A Brief Conceptual History of Cometary Science

M. C. Festou  
Observatoire Midi-Pyrénées

H. U. Keller  
Max-Planck Institute für Aeronomie

H. A. Weaver  
Applied Physics Laboratory of The Johns Hopkins University

The history of cometary astronomy can be naturally divided into five major periods, with each transition marked by an important new insight. Before 1600, comets were usually viewed as heavenly omens, or possibly meteorological phenomena in the terrestrial atmosphere, and were not yet clearly established as astronomical bodies. Then followed two centuries of mostly positional measurements triggered by the stunning discovery of the universal law of gravitational attraction. Two highlights from this period, which lasted until the early nineteenth century, were the successful prediction of the March 1759 return of 1P/Halley’s comet and the discovery of the nongravitational motion of Comet 2P/Encke. The era of cometary physics began with the passage of 1P/Halley in 1835, when spatial structures in a comet were described in detail for the first time. The year 1950 marked the emergence of the modern picture of comets as an ensemble of solar system objects composed of primordial ice and dust, generally on long-period orbits and shaped by their interactions with the solar radiation field and the solar wind. Finally, the space missions to Comet 21P/Giacobini-Zinner in 1985, and especially to 1P/Halley in 1986, provided the first in situ measurements and the first images of a cometary nucleus. While these in situ observations significantly improved our understanding of cometary phenomena, they also posed many new questions for which we are still seeking answers. In this introductory chapter, we only briefly discuss the pre-modern observations of cometary phenomena, which are already well described in the monograph by Yeomans (1991), and focus instead on the advances in cometary science during the past 65 years or so, especially on the developments since the publication of the Comets book in 1982.

1. PRE-MODERN ERA: BEFORE 1950

The word “comet” comes from the Greek “kometes,” which literally means “long-haired,” but the earliest extant records of cometary observations date from around 1000 BC in China (Ho, 1962) and probably from about the same time in Chaldea. Ideas about the true nature of comets are available from the time of the rise of Hellenistic natural philosophy around 550 BC, when the Pythagoreans considered comets to be wandering planets that were infrequently seen, mostly near the horizon in the morning or evening skies. In his Meteorology (ca. ~330 BC), Aristotle relegated comets to the lowest, sublunary sphere in his system of spherical shells and described them as “dry and warm exhalations” in the atmosphere. There is no mention of comets in Ptolemy’s Almagest, presumably because he did not consider them to be of celestial origin, but he described them in astrological terms in his Tetrabiblos. The Aristotelian ideas about planets and comets were upheld for an entire millennium during which there was little scientific advancement in the field of astronomy. The first doubts about the Aristotelian view seem to have been expressed by Aquinas and by Bacon in his Opus Tertium from 1267, although both men, like most of their contemporaries, strongly believed that comets were evil omens.

Toscanelli observed 1P/Halley in 1456 and several other comets between 1433 and 1472 with improved accuracy, inaugurating the renaissance of European observational astronomy after the long post-Aristotle scientific freeze. Brahe’s (1578) exceptionally accurate observations initiated a new era for observational astronomy, as he beautifully demonstrated that the horizontal parallax of the bright Comet C/1577 V1 was certainly smaller than 15 arcminutes, corresponding to a distance in excess of 230 Earth radii, consequently farther away than the Moon. This result raised the question of how comets move, and the suggestion that they orbit in highly elongated ellipses was made in 1610 by the amateur astronomer Lower (Rigaud, 1833). At about the same time, Hooke and Borelli suggested that cometary orbits could be parabolic. Dörrfel was the first to state specifically that the two bright comets seen in 1680 and 1681 (C/1680 V1) were one and the same, before and after its perihelion passage, and that it moved along a parabola with the Sun as its focus, thus providing an explanation for the fact that many comets were seen in pairs, one in the morning and the other in the evening. Newton, in his Principia
Newton, 1687), developed the brilliant tool that could link all these observations together. He applied his new theory of gravitation to show that the comet of 1680 (C/1680 V1) moved in an elliptical, albeit almost parabolic, orbit and that it passed only about 0.00154 AU above the surface of the Sun. Halley (1705) computed the orbits of a dozen well-observed comets and demonstrated the periodic nature of the bright comet of 1682 (1P/1682 Q1). “Halley’s comet” was telescopically recovered by Palitzsch in December 1758, which conclusively proved the validity of Newton’s law of gravity out to the distance of the aphelion at 36 AU, more than three times the distance of Saturn, the outermost planet known at that time.

Cometary astronomy in the eighteenth century witnessed the gradual improvement in techniques for orbital computations (e.g., Brandt, 1985; Yeomans, 1991), and by the beginning of the nineteenth century this had become a rather straightforward task, albeit a somewhat arduous one when planetary perturbations had to be taken into account. Some basic features of the orbital distribution of comets were established, e.g., the extremely broad range of orbital periods. While some comets were found to have orbits virtually indistinguishable from parabolas, others were clearly confined to the inner solar system inside Jupiter’s orbit. As time passed, a concentration of comets moving in similar orbits with fairly low inclinations and with aphelio close to Jupiter’s orbit became more and more obvious, giving rise to the concept of the Jupiter family of comets. Two ideas were proposed to explain the existence of this family: Either there was a continual ejection of comets from Jupiter (Lagrange, 1814), or there was a mechanism of dynamical evolution, called “capture,” whereby the comets would become concentrated into such orbits (Laplace, 1816). It was soon recognized that comets in general, and members of the Jupiter family in particular, suffer by far their largest orbital perturbations due to the action by Jupiter. The restricted Sun-Jupiter-comet three-body problem consequently offered an interesting approximation for the study of their dynamical behavior.

In 1835, 1P/Halley became the first comet for which spatially detailed structures were extensively observed. In particular, Herschel, Bessel, and Struve described the presence of jets, cones, and streamers [cf. reproductions appearing in Donn et al. (1986)]. This led Bessel (1836b), following Olters (1812), to postulate the ejection of solid particles in the direction of the Sun, and that these particles were somehow forced back into a “tail” by an unknown repulsive force acting in the anti-sunward direction. The connection between the Sun and comet tails had been suspected for a long time but never expressed so clearly before. Comets were now identified as physical entities, and not solely masses circulating around the Sun (described later as “visible nothings” by Struve because of their inability to leave any sign of their existence — e.g., against the solar disk or in front of stars — except for their response to the Sun’s gravitational pull). Arago directed his polarimeter toward the tail of Comet Tralles (C/1819 N1) and found the light to be polarized, hence it was “reflected” sunlight. He could not relate this observation to the presence of solid particles or gases because their respective effects on solar light were then unknown. The observation by Bond in 1862 of Comet Donati (C/1858 L1) revealed that the plane of polarization of the light was the Sun-Earth-comet plane, thus definitively demonstrating the solar origin of the scattered light. Around the same time, observations of Comet Tebbutt of 1861 (C/1861 J1, the great comet) by Secchi showed that the center of coma light was not polarized while the outer coma was. Other observations indicated different results, which we interpret today as due to the varying gas-to-dust ratio in the coma. As publications during this period demonstrate, these observations were ahead of their time and could not be properly interpreted. In particular, it is striking to observe that even though in 1940 the polarization mechanisms were all understood [“reflection” or “scattering” by particles, a surface, or gases (see Öhman, 1939, 1941)], no mention of “solid particles” as constituents of cometary coma was made. In his review of 1942, Bobrovnikoff (1942) mentions “meteoric dust” as a possible participant in producing comet spectra, since the nucleus must be made of meteoritic material, but he also said that “the continuous spectrum is still a mystery.”

The link between comets and meteors was made by Schiaparelli (1866, 1867), who observed that the Perseid and Leonid meteor streams coincided with the orbits of Comets 109P/Swift-Tuttle (1862 III) and 55P/Temple-Tuttle (1866 I) respectively. This was the proof that comets were indeed losing solid particles. Bredikhin (quoted by Jaegermann, 1903) further developed the comet tail theory based on an ad hoc repulsive force from the Sun that varied with the square of the heliocentric distance. This became known as the Bessel-Bredikhin mechanical model and was widely used. Finson and Probstein (1968a,b) published a gas dynamic model describing the gas-dust interaction and the solar light-dust interaction that is still in widespread use today. Eddington (1910) introduced the fountain model of particle ejection, in which parabolas represent the outer envelopes of particle trajectories emitted from the sunlit hemisphere of the nucleus. The repulsive force acting on the particles was identified by Arrhenius (1900) as the pressure exerted by sunlight. The corresponding theory was further developed by Schwarzschild (1901) and extended to molecules by Debye (1909). (At this point, it is worth mentioning that neither observations nor ejection models suggested that comets were losing material from the nightside.)

After Halley, Encke (1820) was the second to successfully predict the return of a comet (in 1822). Comet 2P/Encke has the shortest period of all known comets, 3.3 years, which provides similar Sun-Earth-comet configurations every 10 years. The comet was subsequently found to arrive systematically at perihelion about 0.1 days earlier than predicted, even when taking planetary perturbations into account. Inspired by his observations of an asymmetric distribution of luminous matter in the head of Comet Halley in 1835, Bessel (1836a) interpreted this as a Sun-oriented
asymmetric outflow and suggested that a nongravitational effect might arise due to the rocket-type impulse imparted by such an outflow, possibly explaining perihelion shifts such as those observed for Comet 2P/Encke. It would take over a century for this superb idea to be fully accepted by the scientific community because the existence of a solid body at the center of the coma was far from being unanimously accepted. In fact, the theory that comets were a swarm of solid particles was the most favored by scientists at that time.

The first spectroscopic observations of the gas component of comets were made by Donati (1864) and Huggins (1868), who visually compared the spectra of Comets Tempel (C/1864 N1) and Tempel-Tuttle (55P/1865 Y1), respectively, with flame spectra. They found that the bands seen in the comet and in the flame were similar. Huggins also visually noted the presence of a broad continuum, which he identified with reflected sunlight. Sunlight was known since the work of Fraunhofer to be characterized by the presence of numerous absorption lines, in particular the strong H and K lines in the UV region and the Na D lines in the yellow-orange region. Huggins was not a professional astronomer. Nor was Usherwood, who in 1858 took the first photograph of a comet, C/1858 L1 Donati. The bands recorded by Huggins, known as the “carbon” or “Swan” bands, were found in all subsequent observations of comets. The Swan bands so strongly dominated cometary spectra that carbon was immediately believed to be an important constituent of comets. While Draper was taking the second-ever image of a comet, C/Tebbutt (1881 K1, the great comet), Huggins was recording the first photographic spectrum of a comet when observing the same object through a slit spectrograph designed to record stellar spectra. Photography and spectroscopy soon became the standard way of studying comets. In 1882, the Na D doublet was identified in the (bright) comets of that year that passed close to the Sun. A few other emissions were seen, but those would not be identified for a few more decades. In particular, spectra of the tail were taken at the beginning of the twentieth century in Comets Daniel (C/1907 L2) and Morehouse (C/1908 R1). The N2+ emission was found near 3914 Å in the tail spectrum of the latter comet, but it was likely an atmospheric feature as, in the light of modern observations, a careful subtraction of the telluric lines is required to allow any firm identification of these faint emissions. In addition, such identification work is extremely difficult when working with photographic prism objective spectra.

Baldet (1926) published a comprehensive catalog of the spectra of 40 comets obtained since 1864, together with a complete bibliography of all comets observed until that time by spectroscopy. This work, and that of Nicholas Bobrovnikoff on the 1910 apparition of Comet 1P/Halley, which appeared five years later (Bobrovnikoff, 1931), are the first two comprehensive papers of cometary physics published in the twentieth century. Because only very bright comets were studied at that time, all observed comets appeared to have very similar spectra, although differences were qualitatively noted, in particular the relative strength of the continuous and discrete (dubbed “emission lines”) components of the spectrum. These differences were not understood, since neither the emission mechanism of the light nor coma abundances were known. Schwarzchild and Kron (1911) studied the intensity distribution in P/Halley’s straight tail during the 1910 passage and suggested that the emission could be explained by the effect of absorption of solar light, followed by its reemission, i.e., fluorescence. This naturally led to the key result obtained by Polydore Swings (Swings, 1941), who solved the long-standing problem of why the violet CN bands (3875 Å) in cometary spectra did not resemble CN laboratory spectra. Because of the presence of absorption lines in the solar spectrum, the intensity at the exciting wavelength depends on the Doppler shift caused by the comet’s motion relative to the Sun, which thus determines the strength of the emission lines in the comet’s spectrum; this phenomenon is known today as the “Swings effect.” Many papers written during this time describe the spectral properties of specific features in comets, such as the “central point” or “jets,” but are of little value today because of the poor spatial resolution of the observations.

2. MODERN ERA: AFTER 1950

A major evolution in cometary science took place in 1950–1951 with the formulation of three very important ideas within a short timespan. The icy-conglomerate (“dirty snowball”) model of the cometary nucleus was proposed by Whipple (1950). Then, from dynamical studies of the distribution of semimajor axes of comets, came the identification by Oort (1950) of a distant population of comets now known as the Oort cloud. Finally, Biermann (1951) gave the correct explanation for the motions of features in cometary plasma tails caused by their interactions with a flow of charged particles emanating from the Sun’s surface (i.e., the solar wind). None of these ideas resulted directly from new observational evidence, and important parts of them had been proposed earlier by other scientists, but this was the first time that the known facts were effectively combined, leading to a comprehensive description. A new picture of comets, and the existence of a family of celestial bodies, were suddenly revealed at the same time.

2.1. The Icy-Conglomerate Model of the Nucleus

In a series of enlightening papers published between 1932 and 1939, Wurm suggested that since the radicals and ions observed in cometary coma were not chemically stable, these species must be created by photochemistry of more stable molecules residing inside the nucleus [see, e.g., the reviews by Wurm (1943) and Swings (1943) and references therein]. It is Wurm who first proposed the concept of “parent” compounds/molecules in the nucleus, while “daughter” species would be created in the coma by photochemistry. Spectroscopic studies revealed the nature of the daughter species but not the parents. In the 1940s, Swings developed his ideas based on Wurm’s reasoning, and the
key role of these two scientists appears to have been overlooked in the later literature. The presence of CO or CO$_2$, C$_2$N$_2$, CH$_4$, N$_2$, and NH$_3$ in the comet was invoked on the basis that CO$^+$, CN, CH, CO$_2^+$, N$_2^+$, and NH emissions were found in cometary spectra. H$_2$O was also considered as a potential parent molecule, following the discovery of the OH 3090 Å UV emission in 1941. Additional molecules (e.g., C$_2$) were required since all known coma species could not be explained from the above molecules. In 1948, Swings came very close to proposing an icy model for the nucleus by suggesting that the above-mentioned molecules could exist in the solid state in the nucleus (Swings, 1943, 1948). (We note here that quantitative arguments were largely missing from the discussion because of the lack of appropriate observational material and laboratory measurements on the parent molecule properties. The situation significantly improved in that regard about 30 years later.)

A key question was how molecules were stored in the cometary material. Since the middle of the nineteenth century, a great deal of research had concentrated on understanding the nature of the central source of gas and dust in comets, and the alternate theory to Lyttleton’s sandbank model (Lyttleton, 1948) was that an invisible solid nucleus was the source of all observed cometary material. Swings (1942) suggested that molecules similar to those found in meteorites could possibly be stored in the nucleus by occlusion. Independently (probably) of Swings, Levin (1943) developed his desorption theory from the surface of meteoritic material to demonstrate that his sandbank model for the nucleus had a solid basis. An average desorption heat of 6000 cal/mole was deduced from the interpretation of the brightness law followed by comets as the heliocentric distance changes, which was in agreement with laboratory data for the above-mentioned cometary molecules (in the 2,000–10,000 cal/mole range). However, the amount of material that could be desorbed from a swarm of particles with an expected cometary mass could not explain the persistence of comae over several months during single passages, and consequently the survival of comets such as 1P/Halley or 2P/Encke for centuries and millennia. Seeing-limited observations of comets passing near Earth showed a central, unresolved light source of dimensional upper limits in the 10–100-km range (cf. review by Richter, 1963). Upper limits to cometary masses had also been estimated from the absence of evidence for mutual gravitational attraction of the components of Comet P/Biela in 1846, or of any influence on the Earth’s orbit for very close passages, like those of Comet P/Lexell (D/1770 L1). Masses in the 10$^{12}$–10$^{17}$ kg range were surmised (Whipple, 1961).

In an attempt to synthesize all known facts about the cometary nucleus, and with particular attention to the long-standing problem of explaining the nongravitational perihelion passage delays, Whipple (1950, 1951) laid the foundations for the modern model of a solid nucleus. Building on the ideas of Laplace (1813) and Bessel (1836a) (cf. Levin, 1985), Whipple (1950, 1951) described the nucleus as a mixture of ices from which the gases in the coma are produced by sublimation in increasing quantities as the comet approaches the Sun and the surface temperature of the nucleus rises. Meteoritic dust is also released from the nucleus when these ices evaporate, hence the famous expression “icy conglomerate.” The relative proportions of the various ices were discussed only qualitatively; not only was the nature of all the parent species needed to explain the unknown coma composition, but Whipple’s main concern was to explain the non-Keplerian motion of Comet 2P/Encke. Nevertheless, Whipple’s model was hugely influential because of its ability to successfully explain within a single conceptual framework many observed cometary phenomena, such as (1) the large gas production rates [200 kg/s of C$_2$ for 1P/Halley in 1910 derived by Wurm (1943)], for which the desorption model was totally inadequate; (2) the observed jet-like structures in the coma and the erratic activity, impossible to produce if the source of gas and dust was a cloud of particles; (3) the observed nongravitational forces by means of the momentum transfer by the outflow of gas from the nucleus, the sign of the net effect on the orbital motion being dependent on the orientation of the nucleus spin vector; (4) the fact that most comets that pass extremely close to the Sun, e.g., the Kreutz Sun-grazing comet group, apparently survive such approaches; and (5) the fact that comets are the sources of meteor streams. Items (2)–(4) gave particularly strong arguments for a solid nucleus rather than a sandbank structure, a model that had other difficulties in addition to failing to explain the above-mentioned points.

Although Whipple’s model quickly won general acceptance, there were some shortcomings. The main one was described by Whipple himself, namely, the large differences among the latent heats of vaporization of the various ices he thought might be present in the nucleus. He also remarked that the low vapor pressure of water was a serious problem when explaining the observed presence of the OH emission far from the Sun. As a result, the highly volatile material should be removed from the surface layer of the nucleus long before perihelion, in contradiction with the observation of radicals and ions such as CH and CH$^+$ near the Sun. This objection was tentatively removed when Delsemme and Swings (1952) noticed that almost all parent molecules (except NH$_3$) required to explain the observed radicals and ions in comets could coexist in the nucleus in the form of solid clathrate hydrates of H$_2$O, where the volatile “guest molecule” occupies a cage in the H$_2$O crystal lattice. From the stoichiometry of hexagonal ice, the mean value of the occupation number is somewhat smaller than 6. In this way, the highly volatile material does not disappear too rapidly, and is also freed together with less-volatile molecules, which explains why the spectrum remains more or less similar throughout a comet’s apparition. Delsemme and Swings’ ideas also implied that comets contain mostly water molecules, something that was not proven at that time, nor even discussed in any detail.

### 2.2. The Oort Cloud

As a consequence of the nineteenth century work on cometary orbits and the discovery of nongravitational
forces, dynamical studies of individual comets were carried out during the first decades of the twentieth century, with particular attention paid to the influence of planetary perturbations. The earlier results on the non-existence of interplanetary ether were confirmed and statistical considerations about the distribution and dynamical origin of comets naturally followed, including the question of whether or not some comets have “original” hyperbolic orbits (reciprocal semimajor axis \(1/a_{\text{orig}} < 0\)), which would mean that they were not members of the solar system. The work by Strömgren (1914, 1947) and colleagues demonstrated that there were no original hyperbolic orbits among the observed comets; all apparently hyperbolic orbits were actually perturbed into those states by planetary effects, mainly the influence of Jupiter. Comets were not coming from interstellar space. Sinding (1937) produced a list of values of \(1/a_{\text{orig}}\) for 21 long-period comets, which, together with the work by van Woerkom (1948), formed the basis for Oort’s famous paper (Oort, 1950) on the existence of a cometary population residing in the outer reaches of the solar system. The idea of a cloud of distant hypothetical comets, stable against stellar perturbations, necessary to explain the fact that many comets had \(a_{\text{orig}} > 10,000\) AU, had been expressed earlier by Ůpik (1932). Oort (1950) deduced the existence of such a cloud by studying the actual distribution of the semimajor axes of 19 observed comets. He discussed the important excess of long-period comets with \(1/a_{\text{orig}} < 10^{-4}\) AU\(^{-1}\), i.e., with aphelia beyond 20,000 AU, and concluded that while comets can remain in stable orbits out to distances of about 200,000 AU, some of them could from time to time be diverted inward by the perturbations of passing stars. These stellar perturbations should have randomized the orbital inclinations of the comet orbits over the age of the solar system. Oort estimated that the cloud should contain about \(2 \times 10^{11}\) comets to explain the discovery rate of new comets each year. With a mean cometary mass of \(10^{13}\) kg, the total mass of the Oort cloud would be \(-2 \times 10^{24}\) kg, or \(-0.3\) \(M_{\text{gh}}\).

Based on van Woerkom’s (1948) theory of the orbital diffusion caused by planetary perturbations, Oort found that the number of comets with \(1/a_{\text{orig}} \leq 10^{-4}\) AU\(^{-1}\) is much larger than what would be expected from the population of long-period elliptical orbits, which suggested that many of the comets become unobservable after their first passage through the inner solar system. This observational fact still does not have a universally accepted explanation. In a subsequent study, Oort and Schmidt (1951) distinguished between “new comets” (those making their first visit near the Sun’s neighborhood and the planets) and “old comets” (those returning on much less elongated elliptic orbits). The former appeared to be dustier and to brighten more slowly than the latter. These kinds of analyses were revisited later to include the role of interstellar clouds and the galactic tide in delivering new comets to the planetary region. However, the basic concept of the Oort cloud as an outer halo of the solar system has been substantiated by later studies, based on continuously improved cometary orbits by Marsden and co-workers [Marsden and Sekanina (1973), Marsden et al. (1978), and the catalogs of cometary orbits regularly published by the Central Bureau for Astronomical Telegrams].

Our ideas regarding the long-term dynamics of the Oort cloud have evolved considerably over the years. While passages of individual stars dominated the discussion in early investigations, the tidal effects of the galaxy as a whole have become recognized in the last two decades as the prime mechanism providing new comets to the inner solar system from the outer part of the cloud. The dramatic effects that might follow upon close encounters with massive perturbers, such as giant molecular clouds (Biermann and Lüst, 1977), also received a great deal of attention.

Although Oort explored the possibility of having an inner cloud extending inside of 20,000 AU, his preferred model was a thick shell of comets near the outskirts of the solar system. The idea of an inner Oort cloud was proposed later by Hills (1981) for two reasons: (1) to explain the replenishment of the Oort cloud inevitably depopulated by the above-mentioned perturbations, and (2) to provide another source region for Oort cloud comets besides scattering by the giant planets, which is a very inefficient process. These ideas are currently being completely revisited to incorporate the existence and role of the recently discovered transneptunian belt and the various subpopulations of transneptunian objects.

In most theories, Oort cloud comets formed in the Jupiter-Neptune region. The mass of the solar nebula and the time of formation of the planets are key ingredients of Monte Carlo simulations that investigate the transfer of comets from the region of the outer planets to the Oort cloud. The number of comets currently residing in the Oort cloud required to explain the discovery rate of new comets is on the order of \(1–2 \times 10^{12}\) (Heisler, 1990; Weissman, 1996). Recent simulations indicate that comets that were formed in the Saturn-Uranus region currently make up the bulk of the surviving Oort cloud comets, whereas few comets formed near Jupiter are left [see Dones et al. (2004) for a current perspective on the complex formation and evolution of the Oort cloud].

### 2.3. Ion Tails and the Solar Wind

Because the tails of comets can be so impressive, both to the layman and to the professional astronomer, they have been the subject of many investigations. Astronomers of all eras have been struck by the fact that the tail appearances may vary dramatically from one comet to another. In the early twentieth century, it was deduced from the study of the motion of kinks and knots in the tail that the antisolar force acting on straight tails was enormous, up to 1000 times the solar gravity. As early as 1859, Carrington (1859) suspected a physical connection between a major solar flare and enhanced magnetic activity on Earth some hours later. Ideas about the possible existence of a stream of particles from the Sun, perhaps electrically charged, emerged toward the end of the nineteenth century, in particular to explain the excitation of molecules and ions observed in cometary co-.ma. It was also found that cometary ion tails (formerly
called Type I tails) develop closer to the Sun than dust tails (formerly called Type II tails; the comet tail type is directly related to the strength of the repulsive force that obviously varies smoothly, hence the difficulty of any categorization scheme). However, it was only 50 years later that Hoffmeister (1943) provided the crucial observations of a gas tail aberration of 6°, i.e., the angle between the observed tail and the antisolar direction. This was correctly interpreted by Biermann (1951) in terms of an interaction between the cometary ions in the tail and the solar wind (hereafter SW), a continuous stream of electrically charged particles from the Sun with velocities of several hundred kilometers per second. His derived plasma densities were unrealistically high, since electrons were thought to accelerate the cometary ions. Alfvén (1957) removed this discrepancy by introducing the role of a “frozen” interplanetary magnetic field, which is carried along with the SW particles. Until space probes could study the SW in situ, cometary ion tails were the only well-distributed SW probes in interplanetary space and they largely remain so for regions outside the ecliptic plane. The existence of sector boundaries in the SW explains the separation of ion tails from the head of comets; this is one of many phenomena that indicate that the SW properties may change on very short timescales, as described in numerous papers published beginning in 1968 by Brandt and collaborators.

3. SPACE MISSIONS TO COMETS

Following the increased interest in comets that arose in the late 1970s, due in large part to the predicted return of 1P/Halley, another giant leap in our understanding of comet phenomena occurred in March 1986 when six spacecraft (henceforth “S/C”) made in situ observations of this comet. There are undoubtedly “pre-Halley” and “post-Halley” eras (much as historians describe the transition from the Dark Ages to the Renaissance period); for the first time, the nucleus of a comet was seen. However, the first cometary encounter took place six months earlier — on September 11, 1985 — when the International Cometary Explorer (ICE) passed through the tail of Comet 21P/Giacobini-Zinner, ~8000 km from the nucleus. The main results were the confirmation of the plasma tail model, indications about the ion composition, and the detection of a neutral current sheet at the center of the tail. ICE continued on to register the effects of 1P/Halley on the interplanetary medium from a distance of 28 × 10^6 km sunward.

Five spacecraft encountered 1P/Halley in early 1986: Vega 1 (March 6, closest approach distance of 8890 km), Suisei (March 8, 150,000 km), Vega 2 (March 9, 8030 km), Sakigake (March 11, 7 × 10^6 km), and Giotto (March 14, 600 km). Concurrently, an unparalleled, long-term Earth-based observational effort was coordinated by the International Halley Watch (IHW) (Newburn and Rahe, 1990). The IHW archive, with more than 25 GB of data, was released in December 1992 (International Halley Watch, 1992), and the associated summary volume (Sekanina and Fry, 1991) contains detailed information about the data obtained within the various IHW networks. The observations were made by both professionals and amateurs in all wavebands from the ultraviolet (UV) at 120 nm to the radio at 18 cm. It was particularly fruitful to combine space- and Earth-based observations for calibration and long-term monitoring purposes. The earlier cometary models could be tested and refined with the aid of the in situ measurements, leading to many new insights.

The nucleus was observed at close range for the first time; it was found to be larger (equivalent radius 5.5 km) and darker (albedo ~4%) than expected. In the Giotto images, surface features (craters, ridges, mountains, etc.) and source regions were observed (Keller et al., 1988). There were no signs of nightside outgassing. The coma was found to be highly structured on all scales (jets, shells, ion streams, etc.) and the gaseous component was analyzed in situ by mass spectroscopy. Signals at atomic masses of 1 and from 12 to ~55 amu were detected. H_2O was confirmed to represent 85% by weight of the gas phase (see further discussion below), and the likely presence of large organic polymeric molecules was indicated. The dust was analyzed by size and composition and there was an unexpectedly high fraction of very small grains, down to the sensitivity limit of ~10^{-19} kg. Particles rich in metals and silicates were found as expected, but particles rich in H, O, C, and N (“CHONs”) were seen for the first time and were thought to be related to the smallest grains mentioned above (Kissel et al., 1986).

The integrated mass loss experienced by the nucleus at this passage, on the order of 4 × 10^{11} kg (but very uncertain) was ~0.5% of the total mass of the nucleus, estimated at 1–3 × 10^{14} kg. Nucleus images taken in situ and ground-based observations were not sufficient to unambiguously determine the complex (excited) rotational state of the nucleus. A cavity devoid of magnetic field was detected within 5000 km of the nucleus. The various predicted plasma effects were confirmed, including the existence of a bow shock, and the adjacent interplanetary medium was found to be kinematically and magnetically extremely turbulent. Some of the species invoked to explain the mass spectra were produced by gas phase reactions in the coma, as anticipated a decade earlier by Oppenheimer (1975).

The fast flybys of Comets 1P/Borrelly in September 2001 (NASA’s Deep Space 1 mission) and 81P/Wild 2 in January 2004 (NASA’s Stardust mission) have produced two new images of cometary nuclei. In some respects, these two nuclei are very similar to that of 1P/Halley, i.e., all are very dark objects with complex surface structures. There appear to be some significant differences among the three nuclei (e.g., the more spherical shape and possibly “younger” surface of 81P/Wild 2), but the different spatial resolutions of the three investigations may account for some of the apparent diversity. 81P/Wild 2 is covered by what will probably soon be called “erosion craters,” but we must await the full publication of the results, and probably in situ investigations of other nuclei in the future, to understand how these craters are activated and how long they survive. De-
spite the advances enabled by the spacecraft encounters, our knowledge of the composition of the non-icy component will have to await new in situ measurements or, better yet, the return to Earth of a coma dust sample, very likely in 2007 from the Stardust experiment.

4. THE INVENTORY OF COMETARY VOLATILES AND COMPARATIVE COMETOLOGY

Although number density estimates for cometary comae had been derived since the time of Wurm’s investigations in the 1930s, the figures obtained were rather uncertain and their reliability limited by the lack of quantitative information on the excitation mechanisms for the observed emissions. Thus, it is not too surprising that, continuing the earlier investigations by Swings and McKellar, most spectroscopic studies between 1950 and 1970 were devoted to a never-ending attempt to discover and identify new emission lines and bands, as well as unraveling the structure of the ro-vibrational bands of the comet radicals and ions. In that regard, special reference must here be made to the numerous and important contributions from the “Liège school,” reviews of which are given by Swings (1956) and Arpigny (1965). During this epoch, rather complete and fairly accurate models of the fluorescence of the CN, CH, OH, and C₂ radicals were built. The advent of high-resolution spectroscopy in the late 1950s allowed the identification of many unknown lines, most of which were due to C₂ and NH₂. Despite these efforts, it is worth noting that thousands of lines in the optical spectra of comets remain unidentified even today; the most likely candidate molecules responsible for these emissions are S₂, CO⁺, CO₂⁺, and C₃ in the near-UV, C₂ and NH₂ in the optical, and NH₂ and H₂O⁺ in the optical infrared (IR). Many new lines have been recently discovered in the near-IR and radio regions (the submillimeter region is also becoming increasingly accessible), and these domains eagerly await a new generation of cometary spectroscopists.

5. WATER AS THE MAIN CONSTITUENT OF COMETS

In 1958, high-resolution spectroscopy allowed the separation of the terrestrial oxygen lines from the cometary ones and also led to the definitive confirmation of the presence of the isotopic lines of ¹³C, long suspected to be present in and also led to the definitive confirmation of the presence of the terrestrial oxygen lines from the cometary ones (the submillimeter region is also becoming increasingly accessible), and...
taux et al. (1973) showed that the velocity of the H atoms was about 8 km s\(^{-1}\). Following an investigation of the photolysis of water molecules by sunlight, these authors suggested the possibility that the majority of the observed H atoms were the result of the dissociation of OH radicals.

Keller and co-workers reached similar conclusions in a series of independent papers: Keller (1971) discussed the possibility that the observed H atoms in Comet C/1969 Y1 (Bennett) might arise from the direct dissociation of water, ideas that he further developed later (Keller, 1973a,b). However, these investigations, as well as that of Bertaux et al. (1973), were limited by the fact that the parameters governing the water photolysis were not well known at that time. Blamont and Festou (1974) measured both the unknown scale length of OH and the production rate of that radical in Comet C/1973 E1 (Kohoutek). Keller and Lillie (1974) also measured the scale length of OH (in Comet Bennett) and found a value in complete agreement with that found for Comet Kohoutek. An important clue that \( \text{H}_2\text{O} \) was the main source of both the H atoms and the OH radicals came when the velocity of the H atoms was measured directly from Copernicus observations (Drake et al., 1976) and, indirectly, from the analysis of the velocity of H atoms from Ly-\( \alpha \) observations (cf. review by Keller, 1976), and was found to be fully consistent with the water photolysis scheme. For the first time, based on an experimental study of the photolysis of water molecules and simultaneous measurements of the H and O production rates, it was demonstrated that water was the likely parent of most of the H atoms and the OH radicals.

Subsequent systematic observations of OH, H, and \( \text{O} \) emissions in more comets using the International Ultraviolet Explorer (IUE) (Weaver et al., 1981a,b) further strengthened the case for \( \text{H}_2\text{O} \) as the dominant cometary volatile. Beginning with Comet C/1979 Y1 (Bradfield), a long series of high-quality observations of the UV spectra of comets was obtained with the IUE in programs led by A’Hearn, Feldman, and Festou, from which a self-consistent set of OH production rates was derived (e.g., Festou and Feldman, 1987). About 50 comets were observed during the period from 1978 through 1995, and OH production rates were systematically derived for all of them. A comprehensive theory of OH fluorescence, which has been used to interpret the UV observations, was developed by Schleicher and A’Hearn (1982, 1988).

Following the discovery of the 18-cm maser emission of OH (Biraud et al., 1974; Turner, 1974), radio OH emission has been monitored by the Nançay observatory in over 50 comets by Crovisier and collaborators (Crovisier et al., 2002), with observations continuing to the present day with improved sensitivity. These OH monitoring programs established without a doubt the ubiquity of water as the dominant volatile constituent in comets; no comet was found to be deficient in water. In addition, many parameters of the OH radical are derived from 18-cm observations, and information on the kinematics in the coma as well as the determination of the OH production rates are obtained on a regular basis. The methodology for determining OH velocity profiles was worked out by Bockelée-Morvan and Gérard (1984). The detailed mechanism by which comets emit OH photons at radio wavelengths was investigated by Despois et al. (1981).

The radio observations of OH have become the main source of water production rates since 1996. Including both the radio and UV OH observations, water production rates have been derived for about 100 comets. Comets were often followed during a significant fraction of their orbits. The radio and UV determinations of the water production rates do not always agree, as has been discussed by Schloerb (1988, 1989), but these large databases are still extremely useful.

\( \text{H}_2\text{O} \) itself was not definitively detected until its strong IR ro-vibrational emissions were measured by Mumma et al. (1986) in the coma of 1P/Halley during observations from the Kuiper Airborne Observatory, and later from the Vega flyby spacecraft (Combes et al., 1986). The water molecule was also directly detected in 1P/Halley using the neutral mass spectrometer on the Giotto spacecraft (Krankowsky et al., 1986). Non-resonance fluorescence emissions of water at IR wavelengths can now be used rather routinely to monitor water production rates in comets (cf. Dello Russo et al., 2000), but the number of comets observed in this way is still rather small, at least compared to the number whose OH emission has been monitored at radio wavelengths.

There are multiple production pathways for H atoms. Their excitation by the solar Ly-\( \alpha \) line; their interaction with the SW, molecules, and ions; and their kinematics are hard to model. Water production rates can nevertheless be derived from the observation of H lines, as demonstrated in the insightful investigation of the H I UV emission in comets of Richter et al. (2000), which summarizes the most recent work on this topic. One often overlooked conclusion that can be drawn from the many observations of the H comae of comets is that, even with the modeling errors and calibration uncertainties, the production of non-water H-bearing species probably can be no larger than \( \sim 20–30\% \) of the \( \text{H}_2\text{O} \) production rate for most comets observed within 1 AU of the Sun, which leaves \( \text{CO}, \text{CO}_2 \), and a few hydrocarbon molecules, as we shall see below, as the main candidates to supply, after \( \text{H}_2\text{O} \), the bulk of the remaining volatile portion of the cometary nucleus.

### 6. OTHER COMETARY PARENT MOLECULES

In 1970, the known optical emissions were from daughter or granddaughter species (e.g., \( \text{C}_2, \text{C}_3, \text{CN}, \text{CH}, \text{O}, \text{NH}, \text{NH}_2 \)) with uncertain and not very abundant progenitors in the nucleus. Only the O atom was thought to be possibly as abundant as the newly discovered dominant H and OH. For compositional research to develop, technological advances that broadened the wavelength range of the observations were required. The UV window was the first to be explored, followed a few years later by the IR region, and slightly later by the radio region.

Feldman and his collaborators recorded high-quality and high-sensitivity UV spectra of comets during sounding rocket observations of C/1973 E1 (Kohoutek) (Feldman et al.,
1974) and C/1975 V1-A (West) (Feldman and Brune, 1976). The latter observations provided the first detection of CO in a comet and demonstrated that this molecule was one of the most abundant in comets, although we now know that the amount of CO varies greatly from comet to comet. The CO UV emission has been observed in nearly every bright comet since then, first by the IUE and subsequently by the Hubble Space Telescope (HST) and the Far Ultraviolet Spectroscopic Explorer (FUSE). The UV observations also provide access to two other potential parent molecules: the short-lived S\textsubscript{2}, which was discovered during IUE observations of C/1983 H1 (IRAS-Araki-Alcock) (A’Hearn et al., 1982), and CO\textsubscript{2}, which can be indirectly probed via emission in the forbidden CO Cameron band emission (Weaver et al., 1994).

The 1970s witnessed the development of systematic, quantitative observations of optical cometary emissions by means of photoelectric narrow-band filter photometry (by A’Hearn, Schleicher, Millis, and their collaborators) and CCD spectroscopy (by several groups led by Cochran, Newburn, and Fink). A review of the early observations and the observing techniques is given by A’Hearn (1983). In the early 1980s spectrophotometry developed rapidly when fast detectors became available. This method provides both a good separation of band or line emissions and spatial information on the distribution of coma species. In parallel, numerous theoretical studies, aimed at calculating the fluorescence efficiencies of the coma radicals and ions, resulted in the establishment of reliable conversions of observed surface brightnesses into column densities of the different species. The last step in the data analysis process is then the derivation of gas production rates. A systematic survey of the principal optical emissions from 85 comets produced the first evidence for the existence of compositional families among the comets (A’Hearn et al., 1995).

A real breakthrough in the detection of parent species occurred in the mid 1980s with the development of new instrumentation and techniques at IR and radio wavelengths, which could be used to detect ro-vibrational and pure rotational emissions from molecules. Thus, parent species can now be observed directly, leading to more direct information than that obtained from their destruction products. The apparition of two exceptionally bright comets in the mid 1990s, C/1996 B2 (Hyakutake) and C/1995 O1 (Hale-Bopp), together with improvements in instrumentation and the use of the Infrared Space Observatory (ISO), permitted the discovery of more than two dozen parent molecules (Brooke et al., 1996; Mumma et al., 1996; Crovisier et al., 1997; Bockelée-Morvan et al., 2000), among which is the fairly abundant CO\textsubscript{2} molecule, first detected in 1P/Halley both spectroscopically at IR wavelengths (Combes et al., 1986) and with a neutral mass spectrometer (Krankowsky et al., 1986) and suspected to be present in comets for decades because of the well-known CO\textsuperscript{+} bands. Subsequent observations of other comets are leading to interesting compositional intercomparisons (Mumma et al., 2003; Biver et al., 2002), although caution regarding the interpretation of compositional trends is advised because of the small number statistics for most species. Bockelée-Morvan et al. (2004) gives a detailed account of the era that began after the passage of Comet 1P/Halley when ground-based IR spectrometers and radio telescopes were systematically used to investigate cometary composition.

7. COMETARY ORBITS

From the perspective of cometary dynamics, the modern era is defined by the advent of efficient and powerful computers. For the first time, numerical simulations of the orbital evolution of comets over the age of the solar system, including the gravitational influences of close encounters with Jupiter and other planets, stars, and interstellar clouds, have been performed. Computers also revolutionized the work on orbit determination and allowed the linkage of past and recent apparitions of observed comets, as well as the preparation of ephemerides for upcoming apparitions, even for long-lost comets. Whereas Oort had been working on a small sample of comets to build his theory, Marsden et al. (1978) improved the earlier statistics by using 200 well-determined long-period orbits. They found a concentration of inverse semimajor axes corresponding to an average aphelion distance ≤60,000 AU for q > 2 AU, only about half as remote as Oort’s original distance. Besides the apparently abnormal fading of new comets (required because Oort’s peak is too high), a major problem remained: the apparent overabundance of Jupiter-family comets. Everhart (1972) found a possible route of direct transfer from the Oort cloud via jovian perturbations at repeated encounters with the planet, but the efficiency of this transfer was too low to account for the observed number of Jupiter-family comets. An alternative scenario came from orbital integrations of the observed comets by Kazimirchak-Polonskaya (1972): The comets might not be captured by Jupiter alone, but rather by a stepwise process involving all the giant planets. The modern solution is that a disk-like source of comets in the outer reaches of the solar system is required to explain the properties of the Jupiter family of comets (now called “ecliptic comets,” adopting Levison’s 1996 taxonomy of comet orbits). Kazimirchak-Polonskaya’s process naturally explains the existence of the Centaur family.

A major step forward during this period dealt with the modeling of nongravitational effects in cometary motions. Marsden (1969) introduced a nongravitational force into the Newtonian equations of motion with simple expressions for the radial and transverse components in the orbital plane. These involved a function of the heliocentric distance r expressing a standard “force law,” multiplied by a coefficient whose value was determined along with the osculating orbital elements by minimizing the residuals of the fit to positional observations. The radial coefficient was called A\textsubscript{r} and the transverse A\textsubscript{r}. It was recognized that the model might not be physically realistic and that more meaningful parameters might be derived from a more general formalism, but attempts in this direction were unsuccessful (Marsden, 1970). The final update of the model was made in 1973 (Marsden et al., 1973), stimulated by calculations of the H\textsubscript{2}O sublimation rate as a function of r (Delsemme and Miller, 1971).
This was taken as the model force law, expressed as an algebraic function $g(r)$ whose parameters were chosen to fit Delsemme and Miller’s results. Eventually, more realistic models were constructed for the jet force resulting from asymmetric H$_2$O outgassing, including the heat flow in the surface layers of the nucleus (Rickman and Froeschlé, 1983). As a result it was found that the true force law might be very different from the $g(r)$ formula. These efforts led Rickman (1989) and Sagdeev et al. (1988) to the first evaluation of the density of a comet that, not too surprisingly, was quite low and implied a high porosity for the nucleus. The new image of a comet nucleus after the in situ exploration of 1P/Halley, and the intensive efforts of Crifo and collaborators to model the near-nucleus environment, has opened the door to improved models, as described in Yeomans et al. (2004). These efforts will never be completely successful until we finally understand what the expression “activity of comets” really means.

8. THE TRANSNEPTUNIAN BELT AS A COMET RESERVOIR

Around 1950, the Kant-Laplace nebular hypothesis for the origin of the solar system was reconsidered in the light of the chemical compositions of the planets and their variation with heliocentric distance. Edgeworth (1949) and Kuiper (1949, 1951) argued that it is unlikely for the solar nebula to have ended abruptly at the position of Neptune’s orbit, and thus a large population of planet precursors with a generally icy composition had to exist outside the region of the giant planets. Kuiper (1951) claimed that such bodies could be identified with Whipple’s cometary nuclei and suggested that Pluto’s gravitational action (its mass was then thought to be in the 0.11 $M_\oplus$ range) might have scattered the objects into Neptune’s zone of influence, whereupon ejection into the Oort cloud would ensue. Both Kuiper and Edgeworth suggested that the original comet belt just beyond Neptune might still be intact. Fernández (1980) was the first to predict the existence of such a belt in a quantitative way, and he demonstrated that this belt is probably the principal source of the ecliptic comets. This latter idea was later expanded by Duncan and collaborators; Duncan et al. (2004) provides the current perspective on this subject.

The discovery of the transneptunian object 1992 QB1 by Jewitt and Luu in 1992 provided dramatic observational evidence that the basic hypothesis of Kuiper and Edgeworth might be correct. The number of observed “Kuiper belt objects” (KBOs) is now approaching 1000, and the total population likely exceeds $10^5$ members with sizes larger than 100 km. The structure of the belt, its dynamical subclasses, and its evolution are currently the subject of intense research activity by the community.

We also note that dynamical investigations produce new results that seem to contradict the “conventional wisdom” every decade or so, with some evidence that the pace of change is accelerating as computers continue to get faster and new analysis techniques are employed. Thus, we should not be surprised to find in five years that the ideas presented in Comets II on the origin and evolution of the Kuiper belt and Oort cloud have changed significantly.

9. FINAL REMARKS

During the last few decades, theories based on the original ideas of Kant and Laplace have obtained the status of the “standard” theory for the formation of the solar system. Numerical calculations and a wealth of new information on the structure and composition of planetary bodies, as well as data gathered from astrophysical studies of circumstellar disks and extra-solar-system planets, seem to leave little room for alternate theories.

Data from comets have strengthened our understanding of the formation and evolution of the solar nebula and have helped us to see the intimate connections between our solar system, the interstellar medium, and extra-solar-system planets. However, these connections are not always clear and easy to recognize and many mysteries still remain. In particular, our knowledge of the composition and physical structure of cometary nuclei remain rather primitive in many respects, and we can expect surprising discoveries with ever-increasingly sophisticated observations. The imminent return of a sample of cometary matter from 81P/Wild 2 is especially anticipated, as those results will almost certainly guide our future exploration of comets. On the other hand, the Stardust results will likely leave many questions about the icy composition of cometary nuclei, which should be better addressed by the Rosetta mission, due to rendezvous with Comet 61P/Churyumov-Gerasimenko in 2015.

Nevertheless, we offer Comets II as the best compendium of cometary knowledge for the next decade. With the help of our colleagues in the cometary community, we have attempted to put together a volume that presents a logical and comprehensive treatment of cometary science. The structure of the book was laid out with the clear objective of covering most areas of cometary science through a set of contiguous, non-overlapping chapters. In a sense, it is a book written by a team of about 100 collaborative authors. We have included at the beginning of this volume a series of chapters that describe what may happen to interstellar materials that are used to make up a planetary system. Obviously, the path between interstellar molecules/grains and comets is quite long and complicated. It is this long journey and what happens along the way that is described in this book. Enjoy!

REFERENCES


Institute for Physics, Hungarian Academy of Sciences.


Oort cloud formation and dynamics. In Comets II (M. C. Festou et al., eds.), this volume. Univ. of Arizona, Tucson.
Huggins W. (1868) Further observations on the spectra of some of the stars and nebulae, with an attempt to determine therefrom whether these bodies are moving towards or from the Earth, also observations on the spectra of the Sun and of Comet II, 1868. Proc. R. Soc., 15, 529–564.
Levin B. J. (1943) Gas evolution from the nucleus of a comet as related to the variations in its absolute brightness. Comp. Rend. (Dokl.) Acad. Sci. URSS, 38, 72–74.
Marsden B. G., Sekanina Z., and Everhart E. (1978) New oscu-


Comets II


