GA" for "Gallei", the Russian name for Halley.

Both Vega probes were twins, i.e. they were identical in structure and instruments. Each of them had two large solar panels and the following instruments onboard: an antenna dish, cameras, spectrometer, infrared sounder, magnetometers, and plasma probes.

Furthermore, both were sent on an identical double mission with a first stop at Venus and a second one at Halley's comet.

Vega 1 was launched on December 15, 1984 by the USSR, Vega 2 followed on December 21, 1984. Their mission brought them first to Venus where they arrived on June 11, 1985 and June 15, 1985, respectively. Each of them dropped a lander unit there and then headed off for their encounter with Halley.

Encounter with Halley

The journey to Halley went quite smoothly. Vega 1 started to return images from March 4, 1986. Vega 1 was the first spacecraft ever to take an image of a cometary nucleus. Unfortunately, the photo was not sharp. The first images showed two bright areas on the nucleus. It was initially thought that this indicated the presence of double nucleus. Later on, with the help of additional photos by Vega 1 and the other spacecraft of the Halley armada, it became, however, apparent that the comet's nucleus was made up of a single body and not a loose conglomerate of several bodies. The bright areas turned out to be two strong jets ejecting material from the nucleus into space. The images also showed the nucleus to be dark, and the infrared spectrometer readings measured a nucleus temperature of 300–400 K, much warmer than expected for an ice body. The conclusion was that the comet had a thin layer on its surface covering an icy body. The Vega 1 images suggested a nucleus of about 14 km in size and a rotation period of approximately 53 h which matches quite good with modern results. Further details could not be determined due to the probe's poor image quality.

Vega 1 had its closest approach with a distance of 8889 km to the comet on March 6, 1986. Vega 2 followed on March 9, 1986 with about 8030 km. Both passed the comet with a very high relative velocity of 79.2 km/s and 76.8 km/s, respectively. In spite of the very short flyby, Vega 1 managed to take about 500 images and Vega 2 even 700. Luckily, the images of Vega 2 were of much better quality and hence had a better resolution. The few days in between were sufficient to the change the cometary environment. During Vega 2's flyby much less dust was in the coma and give more or less free sight to Halley's nucleus.

Further images were taken by Vega 2 between March 8 and 10, 1986 which concluded the primary mission. The control center soon after lost contact with Vega 2 on March 24, 1986. Since then, Vega 2 has been orbiting the Sun.

Suisei and Sakigake

Suisei, meaning comet in Japanese, and Sakigake, the Japanese word for pathfinder or pioneer, also were twin probes developed by the Japanese Institute of Space and Astronautical Science (ISAS), the predecessor of JAXA.

The construction of both probes was almost identical. They only differed on their respective payloads. Suisei was equipped with a CCD Ultraviolet imaging system and a solar wind instrument. Unlike its twin Suisei, Sakigake carried no imaging instruments in its instrument payload.

Suisei was launched on August 18, 1985 and was planned to measure the amount of hydrogen in the comet's nearby environment. Furthermore, it was supposed to analyze the interaction between the solar wind and the plasma of the cometary coma. Suisei began its UV imaging in November 1985. ITs closest approach was on March 8, 1986 at a distance of 151,000 km from the comet's core.

The twin probe, Sakigake, was Japan's first interplanetary spacecraft and had been designed to observe space plasma and magnetic fields in the interplanetary space. It was launch on January 7, 1986 and had its closest approach with Halley on March 11, 1986 at a distance of about 7 million kilometers.

The spacecraft was during its flyby clearly in front of the comet and was thus to observe the interactions of coma and solar wind.

Giotto

The spacecraft Giotto was a mission of the European Space Agency (ESA). It came closest to Halley compared to the other armada spacecraft. This made it the first spacecraft to make very close observations of a comet.

Originally it had been planned that a US partner probe should join Giotto on its way to Halley. However, NASA canceled their part due to budgetary reasons. Giotto remained alone and the mission was carried out under the aegis of ESA.

Giotto is named after the Italian artist Giotto di Bondone (1266–1337) who had observed comet Halley in 1301 and portrayed it as the star of Bethlehem.

Giotto, the spacecraft, was equipped with several scientific instruments that allowed the optical and spectral observation of the comet and its plasma.

Scientific Goals of Giotto

Giotto was designed as a mature scientific space probe. It was the most advanced vehicle in the Halley Armada. Hence, a many scientific goals had been formulated for the mission. Giotto was supposed to return the first close-up images of a cometary nucleus. Something nobody had seen before and was still open to a lively debate. Then, it should determine the chemical and physical composition of Halley's coma. Additionally, the structure and composition of the dust particles ejected from the nucleus was unknown. Giotto was hoped to bring some light into this subject matter. The ratio of dust and gas was also unknown. The spacecraft should help with this, too. Furthermore, not much besides theoretical assumptions had been known about the interaction of solar wind and comet. It was to Giotto to do close measurements. In summary and without anticipating too much, it succeeded and achieved all the goals. The whole mission was more than a simple success story. Although not everything during the course of the mission went fine. An event close to Halley almost made the mission end in a total disaster.

Giotto and the Halley Encounter

Giotto was launched aboard an Ariana 1 rocket on July 2, 1985 and passed Halley on March 14, 1986 at a distance of only 596 km. However, a heavy impact 7.6 s before the probe's closest approach occurred and almost ended into a disaster. The spacecraft's camera and some other instruments were immediately destroyed. It was also worrying that the dust shield no longer protected the spacecraft. Giotto started to tumble. This was very dangerous, as the flight could no longer be controlled. Additionally, the tumbling caused the antenna, being necessary for communication with the ESA's control center in Darmstadt (Germany). By a hair's breadth Giotto was almost lost. However, within about 30 min it was able to stabilize the probe again. The mission could continue.

Another impact of cometary material destroyed the Halley Multicolor camera. Fortunately not before it could take some photos of the nucleus during the closest approach.

After the flyby of Halley, Giotto's trajectory was adjusted for an Earth flyby. The instruments were turned off on March 15, 1986 in order to save energy. On July 2, 1990 a wake-up signal was sent to Giotto from the control center. The probe became a new target assigned, the comet 26P/Grigg-Skjellerup which it passed at a distance of only 200 km on July 10, 1992. After the flyby the probe was switched off on July 23, 1992 and the mission officially ended.

Scientific Results

The Giotto mission was a great success and yielded a huge amount of scientific data. The images that were taken showed a peanut-shaped and very dark cometary nucleus. The very low albedo of 0.04 made it one of the darkest known objects in the solar system. The darkness was supposed to come from a thick layer of dust covering the nucleus.

Halley, the spacecraft determined, had dimensions of $15 \times 7 \times 10$ km and apparently only 10 % of its surface were active. This was very peculiar since until then it was common belief that the whole nucleus of a comet should be active.

The spacecraft also measured the amount of material and its composition that was ejected by the comet. About 80 % of the abundances were water, followed by 10 % of carbon monoxide and 2.5 % of methane and ammonia. Only very small traces of other material could be found such as hydrocarbons, iron and sodium.

The observed dust could be differentiated into two more or less independent groups. Firstly, the CHON-group composed of light elements such as carbon, hydrogen, oxygen and nitrogen. The second group is formed of elements that are usually the founding material of rocks such as calcium, iron, magnesium, silicon and sodium.

Analysis showed that the comet formed about 4.5 billion years ago. Volatiles that were present condensed at interstellar dust particles and created the basis of the cometary nucleus which then developed due to accretion and gravitational effects. The ratio of abundances of the light elements of the CHON-group excluding turned out to be the same as the Sun's. This suggests, that the comet consists of some of the oldest unaltered material in our solar system.

Giotto was able to observe several gas jets coming from the cometary nucleus. The involved scientists analyzed the jets and found out that the amount of material that was ejected by the jets was about 3 tons/s. This high output of material destabilized Halley's rotation. The comet was wobbling over a longer period of time. Just think about a similar experience on Earth. Imagine a water hose with water that is streaming out with high velocities. The hose will dance around.

Most of the dust particles ejected were very fine, like cigarette smoke particles with masses in the range of 10^{-20} kg to 40×10^{-5} kg. However, considering the two impacts the spacecraft suffered, it becomes apparent that not all of the particles had been that small and fine. On the contrary, quite a few particles must have been larger debris. At least the two objects that hit Giotto.

Further Reading[1]

Introduction

*Armitage PJ (2010) Astrophysics of planet formation. Cambridge University Press, New York
Evans J (1998) The history and practice of ancient astronomy. Oxford University Press, Oxford
Jones BW (2007) Discovering the solar system, 2nd edn. Wiley, Chichester, UK
*Lissauer JL, de Pater I (2013) Fundamental Planetary Science. Cambridge University Press, New York
Taylor SR (2008) Solar system evolution: a new perspective, 2nd edn. Cambridge University Press, Cambridge

Asteroids

Bell J, Mitton J (2002) Asteroid rendezvous: NEAR Shoemaker's adventures at Eros. Cambridge University Press, Cambridge
*Bottke WF, Cellino A, Paolicchi P (2002) Asteroids III. University of Arizona Press, Tucson, AZ
Dymock R (2010) Asteroids and dwarf planets and how to observe them. Springer, Berlin
Yeomans DK (2012) Near-earth objects: finding them before they find us. Princeton University Press, Princeton, NJ

Comets

Eicher DJ (2013) COMETS!: visitors from deep space. Cambridge University Press, New York
*Festou MC, Keller HU, Weaver HA (2004) Comets II. University of Arizona Press, Tucson, AZ

[1] This section is supposed to serve as a starting point for further reading and getting further in-depth details of the topics we have touched in this book. Reading material that requires more theoretical knowledge about physics, astronomy or planetary science is marked with an asterisk.

M. Moltenbrey, *Dawn of Small Worlds*, Astronomers' Universe,
DOI 10.1007/978-3-319-23003-0

Heidarzadeh T (2008) A history of physical theories of comets, from Aristotle to Whipple. Springer, Berlin
Meierhenrich U (2014) Comets and their origin: the tool to decipher a comet. Wiley-VCH, Weinheim
*Schulz R, Alexander C, Boehnhardt H et al (2009) ROSETTA: ESA's mission to the origin of the solar system. Springer, Berlin
Seargent DAJ (2008) The greatest comets in history: broom stars and celestial scimitars. Springer, Berlin
Stoyan R (2015) Atlas of great comets. Cambridge University Press, Cambridge

Trans-Neptunian Objects and Dwarf Planets

*Barucci MA, Boehnhardt H, Cruikshank DP (2008) The solar system beyond Neptune. University of Arizona Press, Tucson, AZ
Brown M (2012) How i killed Pluto and why it had it coming. Spiegel & Grau, New York
Davies J (2001) Beyond Pluto: exploring the outer limits of the solar system. Cambridge University Press, Cambridge
Jones BW (2010) Pluto—sentinel of the outer solar system. Cambridge University Press, Cambridge
Stern A, Mitton J (2005) Pluto and Charon, 2nd edn. Wiley-VCH, Weinheim
Weintraub DA (2007) Is Pluto a planet? Princeton University Press, Princeton, NJ

Index

A
Albedo, 44–46, 57, 58, 60, 61, 71, 86, 113, 147, 162, 167, 171, 172, 178, 180, 193, 197, 199, 205, 209, 210, 218, 250, 266
Anti-pluto, 171
Apollo, 25, 31–33, 253
Asteroid families, 36–39, 155
Asteroid main belt, 3–5, 7, 12, 13, 17, 25, 28, 29, 37, 38, 41, 44–46, 50, 57, 58, 61, 69, 98, 99, 101, 141, 148, 152, 155, 177, 194, 196, 197, 200, 201, 209, 221, 233, 243, 247, 257
Asteroids, 3–5, 7, 10, 12, 13, 15, 17–66, 68–71, 80, 81, 94, 98, 101–104, 106, 114, 121, 123, 141, 144, 147, 163, 175, 176, 179, 194, 196, 197, 209, 216–218, 220–224, 228–230, 233, 235–238, 240, 243, 246, 247, 249–255, 257, 263
Aten, 25, 32–33, 47
Atmosphere, 27, 55, 75, 77, 80, 86, 87, 89, 114, 117, 119–121, 127, 138, 139, 159, 162, 168, 169, 183, 184, 186–189, 192, 199, 201, 205–207, 214, 216, 226, 232, 239, 240, 242, 243, 256

C
Centaurs, 3, 22, 26, 51–56, 58–71, 143, 144, 160, 163, 215, 217, 240
Ceres, 17, 20–22, 34, 37, 39, 57–59, 61, 175, 177, 186, 194–202, 233, 236–238
Chariklo, 54–56, 66–68, 218, 240
Charon, 146, 180–182, 184, 186–193, 240–242, 244–246
Chiron, 54, 55, 64–68, 143, 144, 218, 240
Coma, 30, 31, 55, 65, 66, 77, 81–83, 86–94, 112, 124, 125, 127, 129, 131, 133, 135, 136, 139, 221, 226, 227, 255, 258, 264–266
Comet families, 99–102
Comets, 1, 3, 4, 17, 18, 21, 22, 25, 30, 31, 45, 51–56, 65, 73–118, 121, 122, 124–140, 142–145, 147, 157, 163, 166, 195, 204, 209, 215, 220, 221, 223, 224, 226–228, 230–233, 247, 250–252, 255–267
Cubewanos, 147, 153, 154, 158–161, 209, 212, 246

D
Dactyl, 68–71
Dawn, 20, 36, 60, 61, 63, 64, 124, 194, 198–201, 233–239
Deep impact, 259–263
Detached objects, 164–170, 204
Dirty snowballs, 77, 79, 143
Dwarf planets, 3, 4, 17, 20, 34, 37, 40, 56, 57, 59, 61, 69, 143, 146–148, 152, 155–158, 161, 162, 167, 171, 175–185, 187–215, 233, 236–240, 242

E
Earth Trojans, 41, 48
Ellipses, 4, 76, 94, 95, 100, 114, 124, 134, 138, 144
Encke, 76, 78, 84, 98, 113–114
EPOXI, 260–263
Eris, 148, 161, 167, 172, 175, 177, 181, 202–207, 212, 213, 215

M. Moltenbrey, *Dawn of Small Worlds*, Astronomers' Universe,
DOI 10.1007/978-3-319-23003-0

Eros, 18, 28, 29, 216, 217, 246–250, 252, 255
Extinct comets, 4, 30, 81, 104, 105, 144

F

Frost line, 10, 11

G

Gas giants, 1, 7, 11–15, 22, 23, 32, 35, 38, 43–45, 50, 53, 54, 56, 65, 94, 97, 98, 100, 102, 107, 111, 112, 115–117, 120, 126, 134, 140, 141, 144–146, 148–150, 154, 163, 164, 169, 178, 185, 194, 196, 204, 224, 244, 256
Giotto, 77, 79, 84, 85, 108, 109, 112, 113, 223, 251, 264–267

H

Hale-Bopp, 75, 78, 82, 87, 91, 94, 106, 123, 125, 127–137, 233
Halley, 55, 65, 73–79, 84, 85, 99, 100, 105, 107–114, 128, 131, 134, 223, 251, 264–267
Halley Armada, 84, 109, 112, 264–267
Haumea, 69, 155, 177, 181, 194, 202, 204, 207–213
Hubble Space Telescope (HST), 83, 103, 115–120, 122, 121, 123, 152, 167, 172, 182, 187, 190, 191, 193, 198, 199, 203, 205, 236, 237, 245, 262
Hyakutake, 93, 98, 123–129, 220, 233, 247
Hydra, 189, 193
Hygiea, 34, 37, 39, 57

I

Ice dwarfs, 3, 155, 176, 177, 182, 185, 204, 207
Ice giants, 12, 14, 24, 56, 100, 142, 145, 146, 150–152, 155, 157, 159, 162, 164, 180, 185
Ida, 37, 68–71

J

Jets, 2, 80, 82–84, 263, 264, 267
Jupiter, 1–4, 7, 10–15, 18–20, 22–26, 30, 31, 34, 35, 37–39, 41–43, 50, 53, 56, 60, 68, 81, 86, 94, 95, 98–102, 104–106, 110–112, 114–118, 121–123, 131, 134, 137, 140, 141, 143, 145, 148–152, 154, 163, 165, 169, 175, 176, 195, 196, 201, 202, 224, 243–245, 256, 259, 261
Jupiter Trojans, 22, 26, 41–46, 49, 50, 52, 58

K

Kerberos, 189, 190, 193
Kirkwood gaps, 35, 38, 62, 152, 201, 230
Koronis, 37, 69, 71
Kuiper belt, 4, 5, 7, 13–15, 56, 58, 67, 94, 96, 98, 100, 112, 132, 134, 137, 141–143, 145, 147–161, 163, 169, 172, 181, 185, 209, 212, 224, 240, 243, 245, 246, 259

L

Lagrange, 40, 42
Lagrange points, 26, 40, 41, 43, 45
Long-period, 41, 59, 144, 145, 163, 166, 204

M

Makemake, 177, 181, 194, 202, 204, 209, 212–214
Mars Trojans, 48–50
Mean motion resonances, 14, 15, 23, 24, 32, 34, 35, 37, 39–41, 44, 50, 51, 56, 67, 148, 150, 155, 156, 171, 184

N

Near Earth Asteroids (NEA), 25, 26, 28–34, 47, 54, 215, 246, 252
Neptune Trojans, 51, 52
New Horizons, 152, 157, 159, 184, 186, 189, 191–193, 233, 239–246
Nice model, 13, 43–45, 50, 58, 151
Nix, 189, 193
Non-periodic, 106
Nucleus, 55, 65, 77–89, 91–93, 95, 106, 109, 112, 113, 116, 117, 127, 128, 131, 135, 136, 138, 224, 226, 230, 233, 251, 258–267

O

Occultations, 61, 66, 67, 186, 189, 205, 206, 213, 214, 216–218, 242
1992 QB1, 3, 22, 146–148, 153, 159, 160, 181
Oort cloud, 13–15, 95–98, 100, 112, 124, 126, 127, 131, 132, 134, 137, 138, 141, 142, 144, 145, 150, 162–164, 166, 169

Orbital resonances, 6–7, 23, 47, 144, 153, 155, 159–163, 244
Orcus, 156, 171–173, 177, 202

P
Periodic, 6, 76, 95, 97, 100, 106, 107, 110, 112–114, 259
Planet X, 142, 175, 177–180
Planetesimals, 10–15, 18, 23, 24, 43, 44, 50, 59, 80, 96, 97, 142, 148–151, 165, 169, 185, 197
Plutinos, 56, 148, 155, 156, 171, 185, 186, 212
Pluto, 3, 4, 6, 34, 56, 111–113, 141–143, 146–148, 152, 153, 155–162, 169, 171, 172, 175–182, 215, 233, 239–246
Plutoids, 155, 176, 177, 182, 185, 204, 207
Proto-Kuiper belt, 13, 14, 23, 44, 45, 148–151, 161, 185, 202, 212
Protoplanetary disk, 8–13, 18, 23, 24, 80, 94, 146, 149, 156, 165, 169

Q
Quaoar, 158–159, 177, 202

R
Regolith, 70, 71, 198, 249, 255

S
Sakigake, 109, 265
Scattered disk, 13, 14, 56, 98, 100, 112, 141, 148–152, 159–164, 169, 212, 224, 240
Sedna, 164–170, 175, 177, 202, 204, 205
Sednoids, 164
Shoemaker-Levy 9 (SL9), 80, 81, 94, 100, 105, 114–122
Short-period, 56, 95, 98–100, 104, 105, 107, 112, 114, 116, 144, 145, 163, 250
Siding Springs, 137–140
Snowy dirtballs, 77
Spectroscopy, 45, 52, 126, 127, 135, 157, 172, 187, 189, 191, 219–220, 226
Stardust, 223, 253, 255–259, 262
Styx, 190, 193
Sublimation, 30, 66, 67, 80–83, 86, 87, 89, 101, 109, 127, 131, 136, 137, 144, 169, 198, 201, 202, 206, 214, 239
Suisei, 109, 264, 265

T
Tholins, 54, 168, 172, 213
Tisserand's parameter, 99
Titus-Bode law, 18, 19, 195
Trojan asteroids, 26, 40–52, 176, 240
Twotinos, 156

U
Uranus Trojans, 51

V
Vega 1, 223, 264, 265
Vega 2, 77, 79, 223, 264, 265
Venus Trojans, 46–48
Vesta, 20, 34, 36, 39, 57, 59–65, 196, 233, 235–238