Comets and the early solar system

Donald Brownlee

The material in comets hasn’t changed much over the past few billion years. Spacecraft that were sent to image, probe, and sample passing comets offer a unique look back in time.

Don Brownlee is a professor of astronomy at the University of Washington in Seattle.

Active comets are fascinating ephemeral bodies that have both thrilled and terrified people for millennia. As a bonus, when Earth passes through a comet’s debris trail, we see a dazzling meteor shower. The phenomena of comets are spectacular, but what they can tell us about the early solar system is extraordinary. Comets are the best-preserved samples of the materials that were involved in the construction of the solar system.

The Sun, planets, and smaller bodies of the solar system all formed from a short-lived disk of material that for centuries has been called the solar nebula. The disk formed by the gravitational in-fall of gas, largely composed of hydrogen and helium, and small solid grains containing the majority of the elements in the periodic table. It lasted a few million years, only a factor of 10 longer than the free-fall time of interstellar matter onto the disk. Within those few million years, solids in the disk assembled to form kilometer-sized planetesimals, a landmark event in the assembly of larger bodies. Almost all of the planetesimals, the solid building blocks of planets, either were consumed by planet formation, were ejected from the solar system, or fell into the Sun. But a tiny fraction survived and are now observed as asteroids and comets, the only bodies still in existence that preserve solid materials involved in the solar system’s formation 4.567 billion years ago.

Comets are distinguished and defined by their cometary activity, the ejection of detectable amounts of dust and gas, as seen in figure 1. Asteroids are the source of most meteorites. Both asteroids and comets are vital sources of information about the early history of the solar system. Recently, the NASA spacecraft Stardust and Deep Impact have probed comets in ways that have yielded exciting new clues about the materials and processes of the solar nebula.

Asteroids versus comets

There are only a few safe havens in the solar system where samples of the original planetesimals have survived in stable orbits for billions of years. Those regions are the familiar main asteroid belt between 2.1 and 3.3 AU from the Sun, the Jupiter Trojan region near 5.2 AU with the same orbital period as Jupiter, the recently discovered Neptune Trojan region at 30 AU with the same orbital period as Neptune, the Kuiper belt in the plane of the planets just beyond Neptune, and the spherical Oort cloud well beyond the planets.

The main-belt asteroids formed by accretion of material in the transition zone between the inner, Earth-like planets and the giant planets. That region is near the so-called snow line that separated regions of the solar nebula where water existed as vapor and where it condensed into ice. Many of the main-belt asteroids probably once contained some quantity of ice, even though the presence of ice has been a classic textbook discriminator between asteroids and comets. But there is evidence that during the first few million years of solar-system history, all the meteorite-producing asteroids were heated internally by the decay of short-lived radioactive isotopes to temperatures ranging from the melting point of water to the melting point of silicates and metallic iron. If astronomers had been available to observe asteroids during that time, they would have seen tails and comae, and they would have identified the asteroids as active comets. Even now, a few weakly active comets have recently been observed in the asteroid belt. Those rare comets with asteroidal orbits appear to have formed in place and retained ice for the age of the solar system.

Most of the asteroids in the mid-belt region are low-albedo objects with spectral reflectances similar to carbon-rich meteorites. Almost all carbon-rich meteorites contain bound water in hydrated silicates, phases that formed inside their parent asteroids by interaction with liquid water. The asteroids that best preserve nebular material are dominated by millimeter-sized silicate spheres that formed from individual nebular particles heated to their melting points by violent transient events such as shocks in the nebular gas.

With the exception of the rare asteroidal comets, most comets appear to have formed in more distant regions of the solar nebula and have been stored in reservoirs beyond Neptune for billions of years. Unlike most asteroids, comets are in unstable orbits at the time they are observed. Most comets cross planetary orbits, and encounters with planets limit their inner-solar-system lifetimes to 10^6 years. Comets are usually observed close to the Sun, where sublimation drives cometary activity and mass loss of about 0.1% per orbit. Active lifetimes, limited by mass loss, are even shorter than dynamical lifetimes. Surprisingly, the major life-limiting factor for many comets is splitting or even severe fragmentation—as shown in figure 2—a process whose cause remains mysterious. An active comet has about a 1% chance of splitting in a given year, and it may fragment many times during its active lifetime.

As mentioned previously, a few active comets are known to exist in the asteroid belt. Conversely, it is certain that some fraction of bodies that are considered to be asteroids are former comets that literally ran out of gas. A study of the albedos and orbits of asteroids near Earth has led to the suggestion that some 4% of them are actually inactive comets.
The origin of comets

Comets were traditionally thought to come from the Oort cloud, a hypothesized spherical distribution of more than $10^{12}$ bodies extending out to 50 000 AU from the Sun. Astronomer Jan Oort predicted the presence of that reservoir in 1950 to explain the delivery rate of comets with orbital periods longer than 200 years. Those long-period comets have highly eccentric, nearly unbound orbits with random inclinations. It is estimated that delivery of comets from the Oort cloud, at a rate of a few per year, is caused by gravitational perturbations external to the solar system. The Oort cloud is believed to be populated by planetesimals that were ejected from the planetary disk by perturbations from the giant planets.

However, the Oort cloud cannot produce the observed number of short-period comets. Most short-period comets have low-inclination prograde orbits with periods close to six years. Those whose orbits extend to near Jupiter are called Jupiter-family comets. It is possible for Oort-cloud bodies to enter JFC orbits via three-body interactions with Jupiter and the Sun, but not at the required rate. That problem was first pointed out in 1980 by Julio Fernández, who suggested that the short-period comets were instead derived from a flattened distribution of bodies just beyond Neptune, now known as the Kuiper belt. Perturbations can transport bodies from the Kuiper belt via successive planetary encounters to JFC orbits. Numerical simulations of orbital evolution have shown that the Kuiper belt can produce the observed short-period comets. Bodies in the Kuiper belt are not comets, because they do not exhibit cometary activity, but they would become comets if perturbed closer to the Sun. They are given the inglorious name “Kuiper-belt objects.”

In a triumph of theory and observation, the first KBO other than Pluto was discovered in 1992. There are now more than a thousand known KBOs, and new surveys such as the proposed Large Synoptic Survey Telescope are expected to discover some $10^5$ KBOs out of an estimated total population of more than $10^9$. The observation of the distant bodies by sunlight reflected from their dark reddish surfaces is aided by the number of large bodies. The largest KBOs are more than 2000 km in diameter, and there are many more 500-km KBOs than there are asteroids of that size. As discussed in an article by Michael Brown (PHYSICS TODAY, April 2004, page 49), the study of the Kuiper belt has opened up a new field in astronomy that provides a rich source of information on dynamical processes, including convincing evidence for large-scale outward drift of the outer planets.

Dust, gas, and rocks

The dust-to-gas ratios of comets have long been estimated by measuring the ratio of sunlight reflected off dust in a comet’s coma to light emitted from excited gas molecules. That method underestimates the solid component, however, because most of the ejected mass is not gas or dust but small rocks that are hard to detect because of their small ratios of surface area to mass. But the mass fraction of small rocks can be deduced from impacts measured by flyby spacecraft, from observations of cometary meteor showers, and from IR imaging.

The presence of rocks shows up in IR images as spectacular lines, like jet contrails across the sky. Gas, dust, and rocks leave comets at different speeds. Gas escapes the nucleus at sonic speed and is soon swept outward at more than 300 km/s by the magnetic field carried by the solar wind. Dust is initially pushed away by the drag force of escaping gas and reaches a size-related terminal velocity determined by momentum exchange with the gas. Dust trajectories deviate from the comet’s orbital path due to the radial component of the light-pressure force and the tangential component that produces drag and causes orbital decay. Rocks more than 100 μm in diameter, with their low area-to-mass ratios, are not accelerated to significant speed by escaping gas or strongly affected by sunlight pressure. Ejected at low speed, they spread along a comet’s orbit as a thin line, drifting from the comet because of their slightly different orbital period. The lines are called dust trails, and a recent compilation of IR data for eight comets gave a range of ejected rock–gas mass
ratios from 1.2 to 4.6 with an average of about 3. The mass of comets is thus dominated by small rocks, not ice or dust.

The ejection of water and rock quantitatively explains a once disturbing aspect of cometary orbits—their failure to obey the laws of orbital motion. As an irony of science, the comet named after Edmond Halley, the man who convinced Isaac Newton to publish his Principia and who arranged for its funding, appears to defy Newton's laws. Halley's comet did return as predicted, but it has never been on time. When it made its highly anticipated 1910 return, it reached its closest point to the Sun more than three days later than was predicted by pure Newtonian motion. Comet orbits always deviate at least slightly from pure gravitational orbits due to the rocket effect created by the anisotropic ejection of matter. The net force of the matter ejected from the solar-heated regions of the comet has a component that is parallel to the orbital path and can either add or subtract from the comet's total orbital energy.

Volatiles

Comets have been best known for their contents of volatiles, the driver of cometary activity. Apart from water, the dominant volatile, only carbon monoxide, carbon dioxide, methanol, formaldehyde, ammonia, and hydrogen sulfide have been observed at levels of 1% or higher in escaping comet gas. Comet gas also contains a complex brew of less abundant species, some of which are produced by decomposition of larger parent molecules. There are similarities between the gas released by comets and interstellar material, but with currently available information it seems that cometary volatiles are a likely mix of interstellar and nebular materials. The origin of volatiles and their evolution inside comets is a potentially complex matter, particularly so for comets that are collisional fragments of larger bodies similar to observed KBOs. Water is the major ice, and it has been suggested that many of the other volatile species are trapped in amorphous ice, an accommodating host that can form below 100 K and release trapped species upon crystallization above 120 K.

Although volatile abundances between comets show similarities, there are differences. For example, most JFCs are depleted in two-carbon and three-carbon species relative to...
Comets at first sight

Although comet tails and comae have been observed for millennia, no cometary body was imaged directly until 1986, when the European Giotto and Soviet Vega spacecraft flew by and took images of the nucleus of comet Halley. Subsequent spacecraft, Deep Space 1 to Borrelly, Deep Impact to Tempel 1, and Stardust to Wild 2, provided not only nucleus images such as the one in figure 3 but also insight into comets’ material properties and the origin of cometary activity.

The Borrelly images showed a dark nucleus and highly collimated jets of escaping dust. The jetting phenomenon, whereby water vapor was escaping at sonic or supersonic speed, was interpreted as volatilization of ice in localized regions beneath the surface. The surface images showed subtle features indicative of vertical cliffs and mesas—suggesting that solar heat may have reached ice-bearing material along crumbling cliff faces. The black surface of the comet appeared to be ice free.

The Deep Impact and Stardust flights both had the primary goal of studying materials from comet interiors, but both also provided high-resolution surface images of Jupiter-family comets. The surfaces of the two comets visited were quite different from each other. Tempel 1, imaged by Deep Impact, was relatively smooth but contained discrete units with different properties and presumably different histories. Figure 3 shows two smooth areas that appear to be related to some type of flow. Other images show large-scale layering, areas of surface deflation, and abundant large-scale heterogeneity. Tempel 1 appears to have classic impact craters, features not seen on the three other comets that have been imaged. The presence of craters implies that at least some of the comet’s ancient surface is preserved, presumably dating to the time the comet was in the Kuiper belt. The lack of classical impact features on the other comets suggests that their ancient surfaces have been lost through cometary activity. IR observations by Deep Impact detected only trace amounts of surface ice, which showed that cometary activity is, remarkably enough, driven by sublimation of ice below the surface.

Wild 2, the oblate comet visited by Stardust, is dramatically different from other imaged comets, asteroids, or moons. Its surface is rough on all spatial scales. The 4.5-km body, shown in figure 4, has several deep kilometer-size depressions that do not resemble impact craters. All the depressions lack raised rims, and some have flat floors, vertical walls, or even overhanging walls. The comet also has pinnacles, mesas, and other remnant features suggesting significant erosional loss of previous surface material. The truly wild surface of Wild 2, along with its distinction from other comets, is a mystery. It clearly has a young surface, so perhaps its extreme topography relates to the dramatic loss of volatiles in its recent past. The depressions might be related to impulsive mass-loss events like the ones in 2007 that caused comet Holmes to become a million times brighter in a single day. The extremely rugged surface of Wild 2 and the ubiquity of depressed surface features suggest that essentially all surface regions have been active at one time or another. If that is the case, then active regions turn on, deplete their near-surface volatile source, turn off, and then somehow become active again later.

Getting a closer look

Significant new information on the nature and origin of the nonvolatile components in comets has come from the recent NASA Stardust and Deep Impact missions. A common expectation was that the two visited comets would be largely collections of ice and preserved submicron interstellar grains coated with organic mantles. It was also commonly expected that comets formed in isolation from the inner solar system.

The two spacecraft differed significantly in the ways that they gathered information on cometary solids. Deep Impact cratered the surface of Tempel 1, creating a cloud of ejected subsurface material that could then be studied spectroscopically by Earth- and space-based telescopes. Stardust collected solid particles ejected from active regions on Wild 2 and actually returned samples to Earth, where they could provide information at the atomic scale that cannot be obtained by remote-sensing methods. Deep Impact inferred the presence of forsterite (Mg2SiO4) and enstatite (MgSiO3) along with many other phases. The same two minerals were found as major phases in the Stardust samples. Astronomers have also observed spectral evidence of forsterite in other comets and in dust disks around other stars.
Crystalline forsterite in comets and disks was thought to have formed by thermal annealing of interstellar “amorphous forsterite” (an oxymoron in the mineralogical community) heated to about 1000 K in the regions near stars. That hypothesis nicely explains the spectral data showing that silicates in interstellar space are amorphous, whereas silicates orbiting in disks around stars contain a major crystalline component. However, it is at odds with laboratory studies of the samples returned by Stardust.

Annealing is a solid-state process that allows atoms to diffuse slightly but does not totally transform the material. When a collection of submicror amorphous particles is annealed, the most likely result is an amorphous material with small crystals embedded in it. To form large crystals with no surrounding glass, like the 5-μm silicate crystals found by Stardust, the precursor glass would have to have had perfect mineral stoichiometry. It is a challenge to imagine how that could happen.

Forsterite is a magnesium silicate with very little iron. The forsterite crystals in the Wild 2 samples contain minor elements, such as aluminum, calcium, manganese, and chromium, in a range of abundances controlled by the elements’ ability to be incorporated into a crystalline lattice. But on average, interstellar grains must have approximately equal amounts of iron, magnesium, and silicon, like the Sun, and their minor-element abundances are not controlled by crystal chemistry. It seems unlikely that submicron glass particles of that composition could be annealed to form large, nearly iron-free crystals with distinctive minor-element patterns. Instead, the Wild 2 materials surely formed from matter that was transformed by vaporization, melting, or heating above 1000 K.

The minor-element abundances in the Wild 2 crystals can be used as fingerprints to infer origins and relationships to similar phases in meteorites. For example, both the comet and primitive meteorites contain rare forsterite crystals with more manganese than iron. That phase in meteorites is commonly thought to have formed by condensation from nebular gas, a process that occurs at about 1400 K—it’s never been suggested that it formed by annealing.

Isotopic measurements provide additional information about the origin of the Wild 2 materials. Many interstellar grains retain the isotopic signature of the stars from which they formed. The solar nebula, a mixture of materials from many different stars, had its own isotopic composition. Since Wild 2 originated in the outer part of the solar system, it was thought to be made up of isotopically varied interstellar grains, so Stardust would be sampling stardust. But actually, the great majority of the Wild 2 materials have solar isotopic composition, consistent with nebular origin. In fact, isotopically distinct pre-solar grains appear to be less abundant in the comet than in primitive meteorites that originated much closer to the Sun.

The Wild 2 samples also contain a remarkable array of other minerals, including sulfides, oxides, silicates, nitrides, and metal alloys. The particles larger than a few microns tend to be small rocks composed of several minerals and sometimes glass. Most of the observed components are nearly identical to the minerals found in primitive meteorites, but the diversity of materials in the comet exceeds that found in similar volumes of specific meteorite types. It seems that the comet accumulated materials from many different regions of the early solar system. The presence of so many materials that must have formed at high temperature is strong evidence of mixing between the inner solar nebula and the Kuiper belt. Based on the Stardust measurements, it appears that the majority of Wild 2 grains larger than a micron were transported from the center to the edge of the solar nebula just beyond Neptune. Although many of its components formed at high temperature, the comet itself is a primitive mix of well-preserved components. The mix of minerals is highly un-equilibrated in the sense that there is no evidence for thermal diffusion of atoms between adjacent grains, an effect seen in almost all asteroidal samples.

A remarkable finding in the Stardust sample was a calcium–aluminum inclusion. Also found in primitive mete-
orites, CAIs are composed of elements and minerals with high condensation temperatures. They are the oldest solar-system solids by about 2 million years. Another distinguishing feature of CAIs is their unusual oxygen isotopic composition, with more \(^{16}\)O and less \(^{17}\)O and \(^{18}\)O than other solids. It has been suggested that perhaps the “peculiar” oxygen isotopic composition of CAIs is actually the same as the Sun’s, and that other solar-system solids are isotopically heavier because of photochemical processes that occurred in the solar nebula. That suggestion is strongly supported by the first oxygen isotope measurements on solar wind samples returned to Earth by the NASA Genesis spacecraft.

The CAIs’ elemental and isotopic compositions suggest that they formed in a localized region of the solar nebula during a period as brief as 20,000 years. The discovery of CAI material in a Kuiper-belt comet—in a 20-μm Wild 2 particle nicknamed Inti, shown in figure 5—is highly significant. It constitutes nearly irrefutable evidence that material from the inner solar system was transported beyond the orbit of Neptune.

The Grand Radial Express

The discovery that comet Wild 2 is dominated by high-temperature materials that apparently formed in the inner solar system is evidence that there were effective processes for moving grains and small rocks from the center of the solar nebula to its edge, just beyond Neptune. The processes involved in transporting high-temperature minerals to the coldest, most isolated part of the solar nebula could be called the Grand Radial Express. Although they were not previously given appropriate acceptance, several proposed models have predicted transport processes both in and above the nebular plane. One of them, called the X-wind model, actually predicted that CAIs would be found in comets.

When stars and their accretion disks form, they are known to eject copious amounts of matter back into space. When the Sun formed, the amount of ejected material may have been as much as 0.3 solar mass, or 30 times the mass involved in forming the planets. According to the X-wind model, magnetic forces near the Sun stopped the inflow of mass along the disk midplane and forced it to follow field lines toward the Sun’s poles. When solids were lifted out of the midplane, they experienced severe radiation heating, and some were accelerated to high speed by the outward flow of the X wind. The X-wind model and the origin of outflows from young stars have stirred considerable debate, but there is no doubt that winds occur and that they involve an appreciable amount of mass. However, radial transport in the midplane of the disk due to turbulence or instabilities is currently the most popular model for large-scale radial transport in the disk.

As bodies that have retained frozen volatiles for the entire age of the solar system, comets offer cryogenically preserved samples of the materials that existed in the early solar system. The results from recent missions suggest that comets are constructed of low-temperature and high-temperature materials that formed in totally disparate environments. The frozen volatiles were formed at the edge of the solar system or in presolar environments, but most of the rocky materials were white-hot at the time of their formation and appear to have formed by violent processes in the hot regions of the solar system. The mixing of the high- and low-temperature components required large-scale radial transport of material from the central regions to beyond the most distant planet. In a sense, the solar nebula was turning itself inside out. It is likely that comets contain solids from the entire solar nebula, including fragmental debris generated by the planet-like ob-

References