Our cometary environment

1997 Rep. Prog. Phys. 60 293

(http://iopscience.iop.org/0034-4885/60/3/001)

Rep. Prog. Phys. **60** (1997) 293–343. Printed in the UK PII: S0034-4885(97)07654-9

Our cometary environment

W M Napier† and S V M Clube‡

† Armagh Observatory, College Hill, Armagh BT61 9DG, UK

‡ Department of Nuclear and Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, UK

Received 8 July 1996

Abstract

The encounter of a small armada of spacecraft with Halley's Comet in 1986, the disintegration and multiple impact of Comet Shoemaker–Levy 9 on Jupiter in 1994, and the application of new technologies to the detection of distant solar system bodies, have led to great revisions in the understanding of comets. Further, rapid improvements in computing power and numerical techniques have permitted the dynamical evolution of comets and asteroids to be followed far into the future and past, and the relationships between families of small interplanetary bodies to be explored. The small body environment is now generally recognized as strongly interacting with the terrestrial one, and may be hazardous on timescales of human as well as geological interest. We review our current understanding of the cometary environment, with particular regard to the hazard it presents. It appears that many comets are handed down from the Oort–Öpik cloud, which is dynamically sensitive to the galactic environment, through the planetary system into Earth-crossing orbits. Thus, the terrestrial environment is subject to stresses which vary cyclically on a number of timescales from planetary to galactic.

0034-4885/97/030293+51$59.50

Contents

1. Near-Earth cometary incursions

The appearance of a comet in the sky is usually unexpected; it irrupts into the orderly progression of the heavens; it is a portent of dramatic change. The portentous nature of comets has been known from the very earliest times but our attitude to comets in modern times has become more circumspect, even to the extent that 'portents' may be considered unscientific: it is important to understand the reason for this change and to keep the issue at stake in perspective. Thus, the portentous nature of comets can still be real enough and open to scientific study, the point here being that mankind is upset as much by the expectation of cosmic catastrophe as it is by catastrophe itself. Both are more common when the orbital commensurability of an active or defunct comet of the very shortest orbital period (a few years, say) is such as to bring the Earth repeatedly into contact with a particular source body's 'trail' of more significant meteoroidal debris, sharply increasing the fireball flux to Earth (e.g. Clube 1995). Such periods of repeated contact last for some 50 yr and arise every other century or so. Indeed, it is 'blazing stars' more than comets which presage the possibility of dramatic change on Earth and it is only during the (scientific) early modern and modern periods of history, essentially since the time of Newton and Halley, that both the *line* of a cometary orbit as well as a comet's *presence* have been recognized as important so far as the portentous behaviour of comets is concerned. It is reasonable in fact that direct cometary impacts are no longer greatly feared since the intervals between such events on Earth are $\sim 10^7$ yr. However the intervals between groups of commensurable encounters with cometary debris trails are more like $\sim 10^2$–10^3 yr and the incidence of fear on such occasions is by no means eliminated. It follows that it is also reasonable that scientific consideration be given to the cohesive strength of the larger meteoroidal debris of comets ($\sim$50–500 m, say) so as to predict their (dramatic) physical effects, e.g. low-level multi-megaton airbursts, high-level reflective dust veils and (cosmic) chemical loading of the atmosphere. The study of these derivative phenomena based on comets is currently still in its infancy but can be expected to grow in future as its significance for mankind and civilization is more widely appreciated. The subject will be touched on from time to time during the course of this article but for the most part we are concerned with the observed behaviour of the primary bodies alone: comets.

Crossing the sky, moving against the starry backdrop over a period from days to months, a comet will develop one or more tails as ices sublimate from its nucleus, releasing trapped dust which is then accelerated away by solar radiation pressure. Most comets are faint and have sometimes been described as wraith-like or ghostly in appearance. Rare, giant comets may develop smoky, dark-red tails which may straddle the sky, seeming to split it in two, and within which are fine striations which change from night to night. Such objects may sometimes even be visible by day, close to the Sun: the Biblical 'pillar of cloud by day, pillar of fire by night' which guided the Israelites might have been one such. Multiple tails may develop as a trail of large fragments is formed and the dust is accelerated away by solar radiation pressure, whereas ions in the coma of gas around the nucleus are swept away by the solar wind. Dust tails may be broad and curved; ion tails are straight, anti-solar in direction, and may attain lengths of ~ 1 au (an astronomical unit, 1.5×10^8 km, is close to the mean Earth–Sun distance).

By 1995, almost 1500 cometary apparitions had been catalogued. About half of these

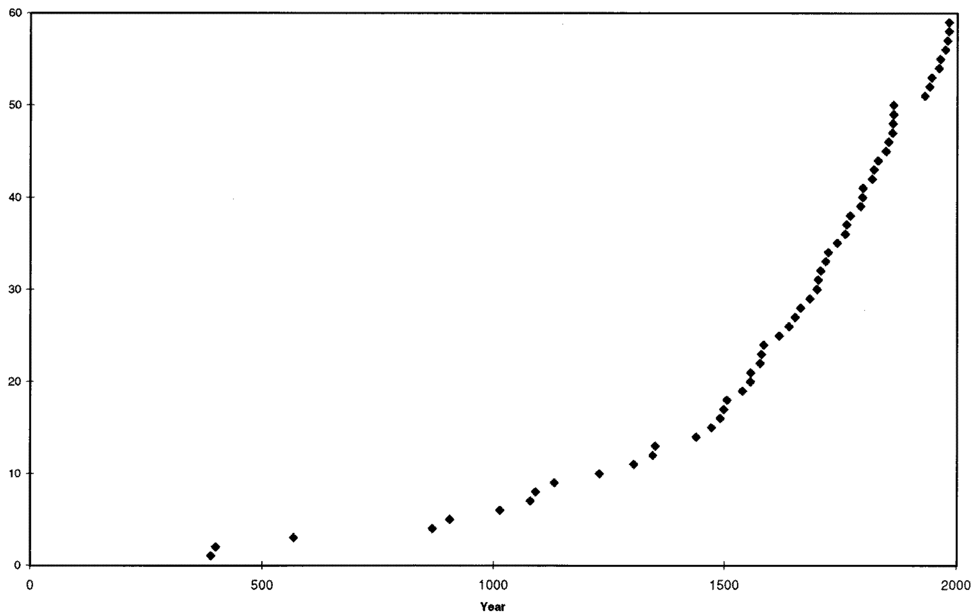

Figure 1. Cumulative comet discoveries within 0.2 au of the Earth for the past millennium. The rate has been steady since about 1700. Courtesy S Manley.

were single passes: the comets have never returned nor for the most part are expected to, for some millions of years. The other half are multiple apparitions, due to the repeated returns of about 180 known periodic comets.

The discovery probability of long-period comets ($P > 200$ yr) is strongly influenced by observational selection effects, including accidents of geometry as well as intrinsic magnitude. Human awareness of the night sky is likely to be an important factor, but is difficult to quantify. Figure 1 shows the observed close approaches to Earth of long-period comets from mediaeval times: it can be seen that there is a steep secular increase, reflecting probable completeness of discovery of naked-eye objects within 0.2 au of the Earth. However there are strange gaps (e.g. the dearth of discoveries in the first half of the 20th century); also it is not clear whether the telescope made any difference to the discovery rate for about two hundred years after its invention. Rather an apparent mid-18th century surge in comet discoveries (Everhart 1967) is the era of the famous comet hunters: Caroline and William Herschel, Pons, Messier and others. As such it almost certainly marks a new-found enthusiasm, but such might cause us to reflect whether it was prompted by phenomena correlated with the recorded increase in the fireball flux during the latter half of the 18th century. Historically such increases appear to be associated with objects in the broad meteoroid stream known as the Taurids and it is curious that this epoch coincides with the first sighting of the comet with the shortest known period—Comet P/Encke—which was then found to be closely associated with this stream (Whipple and Hamid 1952, section 5.4) as well as, probably, the Tunguska meteoroid of 1908 (Kresak 1978a). The question arises perhaps whether unexpected impacts will reactivate defunct comets in the inner solar system from time to time and distribute meteoroidal debris somewhat more widely than the concentrated orbital trails (Asher and Clube 1993).

However this may be, several attempts have been made to estimate the true flux of

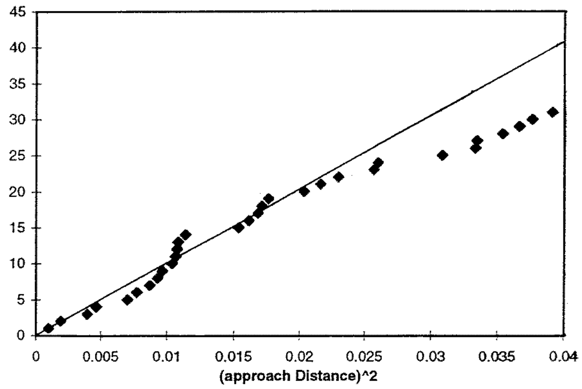

Figure 2. Cumulative radial distribution of all Earth-approaching comets. Courtesy S Manley.

comets into the planetary system or the near-Earth environment. The simplest of these (Fernández and Ip 1991) is based on the fact that the discovery rate of comets approaching to within 1 au of the Sun has been constant since about 1788. Objects passing this close are supposed to be bright enough to be discovered, and allowance need only be made for unfavourable geometry. It is inferred that $\sim$3 comets pass within 1 au of the Sun each year. Everhart (1967) used 256 comets discovered over the period 1840–1967, and a more sophisticated approach involving both magnitude and perihelion (q) distribution functions. He derived a flux of 11.6 Earth-crossers per annum within 1 au of the Sun. This, however, involved an extrapolation of the brightness distribution towards fainter, unseen comets. A recent approach due to Manley (personal communication), following Sekanina and Yeomans (1984), involves examining the radial distribution of those long-period comets which have approached close to the Earth (within 0.2 au) over the last 300 yr. Since these are apparently brighter, they should represent a more completely sampled distribution. Even these, however, show incompleteness of discovery (figure 2). Manley finds that there may have been $\sim$40 close approaches over the last 300 yr, corresponding to a long-period flux of 3.0 comets au^{-1} yr^{-1}. There may be a real dearth of faint comets, corresponding to a diameter cut-off at 0.3–0.5 km, in both long-period and short-period systems; but uncertainties in the discovery geometry of very faint comets make any such claim unsafe at present.

If attention is paid only to comets brighter than 7th magnitude, then various authors converge on predicted Earth-crossing fluxes in the range 0.8–0.9 au^{-1} yr^{-1}. Fernández (1982), using Everhart's basic approach but restricting himself to dynamically new comets of absolute magnitude 7 or brighter, inferred that 1.5 such pass within the orbit of Jupiter every year.

2. The system of long-period comets

2.1. The Oort cloud

In 1705 Edmund Halley, determining the orbits of 24 comets to be parabolic to within the uncertainties, argued that nevertheless they would turn out to be elliptical; and he wrote that 'For so their Number will be determinate and, perhaps, not so very great. Besides, the Space between the Sun and the fix'd stars is so immense that there is Room enough for a Comet to revolve, tho' the Period of its Revolution be vastly long.' Halley was thus the first astronomer to propose that a swarm of long-period comets existed around the Sun.

In 1783, William Herschel discovered that the Sun is moving with respect to the nearby stars, a result which he confirmed in 1805. Around 1860, Schiaparelli deduced that because there was no sign of this solar drift in the orbits of long-period comets, then the Sun must be virtually at rest with regard to them. He thus, independently of Halley, proposed that the Sun is surrounded by an extended cloud of comets. The implication was that comets are part of the solar system, a view which ran counter to Laplace's hypothesis that they arrive from interstellar space. In spite of the obvious logic of Schiaparelli's argument, however, Laplace's interstellar hypothesis for the provenance of comets remained dominant in astronomical thinking, at least until the end of the 19th century.

In 1932, Öpik considered the effect which passing stars would have on comets, and concluded that they would be impulsively perturbed. He too deduced that the Sun must be surrounded by a cloud, made up of comets whose aphelia had been increased by such perturbations, taking comets out of the realm of planetary influence. However, he did not particularly note that stellar perturbations would generate a two-way traffic, causing some orbits to diffuse back into the planetary system; and so he apparently failed to identify the hypothetical cloud as the source of the observed long-period comets. This final step is generally considered to have been taken by Oort in 1950.

Oort proposed that the planetary system is surrounded by a vast spherical cloud of comets extending more than half way to the nearest star. He thus enormously extended the boundary of the solar system, from $\sim$30 au to $\sim$200 000 au. The hypothesis was a bold extrapolation: based on the orbits of only 19 long-period comets with accurately known elements, he inferred a cloud population of 2×10^{11} comets. The 19 orbits were the 'original' ones, corrected for the disturbing effects of the planets and so describing the motion of the comets before they entered the planetary system. Oort found that the original inverse semi-major axes $1/a$ of these orbits were sharply peaked close to zero energy, over half being in the range $0 \lesssim 1/a \lesssim 50 \times 10^{-6}$ au^{-1}. Since the specific orbital energy $E = -GM_\odot/a$, G the gravitational constant and $M_\odot$ the mass of the Sun, the implication is that the distant comet cloud is only just gravitationally bound to the Sun.

Even a single passage of a comet through the planetary system may disperse the energy by $\sigma_E \sim 1000 \times 10^{-6}$ au^{-1}, whereas over half the observed comets had $1/a$ values less than 50×10^{-6} au^{-1}. Oort therefore inferred that these comets were dynamically 'new', that is, they had never before entered the planetary system. He proposed that stellar passages through the comet cloud, occurring every few million years, perturb comets randomly, throwing a small fraction of them towards the planetary system. An order of magnitude increase in the sample since then has confirmed the existence of this sharp peak in the energy distribution. There is in addition a relatively smaller population on more tightly bound orbits ($50 \lesssim 1/a \times 10^6 \lesssim 100$ au^{-1}). A few marginally hyperbolic comets exist but these appear not to be truly interstellar visitors: they are most likely to be either recent ejecta, due to excessive planetary perturbations or non-gravitational acceleration arising from asymmetric surface activity. Some 'hyperbolic' comets may in fact be in truly bound orbits which are not known with sufficient accuracy.

Oort considered that the 2×10^{11} comets made up a spherical cloud of inner and outer radii 40 000 and 200 000 au respectively. At any one time half the comets of the cloud were supposed to lie beyond 70 000 au from the Sun, at which distance the escape velocity is only 0.16 km s^{-1} and a steady attrition of comets from the cloud was expected because of ejection through close encounters with stars. Additionally, comets thrown into the planetary system were lost through splitting, sublimation of volatiles, hyperbolic ejection or collisions with the planets.

The Oort (or Oort–Öpik) cloud concept has survived the substantial increase in the long-period comet dataset. Its estimated outer radius is now taken to be ∼40 000–50 000 au, corresponding to a system of long-period comets with semi-major axes in the range 20 000–25 000 au. The cloud remains a hypothetical construct, however, and it is still suggested from time to time (e.g. Van Flandern 1993) that there could be as few as $\sim 10^7$ comets on orbits of very high eccentricity, generated in a recent planetary explosion or capture event. In the section that follows we discuss the structure of the standard (Oort) model and its later modifications.

2.2. Its detailed structure

Oort assumed that at any distance r from the Sun, the velocity distribution of the comets was isotropic and uniform out to a maximum corresponding to the free-fall speed of a comet dropped from the outer edge R_o of the cloud to r. Under this assumption, the number density ν_r of comets is then uniquely determined through the equation of hydrostatic equilibrium to be

$$\nu_r = A\left(\frac{R_o}{r} - 1\right)^{3/2} \tag{2.1}$$

where the normalizing constant $A \sim 10^{-5}$ au^{-3} is adjusted so that the infall flux of long-period comets expected from (2.1) matches that observed (Bailey 1983). Whether the velocity distribution is in fact uniform depends on the origin and evolution of the Oort cloud, however. Cosmogonies which involve formation in a protoplanetary nebula, or a collapsing solar nebula, would generally involve the initial formation of a dense inner core of comets. Thus, in recent years attention has been given to more centrally concentrated Oort cloud models, with energy distribution

$$\nu_E \propto |E|^{-\gamma} \tag{2.2}$$

where the classical Oort cloud has an energy distribution given by $\gamma = 2.5$. Table 1 lists, for various γ, the semi-major axis and radius distribution, the ν_r being approximately valid only when $r \ll R_o$. Note that in Oort's original model, most of the comets in the cloud lie in the regions furthest from the Sun.

The index γ is fundamental to understanding the origin and evolution of the cometary system. Unfortunately it is as yet a somewhat unconstrained parameter. Models with strongly negative γ, the so-called 'dense inner cloud' models, have been proposed either to replenish comet losses from the long-period (outer Oort cloud) system, or as a source of short-period comets. Likewise, depending on the assumed cosmogony, the postulated geometries vary from compact, disk-like configurations (radius $\lesssim$ 200 au) to relatively extended spheres ($r \sim 10^3$–10^4 au). The distribution $\gamma = 0$ corresponds to an isothermal cloud and might be appropriate to a system whose structure has been determined by relatively gentle, diffusive processes such as the stellar perturbations envisaged by Oort.

Table 1. Semi-major axis a, specific energy E and number density r distributions of comets for various assumed concentration parameters γ. The r-distribution is only approximate.

γ	$\frac{5}{2}$	0	-1	-2
ν_a	$a^{1/2}$	a^{-2}	a^{-3}	a^{-4}
ν_E	$\|E\|^{-5/2}$	$\|E\|^0$	$\|E\|$	$\|E\|^2$
ν_r	$r^{-3/2}$	r^{-4}	r^{-5}	r^{-6}

A dense inner cloud ($\gamma \lesssim -2$) on the other hand, having an abundance of comets with semi-major axes in the range 5000–10 000 au, would (if such exists) be relatively impervious to perturbations by galactic tides or passing massive nebulae. However, a constraint on the characteristics of any such core is set by the need that it should not predict too many short- or intermediate-period comets (Stagg and Bailey 1989). To avoid an overabundance, the core would require to have an inner edge corresponding to semi-major axes (4000, 5000, 6000) au for $\gamma = (0.5, 0.0, -1.0)$ respectively (Bailey and Stagg 1990). A further constraint on the density of a massive inner cloud may be set by the cratering record. Any such core would be penetrated from time to time by stars, each passage yielding an intense flood of comets into the inner planetary system (Hills 1981, Fernández 1992) to be recorded as a cratering record 'spike'.

A comet perturbed by a nearby passing star will suffer a velocity change $\delta v = 2Gm/DV$, D being the impact parameter and (m, V) respectively the mass and velocity of the star. Combining the Sun's motion relative to the local standard of rest with the characteristic velocity of stars, a typical value for V is 30 km s^{-1}. The closest approach distance of a comet to the Sun (the perihelion distance q) is given by $q = a(1 - e)$. For fixed a, and for a system of comets in statistical equilibrium, the proportion of comets with eccentricity $> e$ is given by $F_e = 1 - e^2 \sim 2(1 - e)$ for high eccentricity. Thus, the proportion which approach the Sun to within $\leqslant q_L$ au is $F_L = 2q_L/a$ when $q \ll a$. Hence there is a 'loss cone' of angular radius $2F_L^{1/2}$ towards the Sun such that comets thrown into it by a passing star, when $q_L \lesssim 10$ au or so, become strongly perturbed by the giant planets and are thereby lost from the Oort cloud, some eventually to become collision hazards. A penetrating star is expected to disturb comets in a surrounding tube of (variable) radius D_L along its trajectory, and some proportion of these will be thrown towards the planetary system. The flux of comets arriving in the form of such a shower or mini-shower is given by (Fernández 1992)

$$\dot{n_c} = 2\pi r^3 n(r) F_L/P \int_0^\infty D_L^2 (h^2 + D_\odot^2)^{-3/2}\, \mathrm{d}h. \tag{2.3}$$

Here h represents the perturber's distance from the point of closest approach to the Sun: $\boldsymbol{r} = \boldsymbol{D}_\odot$; $n(r)$ is the number density distribution of comets; and $P \propto a^{3/2}$ is the orbital period corresponding to the most intense phase of the comet shower.

Passage of a star within D au of the Sun will occur at mean intervals of $\sim 30/D^2$ Myr, where D is measured in units of 10^4 au. Thus, an inner cloud of radius 10^4 au will be penetrated at mean intervals of $\sim$30 Myr (which is also about the interval between penetrations by the solar system of molecular clouds of order $10^4 M_\odot$). For a cloud with central condensation $\gamma = -2$, equation (2.3) can be integrated to show that such a passage would create a comet shower about 100 times the background intensity (Fernández 1992). A passage within 3000 au would yield a shower perhaps 100 times more intense again. Such a penetration is expected within $\sim$100 Myr over which the record of large terrestrial impact craters is not too incomplete. The duration of such showers would be of order a few Myr corresponding to two or three orbital periods of the disturbed comets, and such events should manifest themselves in the impact cratering record as a set of co-eval craters.

There are currently $\sim$30 craters with ages $\lesssim 250$ Myr, dated to within 5 Myr. A shower of peak intensity 10 000 times background would yield essentially all these impact craters at one age; one at 10 times background would show up as a set of $\sim$15 co-eval craters; in fact no concentrated cratering peaks are securely observed, although there does appear to be a long-term cycle (section 6.2). Allowing for the uncertain sporadic background of impacts, it seems that the *dense* inner cloud concept is difficult to sustain given the lack of strong peaks in the crater age distribution (Napier 1987, Stothers 1988, Bailey and Stagg 1988) unless its energy distribution index γ (for a particular minimum radius $< 10\,000$ au say) is $\gtrsim -2$. It is nevertheless possible to have an inner population of say 5×10^{12} comets, with an inner edge of semi-major axis $\sim$ 5000 au, distributed as $\gamma \sim 0$. Whether this is sufficient to replenish the outer Oort cloud is at present an open question.

It is possible that dark matter might exist in the galactic disk in the form of lower mass objects ('Jupiters' or very compact nebulae, say) rather than stars. If so, these objects would penetrate a dense inner core more frequently than ordinary stars (Oort 1950); being more numerous, they would sample the velocity distribution more completely, in which case the proportionally larger effects of very low velocity penetrations ($V \lesssim 3$ km s^{-1}, say) might also have to be considered (Clube and Napier 1996). This is discussed more fully in section 6.

2.3. Galactic tides and dark matter

The Sun orbits within the galactic disk, whose half-thickness is $\sim$65 pc for the flattest, Population I component. Objects orbiting about the Galaxy which are separated in space by a small distance experience slightly different galactic forces, this differential effect being the galactic tide. Now the strength of the vertical galactic tide exceeds that of its horizontal component by an order of magnitude and so, to a good approximation, it exerts a restoring force per unit mass on an Oort cloud comet, relative to the Sun, given by

$$f(z) = -4\pi G\rho z \tag{2.4}$$

where ρ represents the local ambient density of the galactic disk at the height of the Sun, and z is the difference in vertical height between Sun and comet. This may be written as

$$f(z) = -C^2 z$$

where the Oort constant $C \sim 0.1$ km s^{-1} pc^{-1}.

The orbital motion in the presence of this restoring force may be found by classical procedures (e.g. Byl 1986, Heisler and Tremaine 1986). A constant of motion exists, given by

$$E = K^2 + 5(1 - K^2)\cos^2\psi$$

where $K^2 = (1 - e^2)$ and $\cos\psi = \cos i \cos\omega$. If $\cos^2\psi > \frac{1}{5}$ ($\psi < 27°$) the galactic argument of perihelion ω precesses; otherwise it librates. Near the stationary solution, the circuit time is $\sim$0.3 Gyr for a long-period comet in the presence of a galactic disk density $0.185 M_\odot$ pc^{-3}. Associated with each of these two families of trajectory, cyclic solutions exist for the perihelion distance of the comet and its galactic latitude. The first-order perturbation theory yields a maximum change δq in perihelion per revolution

$$\delta q \sim 50 a^{7/2} q^{1/2} \text{ au} \tag{2.5}$$

where a is measured in units of 20 000 au and q in units of 100 au. The perihelion distance may thus begin at a very small value, and in the course of a precessional cycle increase to a value comparable with the semi-major axis before declining again.

A comet with perihelion just beyond the edge of the planetary system could therefore, at its subsequent apparition, be thrown into the planetary system. However, such a comet, entering the solar system with perihelion $q \lesssim 10$ au, would rapidly be ejected through gravitational interaction with the giant planets. Depending on the balance between influx and ejection, there will exist a 'loss cone' more or less depleted of comets. From equation (2.5) it can be seen that the tide is relatively much more effective in perturbing the perihelia of the longest-period comets (large a); beyond $a \sim 25\,000$ au, the loss cone is permanently filled; however, the influence of the tide falls off steeply with declining a and is certainly negligible within 5000 au (and probably within 10 000 au) of the Sun. Comets with semi-major axes $\lesssim 20\,000$ au suffer relative perturbations in perihelion $\delta q/q \lesssim 50\%$; such a comet, with perihelion at 20 au at one revolution, say, could be thrown inwards no further than 10 au at its next. Thrown into the region of the giant planets, it would be strongly perturbed and either further thrown into the inner planetary system or hyperbolically ejected. Thus, Jupiter and Saturn impose a barrier, a minimum δq over which comets with $a \lesssim 20\,000$ au cannot leap directly to become observable from the Earth: the remarkable sharpness of the $1/a$ distribution of dynamically new comets is probably therefore a consequence of the galactic tide. In effect the Earth is shielded from the effects of an inner core; if the latter exists, then the sharpness is a consequence of impulsive perturbers and the planets Uranus and Neptune must be subject to extremely intense cometary bombardment.

Direct evidence for the influence of the tide is seen in the latitude distribution of arriving long-period comets. The precessional cycle is correlated with that of the galactic latitude b of the comet:

$$\delta q \sim 2.4a/(5\sin^2 b - 1)\delta b. \tag{2.6}$$

The analysis shows that, allowing for the $\cos b$ dependence of numbers per steradian, the flux of comets should vary as $\sin 2b \cos b$, which peaks at 35°. Studies of the latitude dependence of $\sim$150 long-period comets by Lüst (1984) and Delsemme and Patmiou (1986) revealed clearly depleted zones around the galactic poles and equator, as predicted. The galactic tide is thus seen to be of major importance in filling the loss cone.

Nevertheless, other factors are also at play. The non-zero cometary flux at the galactic poles and equator indicates that close or penetrating encounters with the Oort cloud by stellar or dark matter perturbers are also effective as a whole in filling the loss cone with long-period comets. This too is an important finding, for while the galactic tide is certainly expected to be dominant in the more extended region of the Oort cloud ($a \gtrsim 30\,000$ au), the comet number density at smaller a must also be sufficiently high to ensure a significant observed flux due to impulsive perturbers. *Both* types of perturbation, in other words, apparently play their part in producing the long-period cometary flux (Yabushita 1988).

There are, however, indications of some disequilibrium in the long-period comet flux. Yabushita (1989) has argued that the current influx of long period comets exceeds their loss rate, and so argues for a recently disturbed Oort cloud. In addition there is a large, elongated clustering of aphelia around the antapex of the solar motion (Delsemme and Patmiou 1986, see also figure 3). The signature of the galactic tide would re-assert itself within a few orbital periods of any large disturbance; thus these anomalies, unless due to a continuously applied stress, may also indicate that the cloud was disturbed only a few million years ago. A test is to look for a disturbing mass a few Myr back along the track of the solar motion, the antapex of which lies in the Canis Major–Puppis region. In fact proper motion studies of the Scorpio–Centaurus association (Bertiau 1956) reveal that its convergent point is comparatively close on the celestial sphere to the current solar antapex (figure 3). The association appears to have begun expanding from minimum dimensions of 45×15 pc about 4.5 Myr BP (Blaauw 1991), although earlier epochs of star formation at 7, 11 and 13 Myr have been identified in various subgroups within it. The Sun appears to have passed close to (or penetrated) this star-forming region a few Myr BP. At the time of passage, the association presumably still contained a substantial mass of gas (now seen as surrounding shells of HI). The three positional coincidences (antapex of solar motion, aphelion clustering and Sco–Cen convergent point) may then suggest that the long-period

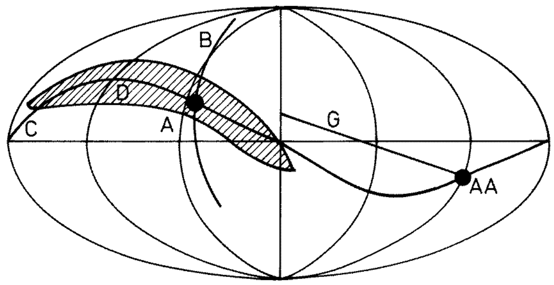

Figure 3. The celestial sphere in galactic coordinates. The shaded region shows the solar antapex clustering. (A,AA) represent the solar apex and antapex respectively. Gould's Belt (a ring of young stars and gas seen edge-on) is represented by G. B is a group of long-period comets with similar orbital elements.

comet system was strongly disturbed during the encounter. The Sco–Cen association is part of the Gould belt system which, should it be connected with other young star systems in the wider solar neighbourhood (say within $\sim$300 pc), may correspond to an expanding ring of gas, dust and young stars of recent origin. According to Olano and Pöppel (1987) the mass of this ring exceeds $10^6 M_\odot$ and was encountered by the Sun 10 ± 2 Myr BP; in which case the encounter, if close, could have deflected the solar trajectory by up to 30°.

2.4. The lifetime of the Oort cloud

For about 30 yr after Oort (1950) had proposed the existence of the comet cloud which bears his name, it was generally held that it had remained *in situ* from the time of its supposed formation and ejection from the planetary regions $\sim$4.5 Gyr ago. Thus, the Oort cloud was held to be a dynamically stable system; the gentle 'gardening' of the cloud by penetrating stars led to a modest evaporation of its comets into interstellar space, and a steady trickle of long period comets into the inner planetary system, where they outgassed and became observable.

However, throughout the 1970s, evidence was emerging that the interstellar medium contains perturbers much more powerful than stars. Sub-millimetre observations were revealing the existence of massive, cold nebulae in considerable numbers ($\sim$4000) throughout the galactic disk. These giant molecular clouds are the most massive entities in the Galaxy. Biermann (1978) proposed that the effect of close encounters with such nebulae would be to strip away the outer regions of the Oort cloud, reducing its radius from the $\sim$200 000 au expected if only stellar perturbations were at work, to the observed $\sim$50 000 au. However, others (Napier and Clube 1979, Clube and Napier 1982) argued for even more severe consequences: numerical simulations (Napier and Staniucha 1982) revealed that, in this new galactic environment, the classical Oort cloud could not have survived for 4.5 Gyr and must be replenished from elsewhere.

The derived mass distribution of molecular clouds in the solar neighbourhood is (Scoville and Sanders 1986)

$$N(M) \propto M^{-1.6} \tag{2.7}$$

with a size spectrum

$$N(R) \propto R^{-2.5}. \tag{2.8}$$

Thus, $\sim$85% of the mass of the system is concentrated in giant molecular clouds of characteristic radius $R \sim 20$ pc and mass $M \sim 3$–$5 \times 10^5 M_\odot$. These giant clouds show considerable sub-structure, generally comprising assemblies of lesser clouds, or clumps (e.g. Blitz and Shu 1980). An archetypal clump is the Mon OB1 cloud surrounding the young star cluster NGC 2264 and associated with the Rosette molecular complex: it has a mass $\sim 20\,000 M_\odot$ and radius 5 pc, and itself comprises about half a dozen fragments of radius 0.5–1 pc, half the total mass being in one fragment and the others having masses in the range 50–3000$M_\odot$.

It is readily shown that passage of the Sun through such a nebular complex would lead to a severe buffeting of the Oort cloud. A grazing encounter at speed V km s^{-1} with a body of mass M solar masses and radius R pc yields an impulsive velocity perturbation $\Delta V = 2GM/RV$ on the motion of the Sun. The *differential* perturbation of a comet relative to the Sun is found to be

$$\delta v = 0.124 \frac{(M/300\,000)(a/50\,000)}{(R/20)(V/20)} \text{ km s}^{-1} \tag{2.9}$$

where (M, R, V) are respectively in units of ($M_\odot$, pc, km s^{-1}) and a is in au. Thus, a single grazing encounter with a Rosette-type molecular cloud will impart a perturbation of order 0.1–0.15 km s^{-1} to a long-period comet, comparable with its escape velocity. However, it is expected that, in the course of its history, the solar system has penetrated giant molecular clouds perhaps 5–15 times (Bailey 1983), and it is clear that the classical cloud would by now have been stripped away by such encounters. A corollary of this fact is an Oort cloud which is itself likely to continually inflate and then rather suddenly disappear on timecales $\sim 10^8$–10^9 yr. Indeed *if* the replenishing source is predominantly a dense inner core ($a \lesssim 10\,000$ au) and the Oort cloud region most sensitive to the galactic tide is re-introduced least efficiently, then the corresponding galactic signatures in the perturbed cometary flux may be absent or present (as now) on such timescales as well.

The expected radius of the current cloud may be estimated by equating the change ΔE of a comet orbit to its specific binding energy E. In general

$$\Delta E = \tfrac{1}{2}[(\boldsymbol{v} + \boldsymbol{\Delta v})^2 - \boldsymbol{v}^2] = \boldsymbol{v} \cdot \boldsymbol{\Delta v} + \tfrac{1}{2}(\boldsymbol{\Delta v})^2. \tag{2.10}$$

The first term produces a random walk and will result in $\Delta E \sim |E|$ after N_o encounters such that $(\boldsymbol{v} \cdot \boldsymbol{\Delta v})\sqrt{N_0} \sim |E|$. The second term is an outward drift yielding $\Delta E \sim |E|$ after N_1 encounters such that $\frac{1}{2}\boldsymbol{\Delta v}^2 N_1 \sim |E|$. It is readily shown that $N_o \sim N_1$. Assuming N encounters at 5 pc with substructures of mass $20\,000 M_\odot$, and applying the escape criterion, the radius of the primordial cloud after N encounters is found to be

$$r = \frac{43\,000}{N^{1/3}} \text{ au}. \tag{2.11}$$

With say eight past encounters in solar system history, the long-period comets still arriving in the planetary system would have semi-major axes $a \lesssim 10\,000$ au, less than half than observed. The $N^{-1/3}$ dependence implies that the initial depletion of the primordial cloud was very rapid, and that the result is rather insensitive to the assumed encounter rate.

More sophisticated analyses have yielded very similar results. Bailey (1986, 1991) for example, finds that the half-life of the Oort cloud against stellar disruption is

$$t_{1/2} \sim 6 \pm 2a_4^{-1} \text{ Gyr} \tag{2.12}$$

and, for clouds,

$$t_{1/2} \sim 15 \pm 0.3a_4^{-3} \text{ Gyr} \tag{2.13}$$

where $a_4 \sim a/10^4$ au. Together, these imply that the critical semi-major axis for which the half-life equals the age of the solar system is only $\sim 10\,000$ au.

Empirical evidence that the classical Oort cloud would be disrupted in the galactic environment comes from the dissociation of binary stars with time in the solar neighbourhood. Poveda and Allen (1975) used a number of age criteria to classify wide binaries as young or old relative to the Sun, and further divided them into systems with separations less than 1000 au, and the rest. They found that amongst the close binaries there were two old binaries for every young one, whereas for wide binaries the ratio was unity. Thus, for stars older than the Sun, dissociation of binaries was taking place down to 1000 au. Encounters with individual stars cannot explain such rapid disruption. More recently, Poveda (1988) has repeated the analysis with 106 binaries within 12.5 pc of the Sun. He found the effects of dissociation to be quite noticeable at 2000 au separation, and that in his sample there were no older binaries of separation $\gtrsim$ 4000 au. The sample is biassed in favour of older stars, whence Poveda concludes that the radius of any surviving primordial remnant cloud cannot be more than 2000–4000 au. This is in good agreement with the result of numerical simulations in which the important perturbers are taken to be the substructures ($M \sim 2 \times 10^4 M_\odot$, $R \sim 5$ pc) within giant molecular clouds. Numerous other studies, referenced in Bailey *et al* (1990), have confirmed that the Oort cloud is indeed dynamically unstable in the galactic environment: it must be replenished from elsewhere, either from an inner core or from the interstellar environment.

3. The systems of short-period comets

3.1. The Jupiter and Halley families

Somewhat arbitrarily, short-period comets are defined as those with orbital periods < 200 yr, corresponding to semi-major axes $a \leqslant 34$ au. In general comets grow tails, losing dust and volatiles, when their orbits acquire perihelia $q \lesssim 2.5$ au, and many are associated with meteor streams observed when the Earth intersects their orbital tracks. They are clearly evanescent bodies, at least in their active phases, the characteristic decay time being of order one to ten millennia; the short-period population must therefore be replenished from some reservoir.

Conventionally, the short-period comets are further divided into Jupiter-family ($P < 20$ yr) and Halley-type comets ($20 < P < 200$ yr; $7.4 \leqslant a \leqslant 34.2$ au). About 90% of the known short-period comets are Jupiter family members, so-called because their dynamical evolution is dominated by frequent close encounters with Jupiter. They comprise a much flatter system than the Halley types ($\bar{i} \sim 18^\circ$ as against $\sim 65^\circ$). Recent work involving long-term numerical integrations (Levison and Duncan 1994) has, however, shown that a definition based solely on orbital periods provides only a rough-and-ready discriminant between the families: certain observed short-period comets will switch allegiance between families perhaps a dozen times before their ejection from the solar system or infall to the Sun, and in fact observed Halley-types tend to have perihelion distances q less than 1 au while the longer period Jupiter-family comets have somewhat larger q, ranging up to the radius of Jupiter's orbit. A dynamically more useful distinction was earlier proposed by Carusi and Valsecchi (1987), who classify a comet as Jupiter-family if $T > 2$ and Halley-family if $T < 2$, where T is the Tisserand parameter with respect to Jupiter. The latter is given by

$$T = \frac{a_J}{a} + 2\sqrt{a(1-e^2)/a_J}\cos i \tag{3.14}$$

where (a, e, i) are respectively the semi-major axis, eccentricity and inclination (with respect to the Jovian plane) of the comet's orbit, and $a_J = 5.2$ au is the semi-major axis of Jupiter's orbit. In the circular restricted three-body problem, objects with $T > 3$ cannot cross the Jovian orbit, remaining forever outwith or within it. With this definition, over 90% of short-period comets remain within the same family during their entire dynamical lifetimes, the exceptions being a handful of interesting marginal cases. The existence of the $T = 2$ barrier suggests that either the Jupiter and Halley families have different source reservoirs, or that they represent different capture streams from the same one.

Levison and Duncan (1994) integrated the then known 160 short-period comets forwards and backwards for 10^7 yr or until the comet was lost from the system. Each comet was integrated from four slightly different starting positions. Each set of four orbits quickly diverged, revealing that the orbits of the short-period comets are chaotic and that their long-term behaviour can be studied only statistically. The median lifetime of the known comets was found to be $4.5 \pm 0.1 \times 10^5$ yr, with only 10 ± 4 objects out of the 640 remaining at the end of each 10^7 yr integration. The Halley-family objects ($T < 2$) have a lifetime of a million years, Jupiter's is about a third of this.

The controlling influence of Jupiter is evident in the frequency distribution of aphelion Q which peaks strongly around the distance of Jupiter, and the corresponding distribution of ω, the argument of perihelion, which clusters around 0° and 180°. These were thought to represent transient initial conditions consequent on recent close encounters with Jupiter (Levinson and Duncan 1994). However, they seem to be statistical equilibrium distributions (Harris, personal communication).

Levinson and Duncan find that the mean inclination of the Jupiter family comets increases with time, attaining a mean $\bar{i} \sim 27^\circ$ on a timescale of order 10^4 yr (as against 18° observed). If the Jupiter family is in a steady state, then presumably the dynamically older family members have a thicker distribution but have faded from sight, this extinct Jupiter family population being an order of magnitude more populous than the visible one. Quinn *et al* (1990) have argued that the relative flatness of the observed population indicates that the source of the Jupiter family must be even flatter. However, the origin of the family is currently uncertain (section 4).

3.2. The Edgeworth–Kuiper belt

The suggestion, made on cosmogonic grounds, that there might exist a ring or belt of comets beyond Neptune was first made by Edgeworth (1943, 1949), who further proposed that this belt might be a source of short-period comets (McFarland 1996). The primordial ring concept was independently proposed by Kuiper (1951), and the term 'Kuiper belt' was introduced for this system by Duncan *et al* (1988). Although the term does not properly reflect priorities, it is widely used and is employed interchangeably with the Edgeworth–Kuiper (EK) belt.

The first searches for trans-Neptunian objects using CCD technology were carried out by Luu and Jewitt (1988) and by Levison and Duncan (1990). Only a few dozen are currently known, but the estimated total population of such bodies larger than 100 km diameter, and lying between heliocentric distances 30 and 50 au, has increased steadily and currently stands at $\sim$70 000. The population of > 1 km bodies may be of order 10^{10}. The system is apparently flat ($\bar{i} \sim 10^\circ$), but as there is strong observational selection against the discovery of high inclination objects the true mean inclination could exceed 30°. Almost half the discovered bodies seem to occupy the 2:3 resonance with Neptune, an orbital niche which is relatively stable (a caveat is that, given the very uncertain orbits, the orbit computers sometimes *put* EK bodies there by fiat!). Pluto itself occupies this resonance and so presumably co-orbits with tens of thousands of bodies $\gtrsim 100$ km across; this leads inevitably to the speculation that the ninth planet is simply a high-mass member of the EK belt circulating in a particularly well-populated region of the orbital phase space.

The first modern indication that the belt might be unstable came from Torbett and Smoluchowski (1990), who followed the orbital evolution of about 200 test particles initially beyond Neptune for 10 million years. Beyond Neptune, in a region where the orbits had perihelia $30 \lesssim q \lesssim 45$ au, they found a chaotic zone such that (a, e) performed a random walk. They thus reaffirmed that Trans-Neptunian objects could indeed, in principle, be a reservoir for the short-period comet system generally.

A few years later, Holman and Wisdom (1993) carried out a more comprehensive study of particle stability in the outer solar system, involving the direct numerical integration of about 7000 particles for durations up to 800 million years. Beyond Neptune, out to 43 au, the test particles were found to have orbits whose semi-major axes remained relatively constant while their eccentricities might increase to the point where close encounters with Neptune, followed by ejection from the region, could occur. The outcome of their trials was a population decaying roughly as $1/t$, with a time constant such that about half of the primordial population would still remain. To account for the observed flux of new short-period comets (taken as one comet per century with diameter say $\geqslant 3$ km), and assuming that one in six Neptune-encountering comets eventually become visible, Holman and Wisdom deduced that the initial comet population in this region of the Kuiper belt must have been 9×10^9 and that half of this original population survives *in situ*. The current mass of the belt, assuming a mean comet mass of say 10^{18}–10^{19} g, is then about 0.7–$7M_\oplus$, as against upper limits of about 0.2–$1M_\oplus$ deduced from perturbations of Halley's Comet (Hamid *et al* 1968, Hogg *et al* 1991, Yeomans 1986). The constraints are clearly tight but there is as yet is no real inconsistency.

Later dynamical studies have progressively extended the integration times, integrations of up to 4 Gyr, virtually the age of the solar system, now having been carried out (Duncan *et al* 1995). These latter reveal the Kuiper belt to be a somewhat complex dynamical system. In agreement with earlier studies, they initially found many comets with low (e, i) and $q \lesssim 35$ au to be in unstable orbits, which must eventually have led to close encounters with Neptune and rapid ejection (10–100 Myr) either to the inner planetary system, or the Oort cloud, or interstellar space. One thus expects a clear zone between the orbit of Neptune and the inner region of the Kuiper belt. Almost all surviving objects within $\sim$41 au should lie close to the 2:3 mean motion resonance with Neptune, and may indeed orbit in this stable region for the age of the solar system. The range $\sim$40–42 au contains several secular resonances and is unstable. Beyond $\sim$42 au it is expected that many comets will be relatively unaffected by resonances with the planets. Thus, planetary perturbations have largely depleted the Kuiper belt within $\sim$32 au, but left it intact beyond $\sim$45 au.

The evolution of the EK belt is strongly affected, not only by planetary perturbations, but also by collisions. Davis and Farinella (1996) consider that the belt has a volume $\sim 10^3$ times that of the main asteroid belt, has perhaps 10^3 times the population (down to say 1 km), and random speeds about 10 times lower. The overall collision rate is thus, according to these authors, similar to that within the main asteroid belt. Their Monte Carlo simulations of these collisions lead to the expectation that bodies $\gtrsim 50$ km in diameter have been battered but not fragmented, and are probably to be thought of as rubble piles. Bodies smaller than $\sim$50 km across will generally be fragments from disruptive collisions. About 10 fragments in the range 1–10 km in diameter are produced each year, a few percent of these being injected into chaotic resonant orbits. To maintain the Jupiter family in a steady state requires an injection of $\sim$0.06 comets yr^{-1} (Levison and Duncan 1994). Thus, if one in two or three cometary fragments ejected from the Kuiper belt ends up in the Jupiter family, the latter can be maintained in a steady state. The adjustments here clearly involve a rather fine blend of new dynamical facts and somewhat more conjectural ideas as to the cohesion and constitution of comets. The system is, however, clearly eroding while the particular EK bodies observed and inferred must be considered to have formed in some quite different environment.

3.3. The Centaur family

A possible difficulty with a pure Kuiper belt source for the Jupiter family arises from the existence of a group of objects known collectively as the Centaurs, of which Chiron is the most massive prototype. Unlike the EK bodies, these are giant planet crossers, orbiting in the Saturn–Uranus region: a specific definition might be that they are bodies with perihelia lying beyond Jupiter (and its sphere of influence) but within the orbit of Neptune, and with aphelia $Q < 100$ au, say. Only a few such have been discovered to date (table 2). There are probably only a few Chiron-sized bodies (diameters $\sim$200 km), but to within a factor of a few there may be 2000 Centaurs ($\gtrsim$ 50 km say). Broad band photometry of some Centaurs and EK bodies has been undertaken by Luu and Jewitt (1996). These authors find them to be dark red, although with a diversity of detail; the colours are consistent with those of organic-rich, irradiated ices with variable impact re-surfacing (*loc cit*). While this similarity of surface properties might suggest that Centaurs are 'leaking' out of the Kuiper belt, there appears to be a problem of population balance. Evolutionary studies of Chiron and Pholus have been carried out (e.g. Hahn and Bailey 1990, Asher and Steel 1993) and reveal that these bodies at least must have arrived from elsewhere on timescales 0.1–1 million years. The *observed* Centaurs (currently six objects) thus appear to require replenishing at a rate of about 10 per million years, at which rate the EK belt (70 000 bodies between 30 and 50 au) would have been depopulated over the age of the solar system. This problem is greatly exacerbated by the need to replenish also the substantially larger (as yet *unobserved*) Centaur population. It seems that either the orbits of Chiron and Pholus are atypically short-lived, or the EK belt is inadequate by one or two powers of 10 to replenish the Centaur system.

Integrations of synthetic bodies in Jupiter-family orbits have been carried out by N Harris (personal communication) in which perturbations from all the planets except Mercury and Pluto are taken into account. It turns out that a significant proportion of Jupiter family bodies cross into the inner Centaur region, and there is in fact a brisk two-way traffic between these domains. It is quite probable therefore that at least the inner part of the Centaur region, and some large proportion of the Jupiter family, derive from some as yet unidentified source. However, orbits with semi-major axes $\gtrsim$ 15 au were not attained in these trials, and on this

Table 2. The Centaurs known to August 1996. Chiron and Pholus have diameters $\sim$180 km, 1994 TA $\sim$30 km and the others are $\sim$60 km across.

Name	a (au)	e	$i\,(^\circ)$
Chiron	13.70	0.38	7
1994 TA	16.82	0.31	5
1995 GO	18.13	0.62	18
Pholus	20.30	0.57	25
1993 HA_2	24.74	0.52	16
1995 DW2	25.03	0.25	4

basis it seems that the Jupiter and Centaur families would need to have been considerably more dense in the distant past to have produced the Kuiper belt (as during the 'late heavy bombardment' of $\sim$3.9 Gyr ago).

4. The source of the short-period comets

The sharpness of the observed (e, q) distribution contrasts with the diffuseness of the evolved, theoretical ones: either short-period comets have very short active lifetimes ($\sim 10^3$ yr), or the system is not in a steady state. Either way, it seems we must ask from whence do the Jupiter and Centaur families come. The EK belt represents an important new input into the question of comet and indeed solar system cosmogony. Indeed it is now widely assumed to be the long-sought source of the Jupiter family (Stern and Campins 1996, Farinella and Davis 1996, Yamamoto 1996) as well as of the more recently discovered Centaur population. However, the contrary view has also been expressed (Zheng *et al* 1996), namely that as a supplier of short-period comets the EK belt is an unnecessary luxury! Certainly it does not seem that either the Halley or possibly even the Centaur families can derive from the belt, and this suggests that there might exist some other reservoir, which would then also be the prime source of the Jupiter family. We consider the likely possibilities here.

4.1. Replenishment from the Oort cloud

The suggestion that the short-period system originates by capture from the long period one was made by Laplace two centuries ago. On this hypothesis, a comet on a parabolic orbit is imagined to lose energy during a close encounter with Jupiter. It is readily shown that a comet captured into the observable region (perihelion $q \lesssim 2.5$ au) by such an encounter will generally have aphelion Q in the range $5 \lesssim Q \lesssim 6$ au, while its longitude of perihelion ω will tend to cluster around 0° or 180°. The observed (Q, ω) distribution of the Jupiter-family comets is indeed of this character. In spite of this, however, single-pass encounters are considered to be an inadequate source of replenishment as they are simply too rare. Everhart (1972) proposed instead that capture might occur through recurring weak encounters with Jupiter. Following the evolution of some tens of thousands of comets for 2000 revolutions each, Everhart discovered a 'capture' region $4 \lesssim q \lesssim 6$ au and $i \lesssim 9^\circ$ within which transiting comets were captured with probability $\sim$0.007. Even this higher rate, however, turns out to yield a Jupiter family population one or two powers of 10 less than that observed, and all estimates of the transfer rate to date which consider only the parabolic flux within Jupiter's orbit have confirmed this discrepancy (e.g. Quinn *et al* 1990, Fernández and Gallardo 1994).

However, these estimates have all involved simplifications introduced by the computationally intense nature of the problem. The possibility of multiple capture, involving the simultaneous actions of Jupiter, Saturn, Uranus and Neptune, has recently been considered by Emel'yanenko and Bailey (1996). These authors followed the evolution of 50 000 bodies in initial near-parabolic orbits with perihelia in the range 0–31 au, with and without the action of non-gravitational forces caused by outgassing. Assuming uniform initial distributions in q and $\cos i$, they derived the probabilities $p_{SP}(q)$ of transfer into the short-period system, in various q-ranges as shown in table 3.

It is clear that, for the Halley family, the capture probability per comet is strongly biassed towards orbits with $q < 4$ au (in fact within $q < 2$ au the authors find a capture probability 0.0195). The capture region of the Jupiter family is significantly enhanced over

Table 3. Capture probabilities of near-parabolic comets into the Halley (p_H) and Jupiter (p_J) families respectively.

q (au)	$\overline{p_H}$	$\overline{p_J}$
0–4	0.0128	0.0002
4–6	0.0013	0.0008
6–10.5	0.0003	0.0003
10.5–18	0.0002	0.0002
18–31	0.0000	0.0004?

that found by Everhart (1972), largely because there is a capture region associated with each of the major planets.

Given these probabilities, the steady-state population of the families may then be estimated from N_c from

$$N_{SP} = \int \bar{L}_{SP}(q)\nu(q)p_{SP}(q)\,\mathrm{d}q. \tag{4.15}$$

Here the integration is carried out over the maximum range of q for which capture occurs, $\nu(q)$ is the number of near-parabolic comets arriving per annum per au, and $\bar{L}_{SP}$ is the mean dynamical lifetime of a comet in the short-period system.

Unfortunately little is known of the near-parabolic flux of new comets at large perihelion distance. Emel'yanenko and Bailey (1996) adopt a rough estimate due to Matese and Whitman (1989):

$$\nu(q)/\nu(1) = 1 + 0.014q^{1.82} \qquad q < 13 \text{ au} \tag{4.16}$$

$$= 5 \qquad 13 < q < 30 \text{ au.} \tag{4.17}$$

For mean lifetimes $L_H \sim 3 \times 10^5$ yr and $L_J \sim 10^4$ yr respectively, and assuming an entry rate of new comets $\nu(1) = 0.8$ comets yr^{-1}, there is a steady-state expectation of $\sim$15 000 Halley family and $\sim$315 Jupiter family comets brighter than $H_{10} = 7$. By comparison the observed numbers are $\sim$20 Halley family and (to epoch 1995) 53 Jupiter family comets with $q < 1.5$ au.

The large uncertainties in the high-q flux make little difference to the estimated equilibrium Halley population (the estimated population, considering only the near-parabolic flux in the range $0 < q < 4$ au, becomes $\sim$12 000) but a large difference to the estimated Jupiter family population (the number drops from 315 to $\sim$6). According to Fernández and Gallardo (1994) the discovery of active Halleys is only 5–10% complete, yielding a true current population of perhaps 100–200 such bodies. The observed Halley family population is thus deficient by about two powers of 10. It seems that this imbalance can be rectified only if the lifetime of a Halley-family comet as an observable entity is $\sim 0.01 L_H$, that is about 3000 yr or say 200 revolutions. Halley's Comet itself has been recorded for $\sim$2000 yr. At present it is an extremely dark object (albedo $\sim 0.03 \pm 0.01$), measuring $16 \times 8 \times 7.5$ km and losing volatiles from a few vents which amount to $\lesssim$ 10% of its surface area. The implication is that there may exist over 10 000 dark bodies (or at least their meteoric debris) in orbit as part of the Halley family. Such a population has not been detected, although in this instance, depending on how rapidly these dark bodies are degraded and dispersed as dust (section 5), it is by no means clear that the absence of evidence is evidence of absence.

A potential difficulty with an isotropic cloud as the major source of the JF comets is that, according to numerical trials by Quinn *et al* (1990), the inclination distribution of the captured comets reflects that of the source region. This contradicts the early work of

Everhart which indicated that only low-inclination comets were captured whatever the initial distribution of the source. If Quinn *et al* are correct, a flat JF distribution could not then arise by capture from an isotropic inner cloud; rather, a flattened source is implicated. However, in order to speed up their computations they exaggerated the masses of the planets by large factors, a procedure which is valid if capture is solely due to low-energy, diffusive processes but not if close encounters are an important element in the process (Stagg and Bailey 1989, Valsecchi and Manara 1996). Emel'yanenko and Bailey (1996) on the other hand, find that, while capture of near-parabolic comets into the Jupiter family can apparently explain the numbers (within the uncertainties), the observed family is much flatter than that obtained in their simulations. Thus, whereas their models produce equal numbers of prograde comets in the range 0–30° and 30–60°, the observed ratio is 7.7:1, and they suggest that the Jupiter family may derive from an inner, oblate Oort cloud. Zheng *et al* (1996) however, employing a Monte Carlo procedure based on capture probability cross-sections, find that the short-period system *is* adequately replenished from the classical Oort cloud. These findings could be further modified if it were to turn out that the secular resonances also play a significant role in the capture process. The issue of the inclination dependence based on dynamical considerations alone is not currently settled.

4.2. A modified capture model

The above arguments presuppose a steady-state balance between the supply of comets into the short-period system, and their loss from it. However, the mass distribution of comets is top-heavy, and the assumption of equilibrium may not be secure: the evidence of the Kreutz group (section 5.2) suggests the possibility that the short-period population might be dominated, at least sporadically, by the disintegration of single, exceptionally large bodies. Chiron, for example, may have $\sim$50 000 times the mass of a typical observed comet, and the tidal disintegration of such a body, say during a close encounter with Jupiter, would lead to a massive temporary increase in the Jupiter family population. This possibility has been investigated numerically by Pittich and Rickman (1994) and by Valtonen and his colleagues (personal communication). The latter computed the evolution of 10^6 comets arriving from the Oort cloud with perihelia $q < 6$ au. Of these, 870 were thrown into the inner planetary region, 104 of them impacted on Jupiter and 48 passed within 1.5 radii of the Jovian surface, where they assumed that comets would disintegrate under the influence of Jovian tides. The orbital evolution of these latter cases was followed; most of them had inclinations less than 25°. The dynamical memory of such an encounter is only 50–100 yr (Pittich and Rickman 1994), whence it may be concluded that the origin and low inclination of the Jupiter-family comets could indeed be accounted for by the recent tidal splitting of a single large comet into thousands of fragments. Clearly, with such an origin, there is no reason to expect the Jupiter comets to be in a steady state at any particular time. The 1889 splitting of P/Brooks 2 at $\sim 2R_J$, and the 1992 disruption of P/Shoemaker–Levy 9 into more than 20 fragments on passage within $1.35R_J$, are evidence that the process is not uncommon (apart from such tidal splitting, comets also break up for no obvious reason, at a rate of about 1% per annum). Depending on tensile strengths and internal structure, the sporadic disruption of very large comets could be a prime mechanism for replenishing the short-period population in bursts (Napier and Clube 1979, Napier 1983, Clube and Napier 1984a).

In summary, the source of the short-period comets remains unsettled. What has changed in the last few years is that a dearth of hypotheses has been replaced by a surfeit: the EK belt, an oblate inner cloud and the classical Oort cloud have all been proposed as source reservoirs. The Jupiter family may not even be in a steady state: if its replenishment is dominated by large transient inputs from disintegrating giant comets, then simple population balance arguments are vitiated. Nevertheless there are indications that the recently discovered EK belt is at best a partial answer to the short-period comet problem. Thus, the Jupiter family may be in a state of rough dynamical balance with the inner Centaur family, whereas the origin of the latter as spillage from the EK belt is problematic. The situation for the Halley family is that, so far, only the Oort cloud seems able to replenish it. There is little direct leakage between the Halley and Jupiter families if defined by the $T = 2$ barrier, but some if one employs the more relaxed definition by period only. Population balance arguments indicate that a large population ($\sim 10^4$) of degassed comets, which might vary from boulder streams to asteroids, should exist within the Halley family. Although only two such bodies have currently been observed, the issue is not yet critical since the great majority of dormant Halleys would be too distant for detection; and in any case their rapid loss from the solar system as dust is also possible. On balance, it seems unlikely that the EK belt will supplant the Oort–Öpik cloud as the principle source of comets an astronomical unit from the Sun. Evidence from the terrestrial record, presented in section 6.2, suggests that the classical long-period reservoir is the dominant source of Earth-crossing comets.

5. Cometary debris

5.1. The disintegration of comets

Once thrown into the region of the outer planets, the dynamical evolution of an Oort cloud comet speeds up greatly. Early studies of this problem (e.g. Lecar and Franklin 1973) generally involved numerical integrations of test particles with the perturbers, Jupiter and Saturn, moving in circular, coplanar orbits. It was found that most test particles underwent close encounters with Jupiter and Saturn within 10^4–10^5 yr, followed by an ejection from the solar system, although a few survived for about a million years. More sophisticated models, using fully consistent N-body integrations involving the four giant planets, have in essence confirmed these results (Gladman and Duncan 1990, Holman and Wisdom 1993).

Once in an orbit of short or intermediate period, with $q \lesssim 2.5$ au, a comet begins to outgas and decay. An understanding of the process is necessary to determine the expected population balance between various families, and in estimating the likely population of any inert bodies which might remain after degassing is finished. The latter is especially important in assessing the celestial hazard presented by the small-body population of Earth-crossers, say bodies in the telescopically inaccessible Tunguska impactor range. Unfortunately, little is known about the mechanisms or modes of decay.

A quantitative demonstration that meteor streams are the product of cometary decay was first made in 1866, when a brilliant display of the annual Leonid meteor shower, in essence a meteor storm, enabled the radiant to be determined. The arriving meteors were found to lie in the same orbit as a faint periodic comet, Tempel–Tuttle, which had been discovered only a few months earlier. Over 100 meteor streams have now been identified, some associated with Earth-crossing asteroids rather than currently active comets. The mass of the meteor stream associated with P/Halley has been estimated at 2.2×10^{17} g (McIntosh and Hajduk 1983), while the mass loss per revolution from the comet is likely to be well in excess of 10^{15} g. The mass of the comet itself has been put at 2×10^{17} g to within a factor of three (Sagdeev *et al* 1988). Thus, the past and future active lifetimes of this 76 yr period comet may each be only a few hundred revolutions.

More generally, the rate of mass loss from a short-period comet is of the form (Kresák and Kresáková 1987) $\dot{M} = ABC$, A an activity factor given by

$$A = (q + qe)^{-5/2}(2 + e^2)/2$$

when the instantaneous mass loss rate varies as the inverse fourth power of its heliocentric distance; $B = 10^{4-2H/5}$ a factor depending on the absolute magnitude H of the comet (its apparent magnitude if 1 au from both Earth and Sun); and C a calibration factor which varies from comet to comet. Beyond ∼1.5–2.5 au, cometary activity is generally slight, although some activity is seen in exceptional comets at much larger distances (such as P/Schwassman–Wachmann I, a comet in a near-circular orbit between Jupiter and Saturn, which has $q > 5$ au).

The calibration factor may be estimated from gas production rates based on UV spacecraft measurements, and dust production from infrared observations (both ground-based and from IRAS). The dust production rates in particular are not well determined. Dynamical and image analyses of the tails and comae of short-period comets are consistent with a power law distribution of grain sizes with index 3.5 ± 0.3; thus, while the total cross-section of the dust is dominated by small particles, the mass is concentrated in large ones. IRAS data, along with *in situ* measures of dust sizes from experiments carried on board the Vega and Giotto spacecraft, have led to significant upwards revisions of the mass loss rates in active comets; thus, for P/Schwassman–Wachmann I a combination of improved modelling and IRAS observations has increased the estimated dust loss rate from $\sim 10^4$ g s^{-1} to $\sim 6 \pm 3 \times 10^5$ g s^{-1} (Fulle 1992). Dust trails have been detected along the orbits of several short-period comets by IRAS (Sykes and Walker 1992), and these appear to be dominated by grains up to ∼1 cm in size. Thus, for an active short-period comet, the data seem to indicate a calibration factor of about 100 million tons of dust per century. Both this factor, and the gas-to-dust ratio (3:1 in P/Halley at its 1986 apparition), are likely to vary substantially from comet to comet.

The characteristic mass flux corresponds to the loss of a few metres of surface at each revolution, and the question arises whether a comet exposed to insolation will simply vanish, will outgas leaving a deep porous mantle or will develop a crust of refractory material which protects the underlying volatiles and ultimately chokes off the escaping gas and dust. P/Halley, at its 1986 apparition, had an active surface comprising a few jets through which gas and dust was ejected at ∼0.5 km s^{-1}. These made up only a few percent of the surface area of the comet; the remainder was inert and blacker than charcoal. It adds a few metres of dust to its crust at each apparition, and it seems very likely that the comet will end up as an inert asteroid. There are currently about 60 asteroids in comet-like orbits or comets in asteroid-like ones. The proportion of comets ending up asteroidal in appearance is currently open. Meteor streams, which are believed to derive from comets, have been associated with all Earth-crossing asteroids which pass close enough to yield them. The small asteroid Phaethon, for example, is co-orbiting with the Geminids, one of the major annual streams. The Apollo asteroid 1979 VA has been identified with P/Comet Wilson–Harrington (1949 III), indicating that the latter has become inactive within the past few decades. The active area on Phaethon can now only be a patch a few metres across, while similar limits have been put on Wilson–Harrington.

The evolution of a comet is not a smooth affair, however, and it may be interspersed by dormant phases and even splitting of the nucleus. The latter in particular appears to be a major mode of evolution. The splitting of a comet (‘a flaming torch of exceptional size’ according to Diodorus Siculus) was reported by Ephorus in 372 BC. Two fragments were formed, each one following a separate path across the sky. Such events have been recorded since the earliest times, comets sometimes being recorded as dividing into four or five pieces.

In the period 1846–1976, 21 comets were recorded as having split. A rate of one split per 90 revolutions for short-period comets, and one per 12 revolutions for long-period ones, has been derived by Kresák (1980). An even higher rate of one or more splittings per comet per century was inferred by Chen and Jewitt (1994), who applied image processing techniques to a sample of 49 comets observed since 1989. However, this latter estimate was based on only three observed splittings and is thus subject to large uncertainty. The phenomenon occurs in short-period and long-period comets, before or after perihelion passage, and at heliocentric distances at least out to Saturn; characteristic separation speeds are $\sim$0.5 m s^{-1}. Apart from the cases of tidal disruption, the splitting mechanism is not understood.

Since the interval between splittings appears to be much less than the visible lifetime of a typical comet, it is likely to be a major mode of comet disintegration. Depending on the extent to which cometary remnants enter an extinct, dormant phase (becoming asteroids), the phenomenon may be relevant to estimates of the current impact hazard. In particular, if a very large comet were to undergo a hierarchy of disintegrations, it is possible that the hazard would be highly variable in time.

5.2. Sungrazers

Comets have on occasion been observed to fall into the Sun, and it is clear that a fragile, icy body, passing close to the surface of a 6000 K mass whose surface gravity is 28 times that of the Earth, is unlikely to come away unscathed. Recent fragments of such an encounter may be represented by the Kreutz sungrazers. This is a group of 30 known comets which pass within two solar radii or less of the Sun's surface. They move in high-inclination, eccentric orbits, close to a common plane, with periods in the range of 500–1000 yr. The known orbits seem to divide into subgroups, consistent with the concept of hierarchic disintegration. The total population of the Kreutz group may be in the thousands, and some members are very large (such as the great comets of 1106, 1843, 1880, 1882 and 1887). The comet of 1882 was seen to break into four or five distinct comets, while that of 1106 AD may have been the parent of one of the sub-groups (comets 1882 II, 1945 VII and 1965 VIII). Marsden (1990) considers that the high-inclination Kreutz group represents the fragmentation products of a very large comet which passed close to the Sun in or before the 4th century BC. The progenitor may have been the Ephorus comet of 372 BC. There is thus observational evidence that large comets may undergo multiple disintegration, the fragments themselves sub-fragmenting into comets whose lifetimes may be measured in centuries or millennia. Such a fate probably awaits several known comets with widely different periods (e.g. P/Hartley–IRAS, P/Machholz and 1932 I).

The motion of a small body in the presence of a binary such as the Jupiter–Sun system is integrable, provided only the secular terms are considered (thus negelcting mean-motion resonances and periodic terms in the disturbing function). In these circumstances the orbit of a small body may be described by three approximate integrals of motion, namely its energy $1/a$, its projected angular momentum on the invariable plane $h_z = a(1-e^2)\cos^2 i$, and a quantity K proportional to its mean inverse distance from Jupiter: $K = e^2(5\sin^2 i \sin^2 \omega - 2)$, where ω represents the argument of perihelion (essentially its angular distance along the orbit measured from the ascending node). The motion can be represented by curves in the (e, ω) plane (Kozai 1962, Bailey *et al* 1992), from which it can be seen that, as in the galactic tide case, there are both librating and precessing solutions (figure 4). The neglect of periodic terms in the disturbing function might seem inadmissible when a comet has a close encounter with a planet, but numerical trials reveal that Kozai librations may often be recognized and persist even through such disturbances (Bailey and Emel'yanenko 1996).

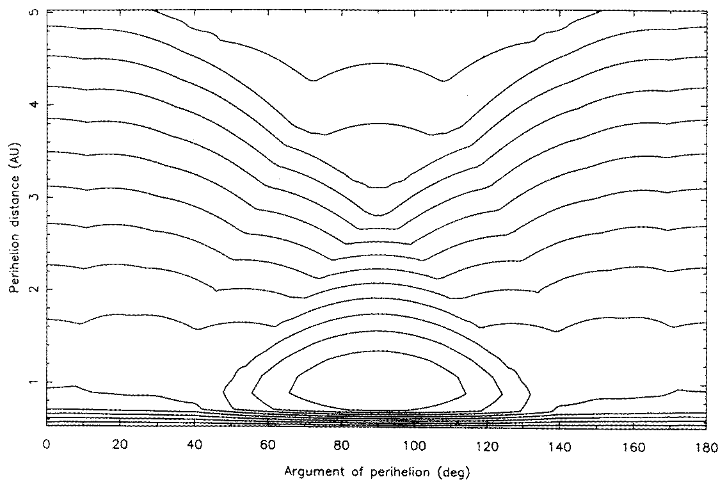

Figure 4. Kozai diagram for Comet P/Halley. Courtesy M E Bailey.

The above secular perturbation theory reveals that comets with small h_Z and high inclination may evolve into sungrazers. Since a and h_Z are constant, evolution proceeds in such a way that $(1-e^2)\cos^2 i \approx 2(q\cos^2 i)/a$ is conserved. Then since $q\cos^2 i \approx q_0\cos^2 i_0$, then the minimum value q_{min} obtained by q is given by $q_{min} \approx q_0\cos^2 i_0$ and a change in i to either extreme (0° or 180°) is accompanied by a decline in q: the perihelion distance becomes small when the orbit has precessed to a low inclination (for a prograde orbit) or a high one (for a retrograde orbit).

Bailey *et al* (1992) found analytically that, whereas only about 0.5% of an assumed isotropic flux of bodies with perihelia uniformly distributed over $0 \leqslant q \leqslant 2$ au would strike the Sun when perturbations were neglected, in fact about 15% are potential sungrazers, achieving this state within 500–1000 revolutions. Their numerical trials revealed that 5–15% of comets with initial $a = 100$ au and $q \leqslant 2$ au ended up as sungrazers (defined as having $q < 0.1$ au). About a third of comets with $q < 2$ au with the potential to become sungrazers actually do so. Thus, the rate of infall of comets to the Sun is largely due to secular perturbations rather than random encounters, and the fact that (for example) the Kreutz progenitor had an inclination ∼90° is probably not due to chance. These calculations are based on the assumption that the comets remain intact which is by no means assured.

A significant aspect of this evolutionary sequence is that the prograde situation maximizes the probability that the Earth will encounter cometary debris. A further implication is that many bodies not currently in Earth-crossing orbits may nevertheless become hazards under the influence of Jovian perturbations. Two fast-moving asteroids in such orbits are now known: (5335) Damocles was discovered in 1991, in an orbit ($a \sim 11.9$ au, $e \sim 0.87$, $i \sim 61.9°$) not unlike that of a Halley-type short-period comet, but unusual amongst the known asteroids. Its size is estimated to be 7–15 km; while 1996 PW, discovered in 1996, has $q \sim 2.5$ au, $i \sim 30°$ and the highest eccentricity of any known asteroid ($e > 0.99$), with an orbital period $\sim$ 5000 yr. Observational selection effects militate strongly against the discovery of such bodies, which spend most of their time in the outer solar system and far from the ecliptic. The total population of such bodies greater than a kilometre or so across may be in excess of 10^4. Asher *et al* (1994) found the orbits of such bodies to be chaotic; they spend perhaps a quarter of their lives in Earth-crossing orbits.

5.3. The Shoemaker–Levy 9 comet

On 25 March 1993 a 'squashed comet'—a bar of light about an arc minute in length—was discovered in the course of a long-term photographic search for asteroids and comets undertaken with the 46 cm Palomar Schmidt telescope. The object was only four degrees from Jupiter, aligned towards the planet and co-moving with it; subsequent observations revealed it to be a string of comets, each with a short inclined tail. As positional measurements accumulated, coupled with pre-discovery photographs, the calculated orbit was progressively sharpened up, and on 22 May 1993 Marsden announced that Periodic Comet Shoemaker–Levy 9 (SL9) would probably strike Jupiter in July 1994.

Backtracking the orbit, it emerged that close approaches to Jupiter had occurred in 1972 and 1970, prior to which the orbit is very uncertain because of the large angular deflection induced by such encounters. The comet had probably orbited Jupiter for many decades, and possibly even for centuries, before its final impact on to the planet. The jovicentric orbit had a period 2–3 yr (semi-major axis ∼0.16–0.2 au) on which a 22 yr Kozai oscillation was superimposed. In general a comet in an initially high-inclination orbit, moving within the Jovian sphere of influence, is subject to Kozai evolution, and in analogy with the sungrazers has its angular momentum systematically removed by the secular perturbations of the Sun. It may collide with Jupiter when its perijove reaches $q_J < R_J$, where Jupiter's mean radius $R_J \sim 0.0005$ au. The capture of comets into temporary satellite orbits around Jupiter was known to have taken place a few times before, but these orbits were generally short-lived, although they sometimes led to close encounters with the planet. Thus, P/Gehrels 3, in a loose orbit around Jupiter, approached to within $2.8R_J$ in August 1970 (curiously, at the same time as the previous extremely close approach of SL9). SL9, however, was unique amongst known comets in the long duration of its temporary capture.

Kary and Dones (1996) investigated the impact statistics of short-period comets on to Jupiter by following the motion of about 50 000 synthetic comets, distributed like the known Jupiter family, for about 100 000 yr. Over half of these comets were temporarily captured in orbit around Jupiter, although only about 300 of these captures were long-term (> 50 yr). There were 750 impacts, of which a quarter were by comets in bound orbits. From these figures the impact rate of $\geqslant 1$ km comets on Jupiter, from Jupiter-family comets, was estimated to be about one in 240 yr although with a very large uncertainty. Assuming breakup occurs within the Roche limit $2.4R_J$, this corresponds to a fragmentation event every century or so (close to Jupiter, gravitational focussing causes the comet flux to scale roughly linearly with distance from the planet). Cometary fragmentation by Jupiter may thus be rather common, although the subsequent chain of impacts on to the surface of Jupiter was unusual. Impacts of tidally fragmented comets were estimated to take place on a timescale of about one to ten millennia, only ∼5–10% of which were due to comets in bound orbits. Thus, the specific SL9 sequence of events was estimated to occur only once every 10 000–200 000 yr.

A novel estimate of the fragmentation rate of comets in the Jovian environment was given by Melosh and Shenk (1993), who attributed the crater chains on Ganymede and Callisto to the impacts of recently fragmented comets. Thirteen such chains have been recognized on Callisto and three on Ganymede, nearly all on the Jupiter-facing hemispheres of these satellites. These linear chains are from 40 to 620 km in length and may comprise up to 20 contiguous craters. The longest, the Gipul catena on Callisto, comprises ∼20 craters 20–40 km in diameter and may have been formed by the fragments of a ∼5 km diameter comet. Given a 4 Gyr age for the surfaces of the satellites, Melosh and Shenk inferred a mean recurrence time of ∼ 80 yr between fragmentation events due to close encounters with Jupiter, in good agreement with the best estimate of Kary and Dones (1996).

The linear dispersion of the fragments from SL9 suggests that the breakup involved a radial separation of the original nuclear fragments, and from ground and space-based measurements of the separation of the brightest half dozen or so fragments, Sekanina *et al* (1994) concluded that they had separated from each other only two hours after the comet had passed 0.0006 au from the surface of Jupiter on 7 July 1992. Models for the SL9 nucleus range from a large, coherent, extremely weak nucleus, through a rubble-pile to a strengthless aggregate of dust. If the disintegration of the comet was due to Jovian tides, a tensile strength of $\lesssim 10^{-3}$ bar is implied: in Earth's gravity, SL9 material more than a few cm high would crumble under its own weight. The fact that the comet broke into individual nuclei rather than a stream of meteors seems however to suggest some internal cohesion, consistent with the rubble-pile model (Asphaug and Benz 1996, Scotti and Melosh 1993). The diameter of the original nucleus can only be estimated indirectly, since most of the light came from the comae surrounding the nuclei; most estimates put the original nucleus at 1–2 km in diameter, with densities in the range 0.2–0.6 g cm^{-3}. A middle-of-the-range estimate for the impact mass and energy of each larger fragment is $\sim 10^{14}$ g and $\sim 10^5$ Mt respectively (Asphaug and Benz 1996).

Over the week 16–22 July 1994, about 20 fragments struck Jupiter at about 60 km s^{-1}. The impacts took place a few degrees over the visible limb of the planet, but rotation took the impact sites into view within about 15 minutes of each collision. Fragment A (the first to strike) was seen from Earth as a spot about as bright as Io, and this and subsequent impacts were easily detected from Earth. The Galileo spacecraft was fortunately well placed to detect the impacts as they occurred. Probably the most energetic collision was that of fragment G. Near-infrared spectrophotometry using the Galileo instrumentation was consistent with the creation of a fireball 40 km in diameter, at a temperature 2200 K, within 20 s of impact. Within 40 s of impact, the fireball had cooled to 1500 K and expanded to 80 km. The Hubble Space Telescope detected the impact plume (visible also with some ground-based telescopes) as it rose above the horizon, largely channelled back along the entry corridor of the comet. The plume reached 3000 km altitude within 10 min, as it expanded, before flattening out into a thin pancake; the latter had attained over 10 000 km in diameter by the time it re-impacted back on to Jupiter at 10 km s^{-1} 20 min after impact, heating much of this wide region to $\sim$1000 K. In general the most intense infrared flashes, characteristically $\sim$100 times brighter than the initial fireballs, were generated by the fall-back of impact ejecta. After each of the larger impacts, a bright ring of 3–4 μm radiation was observed to spread out from the site, reaching a characteristic diameter $\sim$30 000 km an hour after impact. The depth to which the fragments penetrated is uncertain. The relative dearth of water in the spectra suggests that the fragments did not reach the water cloud level which lies about 20 km below the (probable) ammonia clouds which constitute the dark belts of the planet. Thus, disintegration may have taken place at a few bars ambient pressure. Simple analytic models agree surprisingly well with those computed from high resolution smoothed particle hydrodynamic codes. In such models a 1 km bolide deforms into a 'pancake' and is torn apart when aerodynamic forces exceed its tensile strength, and loses over 90% of its impact energy within an atmospheric scale height (MacLow and Zahnle 1994). However, in these analyses the material strength of the bolide was neglected, and the hydrodynamic codes were two-dimensional, whence azimuthal effects were not modelled. Borovicka and Spurny (1996) pointed out the importance of fragmentation over ablation in meteoric phenomena. Scaling up from the behaviour of bright terrestrial fireballs, they argued that the SL9 fragments probably broke up and deposited their energies high in the Jovian atmosphere, above the clouds. Of the various post-impact phenomena, the most conspicuous were the dark spots, easily visible in a small telescope, and the most conspicuous ever seen on the planet. These are probably blankets of dust, high in the Jovian atmosphere, with settling times of many months or years.

In general, in spite of uncertainties and surprises, the impacts have confirmed the broad correctness of pre-existing theoretical studies of bolide impact, at least with regard to their prompt effects. Simulations of the fireball development on Jupiter, for example, agree well with the plume development observed with the Hubble Space Telescope. Thus, the SL9 'field trial' gives one some confidence that the immediate effects of a terrestrial impact would be much as predicted (e.g. Adushkin and Nemchinov 1994). However, the SL9 incident has drawn attention to the issue of planetary impacts in general, and more specifically to (a) the importance of cometary disintegration in discussions of terrestrial catastrophism; (b) the global nature of the 'splash' following an impact, which on Earth might set off worldwide conflagrations; and (c) the lingering blanket of stratospheric dust, which would reduce insolation with a deleterious effect on climate and food chains (Hoyle and Wickramasinghe 1978, Napier and Clube 1979, Alvarez *et al* 1980).

5.4. The Taurid complex

One of the most remarkable features of the space between the inner planets is the system of meteors and asteroids known as the Taurid complex (TC). Central to the system is Encke's comet, a faint telescopic object a few km in diameter and the only active short-period comet known in a stable Earth-crossing orbit ($q = 0.341$ au). Four of the $\sim$100 known meteor streams are associated with P/Encke, the main streams being the Taurids, a broad November shower with many bright fireballs, and the Beta Taurids, a daytime shower of radar meteors which peaks around 30 June each year. In 1952 Whipple and Hamid used an analytic theory to discuss the orbital evolution of a few photographic meteors ($\sim$1 cm) in the Taurid stream and concluded that they derived from two parent bodies in similar orbits, one of them P/Encke, the other a possible fragment which had separated from it at some time in the past. Jones (1986) used numerical integrations with Jupiter as the only perturber and deduced that the stream was $\sim$0.1 Myr old. Modelling an additional asymmetric devolatilization force in association with the presumed hierarchical splitting of the Taurid progenitor reduced this timescale to $\gtrsim 0.01$ Myr (Asher *et al* 1993).

Meantime, an exceptionally large but displaced dust trail in orbit with P/Encke had been detected with IRAS (Sykes and Walker 1992) while Clube and Napier (1984a) found that several of the known Apollo asteroids were in orbits closely resembling that of P/Encke (a recent listing is given in table 4). Comparison with synthetic near-Earth asteroid distributions simulating the overall orbital distribution of the real system showed that this asteroid grouping is real at a high confidence level (Napier 1993, Steel *et al* 1994). The asteroids, and the Taurid meteor streams, are immersed in a huge swathe of meteoroidal material, the Stŏhl streams, from which about half the sporadic meteors derive (Stŏhl 1983). These streams appear to merge into the zodiacal cloud and to constitute a bridge between the cloud and the Taurids. A second, smaller group of asteroids appears to be associated with the large Earth-crossing asteroid Hephaistos: these have similar (a, e, i) to the Encke-group but significantly displaced longitudes of perihelia ϖ (figure 5).

The Poynting–Robertson lifetime of the zodiacal cloud particles is only $\sim$10 000 yr, and it has long been recognized that there are no currently adequate sources of replenishment. Whipple (1967), Kresák (1978a) and others suggested that Encke's comet had once been an exceptionally large body, whose disintegration had released sufficient dust to create the zodiacal cloud. Clube and Napier (1984a) proposed that a giant progenitor comet responsible for the Stŏhl meteoroidal stream entered the solar system $\sim$20 000 yr ago and underwent

Table 4. Osculating elements of the most probable TC asteroids. Another ∼5–10 outlying members may be identified with less certainty.

Object	a	e	i	ϖ
Comet Encke	2.20	0.850	6.0	160
1991 TB	2.40	0.835	8.6	132
1982 TA	2.30	0.772	12.2	129
1984 KB	2.22	0.764	4.8	146
2201 Oljato	2.18	0.710	2.5	172
1979 XB	2.26	0.713	24.8	161
1987 SB	2.16	0.649	3.0	167

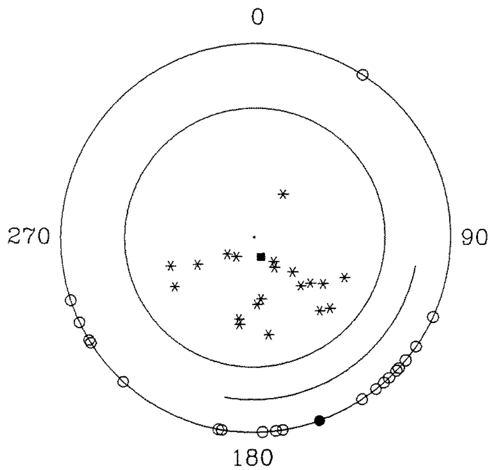

Figure 5. Distribution of longitudes of perihelion of TC asteroids (open circles), Encke's Comet (full circle) and the Taurid meteors (arc of circle). The stars represent the orbital closeness of each asteroid to that of P/Encke using a standard metric (a reduced D-criterion). The asteroids are closer in phase space to the comet than the meteors.

a hierarchy of disintegrations, giving rise to Encke's comet, the associated asteroids, the Taurid meteor streams and the zodiacal cloud. They associated the severe stratospheric dusting consequent on this disintegration with the last ice age, which ended ∼11 000 BP.

The possible long-term dynamical evolution of P/Encke (disregarding any non-gravitational or axisymmetric devolatilization forces) has been studied by Levison and Duncan (1994), Valsecchi *et al* (1995) and others. The comet becomes sungrazing in 50 000–100 000 yr (the orbits of most Earth-crossers are chaotic and the integrations have only statistical validity). In this case the sungrazing phenomenon is only partly due to the Kozai cycle; rather the trends in q are dominated by secular resonances with Jupiter and Saturn. The orbital elements of these planets oscillate slowly, with three fundamental frequencies, namely the precession rates of their longitudes of perihelion (denoted by g_5 and g_6 respectively) and that of their mutual line of nodes on the invariable plane (s_6). These cycles impose precessions of the longitude of perihelion and node of a body in an Encke-like orbit, at rates denoted by g and s respectively; and superimposed on these precessions, the grand cycles of the jovi-saturnian oscillations are also reflected in a spectrum of oscillations in eccentricity e and inclination i of the small body: for example oscillations in e are dominated by three frequencies, g–g_5, g–g_6 and $2(g$–$s)$, the latter being the Kozai cycle. Resonances occur when these quantities are zero, and are denoted ν_5, ν_6 and ν_{16} respectively. Close approaches to the terrestrial planets are also important in that they may induce transitions into chaotic orbits.

The domain of the Taurid asteroids is bounded by the ν_5 and ν_6 resonances, and the 4:1 and 3:1 mean motion resonances with Jupiter. The proximity to these resonances suggested to Valsecchi *et al* (1995) that the Taurid asteroid population is transient, with the resonances constituting entry and exit corridors connecting the Taurids principally to the main belt asteroid system but also to the Jupiter family of comets. For example numerical integration of the past orbit of Mithra (not listed in table 4) reveals that it could have been captured from the Jupiter family $\sim 3.5 \times 10^5$ yr BP through a series of encounters with Jupiter, followed by a steady decline in semi-major axis as time proceeds.

A problem arising from these trials is that secular resonances are found to act very slowly, on timescales a few times 10^5 yr. This is much greater than the generally recognized physically active lifetimes of short-period comets. Whipple (1996) has recently argued for a few times 10^5 yr lifetime for P/Encke, consistent with the result of Jones (1986). Even so, it is difficult to reconcile the long dynamical transition times with the short lifetimes set by the rate at which meteor orbits in the Taurid showers disperse. Indeed this dispersion cannot itself be understood unless we also suppose that each stage in the hierarchical disintegration of the original cometary-meteoroidal source gives rise to an additional, random non-gravitational force producing an additional, significant spread in orbital elements. Thus, differential precession rates would yield a dispersion $\delta\varpi \sim 90^\circ$ in the Taurid streams within $\sim$20 000 yr. A steady creation of the Taurid meteors over the last 100 000 yr would have yielded sets of related streams dispersed around the ecliptic, each set corresponding to a precessional cycle (5–10 000 yr) of the argument of perihelion ω (Asher *et al* 1993, Steel and Asher 1996). An age 20 000–30 000 yr is suggested for the major disintegration period of the Taurid progenitor (or one of its major offshoots), when much of the material now in the Stŏhl stream and zodiacal cloud was initially formed.

These authors found that chaotic dynamical evolution of the TC asteroids yields a dispersion in orbital elements which is, after 30 000 yr, only about a third of that observed (although, in the $\sim$1 Myr timespans considered by Valsecchi *et al* 1996, the asteroid groupings disappear altogether). Thus, non-gravitational forces may significantly affect the dynamics of active comets, these having acted to shorten the semi-major axis of 2P/Encke by $\sim$0.0037 au in the 170 yr since its discovery. Adding simple random walk perturbations to the numerical integrations, based on the current observed non-gravitational forces acting on P/Encke, Steel and Asher found that the observed dispersion in the elements of the TC asteroids may not only be reproduced but that the transition from Jupiter-family to Taurid-like orbits was speeded up considerably, consistent with a fragmentation from a common giant progenitor 20–30 000 yr ago.

The numerical work described above reveals that there are indeed routes from the Jupiter family (or even the main belt asteroid system) into the TC asteroid group which are facilitated by the action of non-gravitational forces. However, since disparate, unconnected objects may enter through these corridors, it is conceivable that the current TC asteroids have no generic connection, and the small ϖ dispersion might then be ascribed to an observational selection effect. Such an effect may well exist (Valsecchi, personal communication), but it could not account for the imbedding of the TC asteroids within a major component of the inner solar system dust population (Taylor 1995, e.g. figure 6). The generic connection between the Taurid asteroids remains a live issue.

Objects within the TC span a wide range of sizes, from sub-micron particles (10^{-12} g: Singer and Stanley 1980), to radar meteors (10^{-3} g: Steel 1995), visual meteors (1 g: Stohl

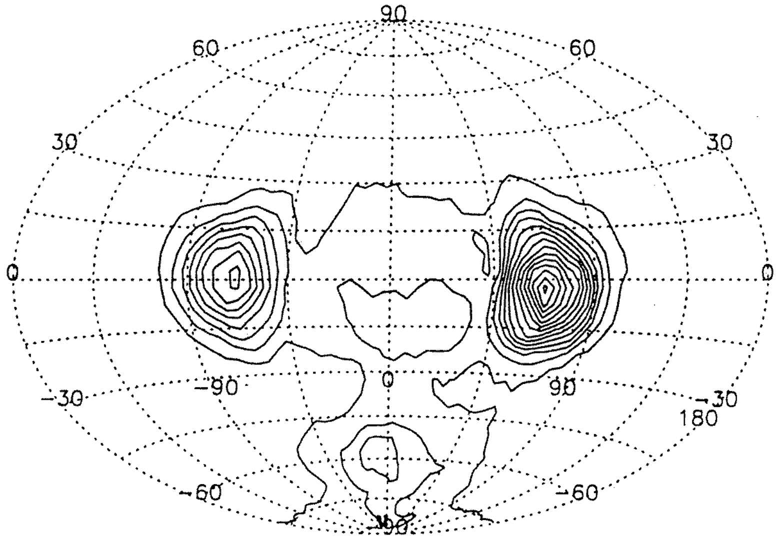

Figure 6. Seasonally averaged Adelaide meteor observations. The concentration into two streams (helion and anti-helion) is clear. These contain the Stohl streams within which P/Encke and the TC asteroids are imbedded. Courtesy D I Steel.

1983) and fireballs (10^9 g: Ceplecha 1992). Larger bodies appear also to exist, singly or in swarms. The Tunguska object of 30 June 1908 had a mass $\sim 10^{11}$ g and an impact energy 10–30 megatons, and probably belonged to the Beta Taurids (Kresák 1978a, but see Chyba *et al* 1993); the Moon was struck by a seismically detected swarm of $\sim 10^5$ g boulders in June 1975 (Dorman *et al* 1978); and it has even been suggested that a Taurid asteroid impacted on the Moon in late June 1178 AD, creating the 20 km Giordano Bruno crater (Hartung 1993). The occurrence of the latter event within the last 10^3 yr is very improbable (although no more so than the occurrence of the SL9 event within the past few years). Very little is known about the fine structure within the complex; for example the possibility that Tunguska-sized objects might be bunched into swarms cannot be excluded. The latter possibility, and the prospect that an ice age may be generated by the arrival and disintegration of a large short-period comet, are clearly vital elements in any assessment of celestial hazards or in the formulation of a defence strategy.

Given the deficiency of current astronomical knowledge, Asher and Clube (1993) have assembled climatic and solar cycle evidence for the long-term generation of meteoroidal dust within the inner solar system dependent on a primary ‘Stohl stream’ source in 7:2 mean motion resonance with Jupiter, having a libration period of $\sim$400 yr. Such a source disintegrating over millennia essentially fills the resonance with an extended swarm of smaller meteoroidal debris for which there is some observational evidence, and which in turn bombards the source, producing a 200 yr modulation in the supply of inner solar system dust eventually reaching the zodiacal cloud. Such a cycle is now well established for the ^{14}C content and growth pattern of some arboreal cellulose (Sonett and Finney 1990, Thomson 1990) while the latter is remarkable also for the presence of higher frequency modulations suggestive of the 7:2 Jovian resonance. Adopting this as a working model, noting that the pre-discovery apparitions of P/Encke have never been found, and recognizing also that P/Encke is a weak source of dust yet associated nonetheless with a separate but exceptionally massive dust trail, the suggestion was advanced that a substantially devolatilized TC progenitor currently lies within this trail. Thus, the model would indicate that P/Encke essentially separated from the librating, unseen progenitor and its then enhanced trail as recently as ∼1786 AD, and subsequently escaped resonance under the influence of a significant non-gravitational force.

The strength of this dust generation model, which is based on a wide variety of apparently disparate astrophysical and geophysical data, resides in its predicted effects, particularly the longer-term interactions of the primary progenitor stream with the Earth during its (broad) nodal intersections with the terrestrial orbit. These occur in pairs at $t \sim$ AD $500 \pm 2500n$, $n = 0, 1, 2 \ldots$, producing a global cooling cycle consistent with an extended warming epoch at the present time and going back several centuries (cf Epstein and Krishnamurphy 1990). This predicted cycle appears now to be present also in both the polar precipitation (dust) and ocean sediment (debris) records of the Holocene and late Pleistocene (O'Brien *et al* 1995, Bond and Lotti 1995). The cycle is thus also continuous with the final stages of the last glaciation, indicating that extreme ice age conditions are themselves correlated with periods of enhanced meteoroidal and Tunguska-like bombardment.

6. Comets and the Earth

6.1. A brief history of catastrophism

The speculation that the cometary environment might be hazardous for mankind has certainly long been admitted. Thus, Halley, having recognized that the comet now bearing his name returned periodically to the inner solar system, realized also that this comet or one like it might on occasion strike the Earth with calamitous results. In fact Newton and Halley, it is now realized, were constrained by their time to work under conditions of considerable censorship. The acceptable part of Newton's scientific output was of course published and has repeatedly proved its worth over 300 yr. The unacceptable part dealt with 'blazing stars' and eschatology, and remained unpublished for some 250 yr. One of the first to examine this material (Keynes 1947) was so taken aback with the contrast as to dub Newton not so much 'the first of the age of reason' as 'the last of the magicians, the last of the Babylonians and the Sumerians'. Thus, it was in fact the founding fathers of the Royal Society in Restoration England who hit upon the 'enlightened' step of deriding the cosmic threat and associated public anxiety, and it is largely an achievement of that particular epoch that cosmic catastrophes were absolutely discarded and the scientific principle of uniformitarianism was subsequently put in place between 200 and 150 yr ago. The point here of course is that catastrophism as a subject can be traced back to the foundation of civilization itself. Indeed there is reason to suppose that the fear of catastrophe has been rather frequently revived by terrestrial encounters with 'mini-SL9' trails (section 5.3), not least during the last apparently major period of global cooling which coincided with the so-called Dark Age about AD 400–600 (Clube 1995 and references therein).

By the late 18th century however, partly as a result of the perceived inconsequential nature of meteor showers, and partly as a result of the growing impatience with failed celestial prognostication, an anti-catastrophist or uniformitarian perception of the celestial environment became established and dominated thinking in the natural sciences for the subsequent ∼150–200 yr. Thus, in geology, the perception emerged that mountains rose and sea levels fell in response to slow-acting, internal forces, while in biology, species competed and adapted within a gradualist environment. From time to time, it was suggested that the impact of a large meteorite might force geological change. In the 1930s three Apollo asteroids were discovered, inducing the meteoriticist Nininger (1942) to suggest that the varied and rapid geological and biological changes which characterized geological boundaries (and often defined them) could have been caused by impacts. The palaeontologist De Laubenfels (1956) proposed that a 'giant meteorite' might have caused the mass extinction of the dinosaurs 65 Myr ago; McLaren (1970) proposed an ocean impact to account for the mass extinctions of corals which occurred in the Late Devonian; and Urey (1973) pointed out that the ages of several tektite groups coincided with geological boundaries. He thought it 'possible and even probable that a comet collision with the Earth destroyed the dinosaurs and initiated the Tertiary division of geologic time.' These and similar proposals, however (e.g. Öpik 1958, Gallant 1964), had little effect on mainstream thinking in the Earth sciences.

Curiously, a second strand of thinking with a celestial import was running through the geological literature for much the same period of time. About 70 years ago, Holmes (1927) had considered that orogenies, vulcanisms and sea-level changes occurred in 30 Myr cycles, overlain by larger events at 250 Myr intervals (figure 7). Large-scale continental movements, often corresponding to the well known geological boundaries, appear to have taken place rather rapidly at such epochs, the intervals themselves being relatively quiescent. Fischer and Arthur (1977) claimed a marine mass extinction cycle of 32 Myr and a longer climatic one of ~300 Myr; and there were many similar claims relating to orogenies, glaciations and so on (for reviews, see McCrea 1981, Williams 1981, Clube and Napier 1986). Several authors recognized that the 30 Myr cycle was similar to the half-period of the Sun's vertical motion through the galactic disk, while the longer cycle was akin to the galactic year. However, no plausible causative mechanism was found; nor were the geological data at the level where rigorous demonstrations of periodicity or otherwise could be given.

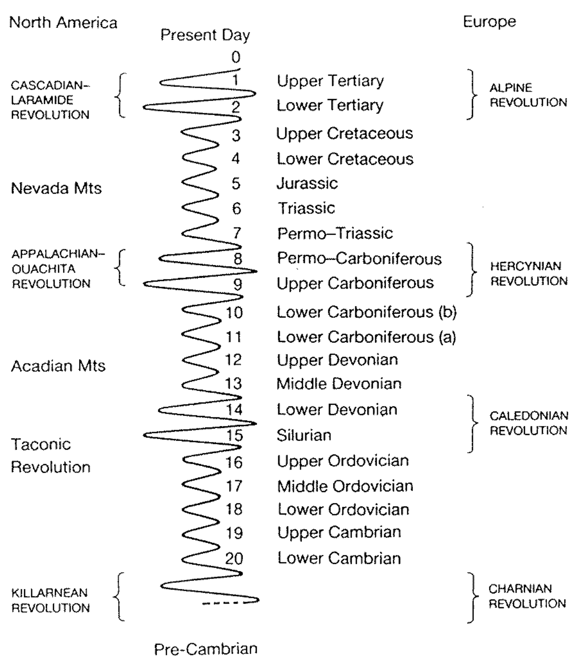

Figure 7. An early representation of a ~30 Myr cycle in the geological record, due to Holmes (1927).

In the 1970s, however, it began to seem that these apparently disparate strands might have an underlying unity. First, terrestrial impact craters were being found in increasing abundance, as were small Earth-crossing bodies. Quantitative assessments of likely impact rates and megatonnages became possible; both were sufficiently high that geological and biological signatures became expectations rather than purely speculative propositions. At the same time it was becoming clear from the discovery of the molecular cloud system that the classical Oort cloud was prone to major disturbances in the galactic environment. The unifying link was based on the assumption that an appreciable proportion of the Earth-crossing asteroids are degassed comets, ultimately deriving from the Oort cloud (Öpik 1963). Thus, periodic disturbances of the Oort cloud as the Sun circulated in and out of spiral arms led to periodic bombardment episodes on Earth, with consequent recurring trauma (Napier and Clube 1979, Clube and Napier 1982, 1984a, 1986). The discovery of iridium of probable extra-terrestrial origin at the Cretaceous-Tertiary boundary (Alvarez *et al* 1980) strongly revived the impact hypothesis, while the claim of a 26 Myr periodicity in the marine fossil record (Raup and Sepkoski 1984) likewise revived interest in a possible galactic connection (and also spawned a number of *ad hoc* scenarios, discussed for example by Bailey *et al* 1990).

6.2. Periodicity in the terrestrial record

Over the bulk of the observed range, and allowing for discovery selection effects, the mass distribution of comets is found to be a power law with population index approximately -1.8 (Hughes and Daniels 1982). There may be a steepening of slope at high mass; nevertheless giant comets appear to strongly dominate the mass influx to the near-Earth system. Best estimates yield a cumulative flux of long-period comets, at the high mass end, given by

$$F \sim 1 \times (d/5)^{-2}$$

comets au^{-1} yr^{-1}, d the diameter of the comet in km (Bailey *et al* 1994). There is no securely known upper limit to d, but several historical comets appear to have been $\gtrsim 100$ km in diameter. A giant long-period comet ($d \gtrsim 100$ km) is therefore expected to cross the Earth's orbit about once every 400 yr, a Chiron-sized body ($\sim$300 km) once within the timescale of civilization. The injection rate of giant comets from a chaotic, trans-Saturnian orbit into a stable Earth-crossing one is of order 10 Myr^{-1} (*loc cit*). The apparent inability of the EK belt to replenish the Centaurs in chaotic orbits within $\sim$15 au of the Sun suggests that, ultimately, most of them may have an Oort cloud source, and hence on long time-scales the Jupiter and sub-Jovian, as well as the Halley, families may have a time-dependent flux that reflects the Sun's variable ambient galactic environment.

Equation (2.4) reveals that the strength of the vertical galactic tide acting on a comet is proportional to the ambient density. The effect of varying the ambient density is to vary the rate of the precessional cycle, and hence the flux into the loss cone, *pro rata*. The long-period comet flux thus samples the instantaneous local density as the Sun oscillates vertically, and provided the amplitude of the oscillations is comparable with the half-thickness of the galactic disk, a significant cyclic variation in flux will ensue (Napier 1987, Bailey *et al* 1990). Detailed analyses (Matese *et al* 1995, Clube and Napier 1996) confirm this proportionality to a high degree of accuracy over a wide range of Oort cloud models ranging from classical ($\gamma = 2.5$) to condensed ($-2.5 \leqslant \gamma \leqslant 0.5$).

According to Bahcall and Bahcall (1985), the half-period of the Sun's vertical motion lies in the range 26–37 Myr, with maximum vertical excursions in the range 49–93 pc. The longer half-periods were obtained for models in which a spheroid of half-height 700-

pc contained substantial unseen material. However, Gilmore and Wyse (1987), from an analysis of K dwarfs at the south galactic pole, found no evidence for significant hidden mass beyond ∼100 pc. The likely range of solar half-periods is then found to be 26–32 Myr, with vertical amplitude in the range 50–82 pc. The greater amplitudes are associated with the longer periods, and the motion is simple harmonic to a high degree of approximation.

The distribution of known matter in the solar neighbourhood comprises a spiral arm population of young material (stars and gas) with half-height and vertical velocity dispersion (∼50 pc, ∼5 km s^{-1}) respectively, a more evenly spaced young stellar population (100 pc and 7 km s^{-1}) and a merging of older components (250 pc, 20 km s^{-1}). The local disk may be crudely modelled by an exponentially declining population of vertical density amplitude $Z_{1/2} \sim 50$–100 pc. For a vertical amplitude of 50 pc in the solar orbit, the long-period comet flux is then expected to have a periodicity ∼26 Myr with a peak-to-trough amplitude in the approximate range 1.5 to 4 (Matese *et al* 1995). A solar motion of amplitude 80 pc yields a periodicity ∼32 Myr with amplitude in the range ∼2.2–5. If these variations in comet flux are reflected in terrestrial signatures, a strong periodicity in geological phenomena is certainly to be expected. The flux variations in these circumstance are primarily due to small changes in the critical semi-major axis—the tidal radius—above which the loss cone is permanently filled. Currently this critical radius is ∼25 000–30 000 au (section 2.1).

Depending on the nature of the primary perturbers, granularity in the tidal force may become important at small dimensions. In that case impulsive perturbations are generated (2.3), and so the galactic disk material disturbs the Oort cloud not only smoothly, through the adiabatic tide, but also stochastically, through penetration by individual gravitating bodies. The deflection δv suffered by a comet due to a perturber of mass m, velocity V varies as m/V, and so the volume of tube disturbed out to $\delta v = 2Gm/(D_L V)$ varies as $(m/V)^2$. Thus, the flux of comets from all penetrating bodies of fixed mass m is given by

$$\dot{N}_c = \int \dot{n}_c V \,\mathrm{d}V \propto \rho m \int_0^\infty f(V)/V \,\mathrm{d}V \tag{6.18}$$

where $[\rho, f(V)]$ are respectively the local number density and velocity distribution of the perturbers. Both density and velocity effects are therefore to be considered.

Figure 8(a) shows the result of a Monte Carlo simulation in which the Sun was taken to orbit through a field of brown dwarf stars ($M = 0.05 M_\odot$) with an ellipsoidal velocity distribution corresponding to an extreme Population I system. Typically one or two dozen such dark matters stars could pass through the Oort cloud every million years. In the simulation shown the vertical amplitude of the solar motion was 70 pc, the half-height of the dark material was taken to be 50 pc with an in-plane density $0.15 M_\odot$ pc^{-3}. It can be seen that the cumulative effect of the mini-showers so produced may easily yield comet flux cycles, which peak as the Sun crosses the galactic plane, with amplitudes of order 4:1.

Since the solar velocity reaches a minimum at the peak of its orbit, out-of-plane oscillations could also, in principle, be attained. The flux varies as $\langle 1/V \rangle$, and since the variation in $V_\odot$ is only ∼2 km s^{-1}, the cycle would be overlain by large stochastic perturbations if ordinary stars were the chief perturbers. For a Population I velocity distribution, with dark matter stars as above, a noisy cycle of amplitude ∼10–20% is expected; this would scarcely be detectable even in a very complete geological record.

There may, however, be circumstances in which the out-of-plane cycle is enhanced. The assumptions of a velocity ellipsoid, a 'local standard of rest', and even an 'infinite plane parallel disk', although convenient statistical descriptions of the contemporary environment, are somewhat incomplete. It has long been recognized that the solar environment is dominated by a small number of 'moving groups' (Eggen 1965). Of these, the Pleiades

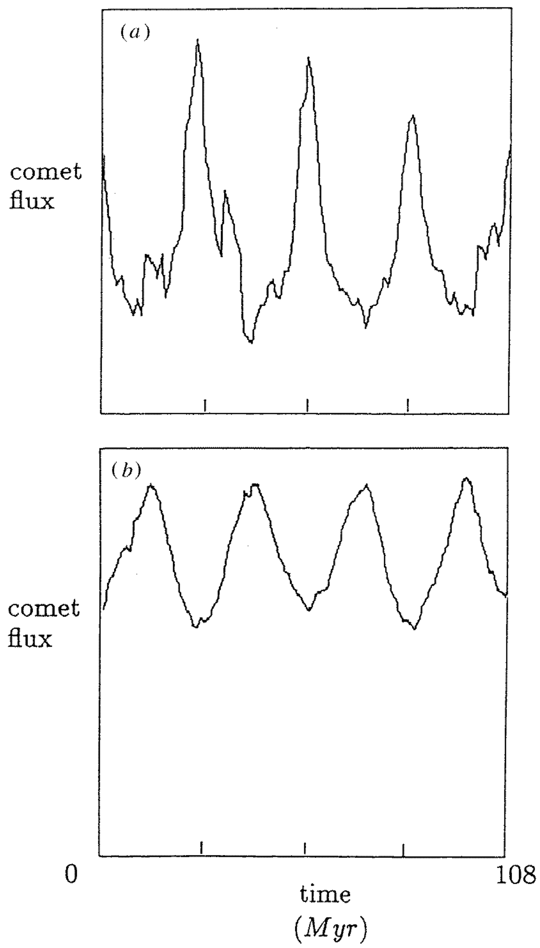

Figure 8. (*a*) The Raup/Sepkoski marine extinction record by genera. (*b*) The Rampino/Caldeira compendium of global geological events (other than the mass extinctions).

(or Gould's Belt) and Coma Berenices (or Orion Arm) groups are representative of the youngest and most massive systems, including giant molecular clouds, which pervade the solar environment within ~300 pc. Of these, the Coma Berenices group appears to represent the nearest spiral arm. The Sun is drifting through this material at a velocity (5, 5, 7) km s^{-1}, while the internal velocity dispersion of young material within the stellar associations belonging to the group is very small, $\lesssim 2$ km s^{-1}. Figure 8(*b*) shows the result of solar drift through low-mass material ('Jupiters') with these properties, from which it can be seen that the out-of-plane oscillation is now quite appreciable. The assumption here is that the abundant low-mass material penetrates a dense inner cloud, the high density of the inner cloud being required to offset the noise generated by stars acting on the outer one. There are clearly *ad hoc* elements in this scenario, which is also dependent on the existence of a dense inner cloud (whose properties are themselves strongly constrained by the cratering record: section 2.1). Thus, if cometary bombardment significantly influences terrestrial evolution, at least one, and possibly two, cycles with a galactic periodicity may reasonably be expected in the record.

With the advent of more comprehensive data in recent years, it has become possible to test the early cyclicity claims in a quantitative fashion. The Fischer/Arthur claim for a cycle in the mass extinction record was supported in a study by Raup and Sepkoski (1984), who tested for periodicity of extinction in a set of 567 extinct marine families, whose extinction dates can be resolved to within $\sim$6 Myr. They identified 12 extinction peaks as having occurred with a 26 Myr cyclicity over at least the late Permian $\sim$250 Myr ago (figure 8). Much controversy attended this claim which, however, appears to have proven robust (Sepkoski 1990, Clube and Napier 1996).

Recently, a compilation of major geological events has become available (Rampino and Caldeira 1992) which includes the Raup/Sepkoski extinction peaks as a subset, anoxic events in the ocean, major sea level changes, large discontinuities in sea-floor spreading, major mountain-building episodes and continental flood-basalt eruptions. The authors of the compilation find a strong periodicity of 26.6 Myr in these varied global events (chance probability $\sim 10^{-3}$), with a recent maximum at about 9 Myr ago. The periodicity survives an analysis, at a similar confidence level, which takes account of edge effects, secular trends, and inconsistency and bias in the standard power spectrum statistic (Clube and Napier 1996). However an examination of the dataset reveals a recent pulse (from a few Myr BP to the present) which is out of phase with the remaining peaks (figure 9). If the data of the last $\lesssim$ 8 Myr are excluded, the remaining periodicity is appreciably strengthened, and shifted in phase to $\sim$11 Myr BP (table 5). The same (P, ϕ) appears at appropriately reduced strength when the dataset is sub-divided in various ways, for example by examining the geological events independently of the biological extinctions, or by progressively truncating the data.

Although the periodicity clearly matches the vertical excursions of the Sun's vertical motion, it is conceivable that some purely internal forcing mechanism is at work with a periodicity or quasi-periodicity which is coincidentally galactic. A crucial further test for the galactic hypothesis is thus to be found in the impact cratering record. If exogenous forcing of the terrestrial cycle is indeed involved, then impact craters should also reveal the same cyclicity at some level. Seyfert and Sirkin (1979), in an introductory textbook on geology, pointed out that the ages of impact craters seem to be bunched, and proposed that impacts occur in distinct 'bombardment epochs' $\sim$26 Myr apart, the craters within each epoch showing little spread in age. They also, as a purely empirical matter, noted that the epochs seemed to coincide with major geological disturbances, and proposed a cause and effect relationship. The cratering database used by Seyfert and Sirkin was heterogeneous and no formal assessment of the confidence level was made. However, Alvarez and Muller (1984), on the basis of only 11 craters $>$ 10 km in diameter, in the age range 5–250 Myr, claimed the presence of a $\sim 28.4 \pm 1$ Myr periodicity. In fact their impact epochs were virtually identical to those of Seyfert and Sirkin, except that the latter considered the Earth to be currently immersed in an impact epoch, whereas Alvarez and Muller held that the last

Table 5. Periodicities and phases found in the terrestrial record. The formal errors, and the last decimal place in the confidence levels C, are not precisely determined.

Record	$P \pm \sigma$ Myr	$\phi \pm \sigma$ Myr	C
Mass extinctions	25.9 ± 0.5	12.1 ± 0.2	0.998
Geology	26.3 ± 0.4	10.8 ± 1.0	0.998
Field reversals	30.0 ± 4.0	13.5 ± 1.5	0.996
Impact craters	27.0 ± 0.3	11.1 ± 0.7	0.994

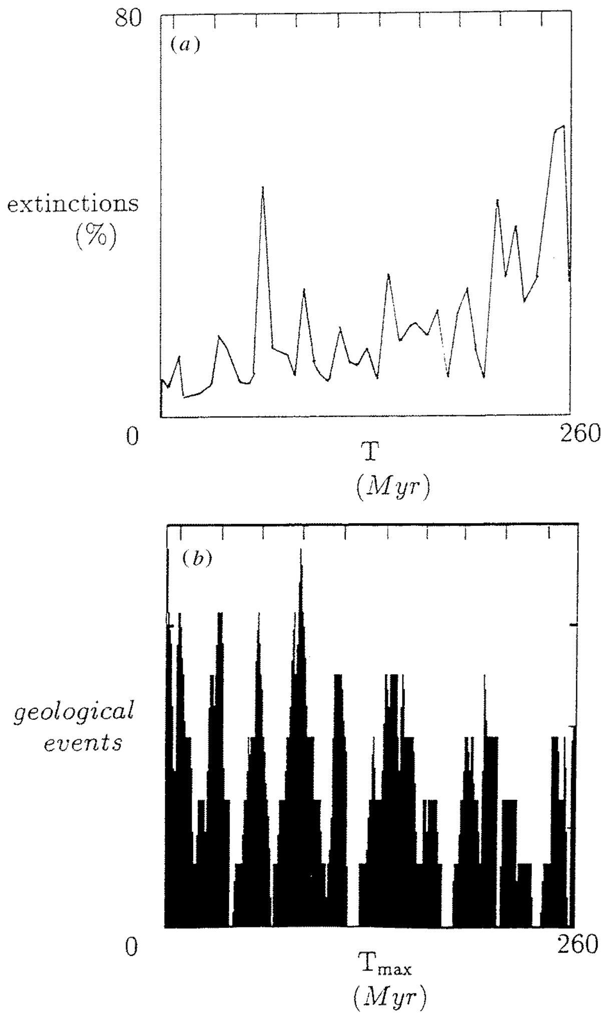

Figure 9. Simulations of comet flux variations due to encounters with assumed dark matter objects as the Sun oscillates vertically through the galactic disk. (*a*) Passage at velocity $(14, 14, 7)$ km s^{-1} through dark matter stars ($M = 0.05M_{\odot}$) with velocity dispersion $(8, 8, 6)$ km s^{-1} and scale height 50 pc. (*b*) Passage at velocity $(5, 5, 7)$ km s^{-1} through a uniform disk comprising 'Jupiters' with velocity dispersion $(2, 2, 2)$ km s^{-1}.

peak occurred $\sim 13 \pm 2$ Myr ago. Subsequent analyses have yielded a variety of claims for cratering periodicity, ranging from ~99% confidence level to none at all.

These analyses addressed the question: 'Is there a periodicity in the cratering record?' However, as a test for the galactic hypothesis, the observed terrestrial cycle should be taken as the template: the appropriate question is then: 'Is there a periodicity $P = 26.3 \pm 0.5$ Myr, with phase $\phi = 11.0 \pm 1$ Myr, in the cratering record?' There are currently 30 non-iron craters younger than 200 Myr, over 10 km in diameter, and with ages measured to formal accuracy $\sigma \leqslant 10$ Myr. Excluding craters < 10 Myr old to take account of the recent out-of-phase geological activity, a weak signal is found, which answers the first question above in the affirmative, but only at a confidence level ~95%. On the other hand the period and phase of the cratering periodicity of this weak signal closely match those of the geological data (table 5), a coincidence which is found to be significant at a confidence level $\sim$99.4%. The signal appears to be dominated by three main bombardment episodes, which took place $\sim$93, 65 and 39 Myr ago.

A continuous record of ^{3}He deposition covering the last 70 Myr has recently been obtained from a Pacific clay core (Farley 1995). The helium is of probable extraterrestrial origin, and being volatile, has most likely been delivered in the form of slowly falling cometary dust. The mass accumulation is variable, by factors of two or three, and comprises peaks at $\sim$65, 51, 38, 7 Myr BP and the present. Four of the five peaks agree with expectations; that at 50 Myr is possibly consistent with the presence of two cycles. High resolution examination of the helium deposited $\lesssim$ 0.5 Myr BP shows a clear correlation with the 100 000 yr climate cycle (Farley and Patterson 1995).

A periodicity of 26.5 Myr is at the extreme of the permissible range derived from local stellar kinematics, and implies that the galactic disk has a substantial dark matter component ($\sim 0.15 M_\odot$ pc^{-3}). This material is required to have a small scale height, and is presumably therefore baryonic in nature. The observed periodicity is thus within the range of expectations, and provides new constraints on the nature and amount of dark matter in the Galaxy. However the phase of the cycle ($\phi \sim 12$ Myr BP) is inconsistent with that of the Sun's vertical cycle, assuming the in-plane component to be the dominant one (figure 9). The expected phase of the in-plane cycle, measured from the present, is

$$\phi \sim t_c - t_i - t_d$$

where t_c represents the time of last crossing of the galactic plane, t_i is the infall time of a comet from the Oort cloud and t_d represents its diffusion time within the planetary system. Currently, the Sun is very close to the galactic plane, having crossed it, south to north, $\sim 2.5 \pm 2$ Myr ago ($t_c \sim -2.5$). With $t_i \sim 1.3 \pm 0.3$ Myr and $t_d \sim 1$ Myr, the solar system should be virtually in an epoch of maximum disturbance ($\phi \sim -0.2 \pm 2.5$ Myr). The observed phase $\sim 11.5 \pm 1.5$ Myr, on the other hand, fits rather better with that expected from the out-of-plane cycle (13.0 ± 0.7 Myr BP).

In the context of the galactic theory, there are two possibilities: either the out-of-plane (velocity) cycle is indeed the dominant one, or the in-plane (density) effect dominates but there has been a recent phase shift of $\sim$12 Myr, presumably due to the recent encounter with Gould's Belt and the Scorpio–Centaurus association. A velocity perturbation $\sim$7–10 km s^{-1} is required, which is feasible for the mass involved. This encounter may certainly be allowed since the corresponding geological record is unquestionably perturbed (figure 8). The Pliocene, for example, incorporates three of the 14 mountain-building events of the last 260 Myr, while the Pleistocene glaciations since $\sim$2.5 Myr BP plausibly mark the onset of the current galactic plane crossing. However, the expected phase shift cannot be securely predicted and its good match to a quarter cycle shift is surprising. Either the density effect requires an *ad hoc* adjustment in phase (connected to the recent encounter with a massive nebula), or the motion of the Sun is rather more securely tied to that of the nearest spiral arm and its dark matter content than is currently supposed. In either case, according to this evidence, the Earth is currently in a high-risk period.

6.3. The contemporary hazard to civilization

After the initial 'late heavy bombardment' of the Moon, which lasted $\sim$500 Myr, the lunar cratering rate settled down to a steady value, although with some evidence of fluctuations just apparent at 200 Myr resolution (figure 10). This long-term production rate of large

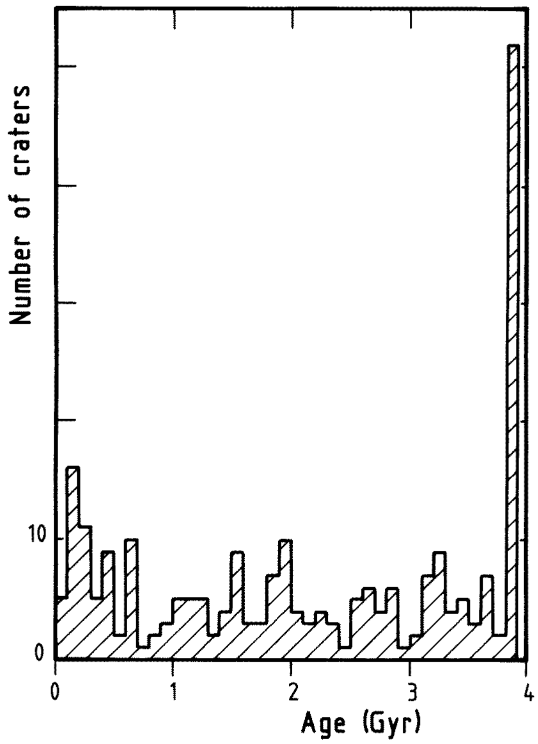

Figure 10. Long-term lunar cratering record. After Baldwin (1985).

lunar craters (> 20 km in diameter) translates to 1.0 ± 0.4 equivalent impacts/Myr on the Earth. The rate has risen to the equivalent of 2.8 ± 1.4 impacts/Myr over the last 120 Myr, in accordance with that directly inferred from the terrestrial cratering record.

Counts of small lunar craters have been used to deduce present-day impact rates (Chapman and Morrison 1994, Clube and Napier 1990), but the procedure is clearly risky as the lunar surface is over 3 Gyr old while the cometary environment appears to be highly variable on timescales much less than 0.1 Myr. Extrapolation of bright fireball data, however, yields similar results, perhaps fortuitously (Kresák 1978b), and leads to an expectation of $\sim$50 Tunguska-like impacts (energy $\gtrsim$ 10 Mt), and perhaps one or two $\gtrsim$ 1000 Mt impacts, within the timescale of civilization (Napier and Clube 1979). The evidence of fireball detection networks is that as one progresses up the fireball size scale, to the megaton energy range, there is an increasing preponderance of weak, cometary bodies (Ceplecha 1994). The extrapolated fireball numbers seem to agree with those derived from the Spacewatch project, an automated sky survey with the capability of detecting bodies down to sizes $\sim$10 m in diameter (Rabinowitz 1993, Rabinowitz *et al* 1993, Scotti 1994, Scotti and Jedicke 1996). Objects of the latter dimensions usually have exceptionally high angular speeds on detection and are followed in near real-time. Their numbers are $\sim$40 times those expected by extrapolation from the size distribution of the larger Earth-crossers, again consistently with a predominantly cometary constitution for the large fireballs (Chyba 1993); a subset of them appear to have low inclinations, almost circular orbits and perihelia close to that of the Earth. However, there still remains great uncertainty about absolute impact rates, in part because the albedos of the Spacewatch objects are uncertain.

As we have seen (section 3), however, 'spontaneous' disintegration of comets is a common phenomenon and may lead to a temporary excess of cometary over meteoritic (i.e. main belt asteroidal) material by a factor $\sim 10^2$–10^4 in the impactor range up to a few km diameter. Such disintegration is generally expected to proceed via short-lived 'asteroids' ($\sim 10^{4.5}$ yr) of cometary provenance, many of which remain in near-progenitor orbits whilst others may deviate more widely into an extensive meteoroidal complex. Small (Tunguska-sized) bodies, deriving from such hierarchic breakup, will tend to spread out along individual orbits over $\sim$200 yr, but will also result in a fairly concentrated core-stream if a significant progenitor remnant survives, even bunching in mean anomaly if they inhabit a mean-motion resonance with Jupiter (Asher and Clube 1993). Nodal intersections with individual deviant orbits and with core-streams will occur quasi-periodically on timescales 10^2–10^3 and 10^3–10^4 yr respectively. Even without orbital bunching, the highly non-Poissonian nature of the intersection probabilities may lead to surprisingly large differences in close-encounter rates over a given timespan (say 300 yr) and corresponding uncertainty in the underlying populations. In cases where the hierarchic breakup is itself catastrophic (e.g. due to impacts on relatively fragile asteroids) and the disintegration products are rather more widely dispersed, we anticipate a bombardment on a shorter timescale (say $\sim$20 yr) due to a relatively broader stream. In short, rather than purely sporadic bombardment by individual objects, the impact hazard is very likely to be structured; but both the mean and the higher moments of this temporal structure are largely unknown.

Atmospheric protection probably holds up to the 10 megaton (50 m) energy range, beyond which airbursts may generate sufficient overpressure at ground level to cause severe damage (Chyba 1993, Hills and Goda 1993). Bodies of over 200 m diameter probably reach ground level; in the case of deep ocean impact the energy spreads out in two dimensions and the wave height declines only linearly with distance. A 200 m body would create a wave 15 m high 1000 km from impact. On reaching shallow water the wave amplitude may increase by a factor $\sim$40 with duration of passage about two minutes: Hills and Goda (1993) point out that such an impact anywhere in the North Atlantic would overrun all low-lying areas (Denmark, Holland, Manhattan etc) bordering the ocean, while Yabushita and Hatta (1994) estimate that a Pacific impact of a 200 m body will occur with $\sim$1% probability in the next century and, if it does, will destroy most of the artificial constructions around the Pacific Rim.

Traumatic though such impacts would be, it seems that on timescales relevant to civilization the climatic hazard due to stratospheric dusting is of even greater significance. Adopting standard rates, it appears that the injection of giant comets from chaotic, trans-Saturnian orbits into stable, Earth-crossing ones is $\sim$10 Myr^{-1}; and when such a comet disintegrates, the Earth gathers up a stratospheric dust veil of optical depth $0.05 \lesssim \tau \lesssim 3$ during the 10^3–10^4 lifetime of the comet (Hoyle 1984, Clube and Napier 1984a, Bailey *et al* 1994). Such an event would probably lead to a full-blown ice age (*loc cit*, Clube *et al* 1996), which would presumably bring civilization to an end over most of the globe. There may, however, be a continuum of lesser climatic events, significant on millennial timescales, and associated either with periodic immersions in the dense IRAS dust trails of short-period comets or with one or more super-Tunguska impacts. As we have seen, a good correlation seems to have held over the past $\sim$Myr between comet dust deposition and the known climatic oscillation of $\sim 10^5$ yr. To assess these 'lesser' events, limited comparisons may be made with nuclear winter studies.

The earliest nuclear winter models (Turco *et al* 1983) were one-dimensional, assuming a hemispheric distribution of smoke and dust, and neglecting feedback effects from cryosphere and oceans. Typically in such runs the initial optical depth is $\tau \sim 4$, of which $\tau \sim 3$ is

Table 6. Maximum growth season temperatures below which various agricultural crops cannot survive. From Ehrlich *et al* (1983).

Crop	Summer temperature
Winter wheat	5 °C
Rice	13 °C
Sorghum	13 °C
Corn	10 °C

due to smoke injected into the troposphere, while fine dust ($\lesssim 10\ \mu$m radius) reaching the stratosphere has $\tau \sim 1$. The overall optical depth declines to $\sim$2 after three months. In these models temperature variations $\Delta T \sim -40\,^\circ$C are reached within about twenty days of the dust injection, and the recovery time of the climate is over a year. Second generation models incorporated snow and ice feedbacks, and took account of the thermal inertia of the oceans, which moderated the land temperature response. Robock (1984) assumed the injection of smoke in a uniform latitude band between 30° and 70° N, and distributed between 1 and 10 km in altitude. Incorporating seasonal cryospheric interactions, he found that land temperatures were $\sim$6 °C colder than normal in the second year after smoke injection. Covey *et al* (1987) modelled variations in sea ice and sea surface temperature, and found that maximum northern hemisphere coolings after 20 days were typically $\sim$25 °C in July, $\sim$15 °C in April and $\sim$5 °C in January; thus, the ocean moderates the cooling by a factor of about two. A third generation of models (e.g. Schneider and Thompson 1987) has moderated these predictions by a further factor of two, yielding $\Delta T \sim 10\,^\circ$C for the baseline July case. This results from the incorporation of two additional factors: (i) the infrared opacity of the smoke allows for a compensatory 'greenhouse effect'; and (ii) the smoke distribution is patchy, allowing sunlight to penetrate from time to time. Thus, these models point to a 'nuclear fall' rather than 'nuclear winter'.

However, there are indications that a 'cosmic winter' might be considerably more severe than a nuclear fall, since the 'third generation' factors do not apply to the chronic climatic effects of prolonged cometary dusting. Thus, whereas in the nuclear winter scenarios tropospheric dust is the predominant absorber of sunlight, and is removed in some weeks, in the 'cosmic winter' situation tropospheric dust is relatively unimportant, the scattering of sunlight being primarily due to sub-micron Brownlee particles in the stratosphere, with a residence time of a year or more. Further, the patchiness simulated in the later models does not occur with cometary dusting. Additionally, dust injection is global rather than hemispheric, and there is no moderation of climatic severity from warm ocean currents entering from the southern to northern hemispheres. Thus, at the qualitative level it appears that the prompt effects of a cosmic winter are more closely simulated by the first or second generation nuclear winter models rather than by the later 'nuclear fall' ones.

The likely biological effects of a first-generation-type nuclear winter have been examined by a number of authors (e.g. Ehrlich *et al* 1983). Table 6 lists the critical temperatures T_c below which various crops are killed during active summer growth. Exposure to temperatures $\lesssim T_c$ for even a few days, during the growth season, may be sufficient to destroy these crops. A temperature drop $\gtrsim 3\,^\circ$C in the growing season (July in the northern hemisphere) is thus more than adequate to bring commercial agriculture to an end. Other factors discussed by Ehrlich *et al* (1983) include the disruption of man-made energy subsidies necessary for major crops and the effect of reduced continental precipitation. Since crop productivity is linearly proportional to the degree of exposure to sunlight, many plants would disappear by virtue of reduced light alone.

Tropical terrestrial ecosystems are less adapted to prolonged darkness or cold, and Ehrlich *et al* (1983) consider that most forests and terrestrial species would disappear. They consider that 'the possibility of the extinction of Homo Sapiens cannot be excluded'.

The effects decribed above are actually more acute rather than chronic; however the latter are likely to have the most far-reaching consequences in biological terms. Unfortunately the chronic effects are dependent on ill-understood feedbacks and have hardly been modelled. A decline of 5 °C in solar insolation for a year or more, which is within the parameter range under discussion, might cause a drastic reduction in monsoon rainfall, with a catastrophic effect on much of the world's agriculture (Barnett *et al* 1988).

Heavy snowfalls are expected over the continents in the course of a stratospheric dusting event (or a nuclear winter). In the case of instantaneous dust injection the period of severe cooling ($\sim$1 yr) is likely to be too short to overcome the thermal inertia of the oceans, which have heat content equivalent to $\sim$3–5 yr of sunlight. However, with more prolonged dusting, exceeding the oceanic cooling time, the oceans will no longer act as a heat buffer. In these circumstances heat transported to the upper atmosphere may be reduced to the point where high albedo ice crystals form in the stratosphere, further reducing insolation. Once in this self-sustaining state, the Earth will slide into an ice age. This mechanism is discussed in more detail by Hoyle (1996) and Clube *et al* (1996).

The dust input described above derives from prolonged meteoric and meteoroidal input, the larger bodies disintegrating on atmospheric entry to $\sim\mu$m-sized dust particles. However, a significant prompt injection of dust from much larger (super-Tunguska) bodies is also expected, at characteristic intervals of some centuries. The dust veils from such events may initially be hemispheric, but could still yield measurable climatic and agricultural effects over some years. The AD 536 dust-veil ($\tau \sim 2.5$) might be of this character as it appears not to be associated with a volcanic acidity signal. A causal relationship has been proposed between this dust veil and the severe cold and famine at this time, the latter documented throughout the Old World and presaging the arrival of the Justinian Plague (Baillie 1994). Both the AD 536 event and the 17th century Little Ice Age are correlated with known surges in the Taurid meteoroid flux, the likely source of the 1908 Tunguska object. The likely climatic perturbations of celestial origin are thus no less significant than than the correlated impacts discussed in section 3.

Table 7 summarizes the characteristic timescales associated with the various hazards to civilization. It should be stressed that large uncertainties, both in statistics and modelling, are associated with all of these figures.

Table 7. Types of hazard from near-Earth environment. Numbers are intended only as a rough guide.

Nature of hazard impact energy (Mt)	Integrated duration of trauma (yr)	Recurrence time (yr)
Stray Tunguska	1–10	50–250
Tunguska swarm	1–10	200–250
Super-Tunguska swarm	1–100	200–2500
Giant comet dusting	1000–3000	10 000–100 000
1 km asteroid	1–100	100 000–500 000

7. The origin of comets

Many ideas about the origin of comets have been put forward over the past 200 or more years (a comprehensive review of these is given by Bailey *et al* 1990). Generally, mainstream opinion has oscillated between ‘interstellar’ and ‘solar system’ cosmogonies, although with numerous variants of each, and with one view or the other dominant for $\sim$50 yr or so (about the active lifetime of an active astronomer!). A solar system cosmogony has been generally accepted since Oort (1950) proposed that an exploded planet might have given rise to both the main belt asteroids and the system of comets (cf Ovenden 1972, Van Flandern 1993). A more popular assumption, after the formulation of the Oort cloud hypothesis, was that comets had grown in the Jupiter–Saturn region of the ‘solar nebula’ before being ejected by these planets into the fringes of the solar system and beyond. Later, the canonical scenario became that comets had grown and were ejected from the Uranus–Neptune region; and with the discovery of the EK belt, the prospect that comets were formed in a disk extending far beyond the planets, possibly out to 1000 au, has been widely discussed. Thus, as knowledge has accumulated, there has been a tendency to consider sites ever more remote from the Sun. Underlying this trend is the growing realization that a very cold ‘aggregation’ environment is involved, similar in both composition and temperature to that of the interstellar medium.

Indeed, for all elements heavier than He, the bulk composition of P/Halley more closely resembles that of the Sun and the interstellar medium than even the most unprocessed meteorites: this implies a relatively cold, but otherwise not strongly constrained, formation site. Organic compounds (such as hydrogen cyanide, methyl alcohol and the CHON particles) are abundant in P/Halley and could not have been formed by the successive condensation of increasingly less refractory material in an initially hot, *slowly* cooling solar nebula (Festou *et al* 1996). Nevertheless the same simple organic molecules are observed in cold molecular clouds, and there is thus good agreement with the hypothesis that comet nuclei are pristine aggregates of the icy, sub-micron, organic-rich interstellar grains found in such nebulae (e.g. Greenberg 1988, Greenberg and d’Hendecourt 1985). The millimetre-wave detection of H_2S in comets Austin 1990 and Levy 1990 has been interpreted to indicate a nebular sublimation temperature $\lesssim$ 60 K. This is consistent with formation in a warm molecular cloud; in the context of conventional solar nebula models, the comet factory would have lain in the Uranus region or beyond.

The detection of ethane (C_2H_6) and methane (CH_4), as well as carbon monoxide and water, in Comet C/1996 B2 Hyakutake has imposed new constraints (Mumma *et al* 1996). The ethane and methane were present in comparable abundances (4% and 7.5% of the water abundance respectively), whereas a solar nebula in thermochemical equilibrium would have yielded an ethane/methane ratio $\lesssim 10^{-3}$. Ethane, however, is readily produced on grain surfaces in the cold, dense cores of molecular clouds, either by photolysis of ice or from the addition of hydrogen to acetylene (C_2H_2), the latter being readily generated in the nebular gas (Mumma *et al* 1996). Possibly, both ultraviolet processing and H-atom addition took place simultaneously. The high ethane abundance could then still have arisen through photolysis in a solar-irradiated inner zone of the solar nebula, say within the orbit of Jupiter, and indeed a spatially variable UV flux could give rise to a diversity of comet chemistries, with methanol-poor comets for example originating in the Jupiter region. It is not clear, however, that the fragile, fluffy nature of comets is consistent with growth in such a robust, probably turbulent region: coma dust may have packing fractions $\lesssim 0.07$ and fireballs of cometary origin may have densities $\lesssim 0.2$ g cm^{-3}; and Donn (1976) has pointed out that collisions between grains at $\gtrsim 0.1$ km s^{-1} would evaporate their ices. The abundance of hydrogen isocyanide (HNC) detected in Comet Hyakutake is very similar to that in quiescent molecular clouds, and is again inconsistent with the equilibrium value expected to pertain in the outer solar nebula whence comets are widely assumed to have originated (Irvine *et al* 1996); however, *ad hoc* possibilities can again be envisaged which might 'fix' this discrepancy. Thus, with possible formation sites ranging from solar nebula to molecular cloud, the chemistry of comets has not so far strongly constrained their formation environments.

Nor can it yet be excluded that comet formation involved before the high temperature phase. The fundamentally ordered layering of the refractories, organics and volatile ices within the grains which apparently make up cometary condensates, and the evidence for vapour phase growth within them (Bradley *et al* 1983), are consistent with a presolar, reducing environment which was initially hot and underwent very rapid cooling. The mean $^{12}C/^{13}C$ isotope ratio in P/Halley dust particles $\sim$100, closer to the solar system value of 89 than the local interstellar one of $\sim$65. Individual grains show a huge scatter, however, from $\sim$0.5 to $\sim$3300 (Šolc *et al* 1987) and there is diversity in the relative abundances of simple organics from one comet to another: thus, it is possible that comets have originated from a variety of condensation environments reflecting a wide range of initial masses. Whether a contemporaneous or primordial interstellar origin is implied for Oort cloud comets probably cannot be settled from the chemistry alone. Indeed while it seems natural to asssociate the hot plasma with the compression lanes which frequent galactic disks, we cannot yet tell whether comets would originate from spiral density waves (section 7.2) or from such spiral ejecta as apparently emanate from galactic nuclei.

7.1. Formation in a protoplanetary nebula

In contrast to the foregoing, which allows for possible comet formation in advance of the Sun, the most popular, and most highly developed hypothesis for comet cosmogony, is that comets and planets formed together within a nebular disk which encircles the pre-existing Sun. Observational evidence, in particular direct imaging with the Hubble Space Telescope, has confirmed that young stars are indeed often surrounded by disks, although the latter are much more active, massive and extensive than usually envisaged in the models. Whereas in most theoretical studies the hypothetical disk is of relatively small mass ($\sim 0.01 M_{\odot}$) and radius ($\sim$50 au), just sufficient to encompass the formation of the planets by the aggregation of grains within it, the observed disks are typically $\sim$2000 au in radius. A typical example is IRAS 1629A, in the Rho Ophiuchi molecular cloud, which has luminosity $\sim 23 L_{\odot}$, temperature $\sim$40 K and mass $\sim 0.24 M_{\odot}$. The material surrounding the source out to $\sim$3000 au appears to be infalling at $\sim$1 km s^{-1} and to have a microturbulent velocity $\sim$0.4 km s^{-1} (Wolstencroft and Walker 1988). Strong bipolar outflows ($\sim$100 km s^{-1}) are characteristic of this and other protostellar and young stellar systems. A number of IRAS sources are associated with dwarf and subgiant stars, and it appears that excess IR emitters, modelled as dusty disks, are very common amongst F,G and K main sequence dwarves.

Within the disk, according to this scenario, grains accumulated first into cometary ($\sim$km-sized) and then planetary bodies (Bailey 1994). To give this primordial nebula a solar composition and yield the planetary masses, the initial mass had to be 0.01–0.02$M_{\odot}$ prior to the loss of volatiles (Hoyle 1960). On this picture the outer planets are regarded as cometary aggregates, at least in part, and so comets are envisaged as primitive building blocks whose chemistry may yield unique insights into the origin of the solar system. It is envisaged that within $\sim 10^8$ yr the primordial comets so formed would have been scattered by the growing planets, to yield the comet cloud now observed.

The required initial mass of the Oort cloud may be estimated by combining three factors: the current cloud population, the mean mass of comets within it, and the emplacement and survival probabilities assuming an origin contemporaneous with that of the planets. We adopt a conventional Oort cloud with an inner edge corresponding to $a_o = 4000$ au, an outer radius $a_t = 33\,000$ au, and a power law index $\gamma = 0$ corresponding to a flat energy distribution (section 2). Then the current cloud contains 5.7×10^{11} comets brighter than absolute magnitude $H_0 = 7$. The nuclear mass of a comet is a very uncertain quantity, in part because the measured brightness of a cometary nucleus will, even at large distances from the Sun, often be contaminated in greater or lesser degree by emission from the surrounding coma. We adopt here the mass-luminosity relation (Bailey 1994):

$$\bar{M} = 4.6 \times 10^{19} (D_{max}/300)^{1.125} \text{ g} \tag{7.19}$$

where D_{max} represents the maximum planetesimal diameter. The emplacement and survival probabilities are respectively $\lesssim 20\%$ for ejection from the Uranus–Neptune region, and $\lesssim 5\%$ for comets with initial $a \gtrsim 10\,000$ au (Bailey *et al* 1990).

Adopting the above with $D_{max} = 300$ km, the current Oort cloud mass turns out to be $\sim 0.03 M_\odot$ or over 20 times the mass of the planetary system (Mendis and Marconi 1986, Marochnik *et al* 1988). To account for the currently observed cloud, allowing for molecular cloud perturbations over the past 4.5 Gyr, an uncertain but very substantial initial mass (of order $10^4 M_\oplus$) must have been ejected from the region of the outer planets. Assuming a metallicity $Z = 0.02 M_\odot$ and 100% efficiency of dust aggregation, the required mass of the progenitor nebula, with hydrogen and helium in place, is $\sim 1.5 M_\odot$; the presence of a dense inner cloud might increase this estimate by one or more powers of 10. Thus, one has to envisage comets segregating out of a stellar mass, within say 1000 au of the early Sun. It is not clear how such a large mass of gas could have evaporated while leaving its comets bound (it would seem more likely to have formed a binary companion of the Sun).

Even allowing for the range of uncertainties of the various factors, it thus appears that the upward revision in recent years of cometary masses, and the further realization that the Oort cloud is unstable in the galactic environment, have put a considerable strain on these models. This suggests that comet formation might have occurred in a massive protoplanetary disk, extending to 1000–3000 au rather than 30–50 au, in better accordance with the disks observed around young stars (e.g. Biermann and Michel 1978). Bailey (1994) has argued that a cold, slowly rotating molecular cloud of radius 0.1–1 pc, mass 1–2$M_\odot$ and temperature ~ 10 K, and threaded by a weak interstellar magnetic field, could collapse into an extended, relatively massive Keplerian disk of radius ~ 1000 au, with a surface density profile $r^{-3/2}$ as required by the normalization to planetary masses. The resulting protoplanetary disk has a few hundred $M_\oplus$ of dust and ice available for the creation of planets and comets. Radial drift, which would presumably be necessary for the creation of a central protostellar condensation, was however neglected in this model. Much of this hypothetical formation region is beyond the reach of significant planetary, stellar and molecular cloud perturbations, and so could not act as a reservoir for the observed cometary systems.

7.2. Formation in molecular clouds

Given that stars form in molecular cloud environments, the question arises whether comets might do likewise even preceding the formation of stars. In that case, one envisages molecular clouds as being seeded with a population of comets or proto-comets, possibly within a hierarchy of aggregates ranging from protoplanetary to star cluster masses. The present-day solar system comets might then have originated within the same molecular cloud or star cluster as the Sun, but otherwise be unconnected with it (Donn 1976), or they might be captured as the Sun wends it way through the Galaxy, penetrating comet-rich molecular clouds from time to time (Clube and Napier 1984b). With no need to eject the original Oort cloud outwards from the protoplanetary regions, the problems associated with this latter mechanism do not arise.

There exist a number of instabilities in the interstellar medium, and one might enquire whether any of these might yield the collapse of grains into comet-sized bodies on a reasonable timescale. Yabushita (1983) proposed that comets might form in the interstellar environment by simple gravitational sedimentation of dust, followed by coagulation. For this to work an extremely quiescent environment is required, and this may exist within Bok globules. These are concentrations of dust a few parsecs in radius and masses up to $50M_\odot$ which may be collapsing. However, the coagulation time, even in quiescent conditions, is 10^7–10^8 yr, and this probably exceeds the collapse time of a globule or its lifetime against disruption by supernova shock waves. It therefore does not seem likely that, in the interstellar environment, straightforward gravitational collapse of dust into planetesimals is a widespread process.

However, other forcing mechanisms may exist. Spiral structure associated with star-forming regions in galaxies may be indicative of an intermittent though largely coherent, rapidly cooled, flow from their nuclear regions, in which case the necessary quiescent conditions during confinement are probably imposed by the ambient galactic medium (cf Clube 1988). Alternatively, one might envisage the driving of grains by radiation pressure, either in the neighbourhood of hot, young stars (Bailey 1987), or in a wider molecular cloud environment (Napier and Humphries 1986, Napier 1990). Radiation pressure may be a much more powerful driving mechanism than gravity, and it has long been recognized that starlight may provide an effective mechanism for segregating gas and dust. Flannery and Krook (1978) pointed out that dust grains, shadowing each other in an ambient radiation field, would tend to be forced together by the differential radiation pressure so created. This effect may be the source of instabilities within nebulae since a small perturbation in dust density might attenuate the ambient radiation field, forcing grains together and so increasing the attenuation and amplifying the effect. Gerola and Schwartz (1976), following an earlier paper by Reddish (1971), pointed out that near-ultraviolet photons might eject adsorbed molecules from interstellar grain surfaces; the momentum transferred to the grain by this 'rocket effect' might be up to two powers of 10 greater than that due to radiation pressure alone. An anisotropic radiation field could thus effectively separate dust and gas. The condition for marginal instability can be written as

$$2K_E + \Omega + \Psi = 0 \tag{7.20}$$

where

$$\Omega = \int_V \rho_g \boldsymbol{r} \cdot \boldsymbol{g} \, \mathrm{d}V \tag{7.21}$$

represents the gravitational potential, and

$$\Psi = \int_V \rho_g \boldsymbol{r} \cdot \boldsymbol{f} \, \mathrm{d}V \tag{7.22}$$

represents the radiative one. Here $(\boldsymbol{f}, \boldsymbol{g})$ represent the accelerations acting on a dust grain due to radiation pressure and gravity respectively, while ρ_g represents the density of an individual grain.

Two characteristic Jeans' masses are imbedded in these equations, one a stellar mass associated with gas and gravitational instability, the other a cometary mass associated with dust and radiative instability. Unlike the stellar case, the collapse of dust may be rapid, especially in the presence of a UV radiation field. Laboratory experiments simulating grains in a cold interstellar environment have revealed that photodesorption of water ice is very efficient ($\sim 8 \times 10^{-3}$) following an incubation period of UV irradiation (Westley *et al* 1995), and could be an important mechanism in dense molecular clouds ($n_H \sim 10^3$–10^5 cm^{-3}). With this efficiency, comets could collapse out of a cooling flow or a quiescent, irradiated molecular cloud in as little as a thousand years (Napier 1990).

7.3. Interstellar comets

McGlynn and Chapman (1989), using only simple arguments and widely accepted hypotheses about star and planet formation, arrived at an estimate for the detection rate of interstellar comets which, as they put it, 'is becoming an embarrassment to the theories of solar system and cometary formation'. We present here a modified version of their discussion.

Safronov (1969), assuming that comets were initially formed in the region of the giant planets and subsequently ejected, found that the initial Oort cloud was produced with only about 1% efficiency, the remaining comets formed in the region being thrown into interstellar space. Later models incorporating diffusive perturbations, inwards transfer of orbits and the like, have increased the overall formation efficiency of the Oort cloud to around 10% (e.g. Fernández 1978) or even higher (Fernández, personal communication), with the remaining comets about equally distributed between destruction and ejection. In addition giant molecular clouds have probably depleted the Oort cloud by an order of magnitude over the age of the solar system. If one now assumes that the initial population of the primordial cloud was not less than the present-day one (5×10^{11} comets brighter than $H_o = 16$, to within a factor of a few), then something like 3×10^{13} comets may have been ejected from the solar system due to the above processes.

Now assume that comet formation is a normal adjunct of star formation and that the solar system is typical in this respect (at the time of writing, eight planetary systems have been discovered in the solar neighbourhood in spite of the difficulties of detection). Then allowing for a largely undiscovered white dwarf population in the galactic disk, yielding a local stellar number density $\nu_c \sim 0.1$ pc^{-3}, the number density of interstellar comets is found to be $\nu_c \sim 4 \times 10^{-4}$ au^{-3}, to within a factor of a few. This is comparable with the number density of Oort cloud comets. Taking the one-dimensional velocity dispersion of interstellar comets to be that of the nearby G0 stars (21 km s^{-1}), and the solar motion relative to the local standard of rest as 16.5 km s^{-1} one finds that an interstellar comet should pass within 2 au of the Sun about once every five years, with a mean speed of just under 40 km s^{-1}. No such comet has been detected, although sky coverage has now been maintained by comet hunters for about 150 yr.

The detection probability of long-period comets over roughly this period has been estimated by Everhart (1967), who considered that about 8000 comets passed within 4 au of the Sun, of which 256 were discovered. Restricting the sample to $q < 2$, when comets outgas, McGlynn and Chapman (1989) thus deduce a detection probability $\sim$0.07 over this period. There is therefore an expectation that one or two interstellar comets with strongly hyperbolic orbits might by now have been detected. Although the constraints are already tight, the absence of any such detections is not yet a critical matter, given the uncertainties. However interstellar comets, if they exist, will be detectable with the greatly enhanced deep sky coverage which projected near-Earth-object search programs will provide: some of these projects are geared towards the discovery of up to 30 bodies/night with diameters $\gtrsim$ 0.5 km. It should therefore be clear within the next few years whether these comets exist in appreciable numbers, or whether some fundamental revision of widely held beliefs is called for.

8. Discussion and conclusions

The current revolution in the Earth sciences, triggered and sustained by that in solar system exploration, is still being worked through. It is now widely recognized that the Earth is a bombarded planet, and that the traumatic effects of large impacts may have had profound consequences for the evolution of life. Beyond this broad consensus, however (and still not all paleontologists accept that impacts have had a significant role in the evolution of life), there is considerable uncertainty. At one extreme the world is perceived as a basically uniformitarian stage, on which the drama of traumatic impacts is occasionally enacted (Alvarez *et al* 1980); at the other it is seen as essentially under the detailed control of solar system and galactic inputs (Clube 1978, Napier and Clube 1979). The differences between these two basic paradigms reflect in part differing assessments of the rôle played by structure in the cometary environment, and by cometary dust in the control of climate: empirical evidence from both interplanetary and geological studies may be expected to yield further new insights into these issues over the next few years. In addition, though, they can imply differing assessments of the role of cosmic (bio)chemistry in biological evolution: essentially none on the one hand yet potentially massive on the other (cf Hoyle and Wickramasinghe 1985). But while cometary science and biological evolution have thus become inextricably linked to the issues of the past *and* current celestial hazard, it could also, as we have seen, ultimately lead to new insights into such fundamental issues as star and spiral arm formation.

Acknowledgments

The authors are especially grateful to Mark Bailey, Julio Fernández, Mike Mumma and Jim Scotti for various discussions, communications, critical readings and comments on different aspects of this review. Nathan Harris and Scott Manley kindly provided us with results in advance of publication. Part of this work was supported by the USAF (EOARD SPC-93-4706) through the good offices of Stu Nozette and Pete Worden. WMN also wishes to acknowledge the hospitality of the Physics Department of Oxford University, where a substantial part of this review was written.

References

Adushkin V V and Nemchinov I V 1994 *Hazards due to Comets and Asteroids* ed T Gehrels (Tucson, AR: University of Arizona) p 721
Alvarez L W, Alvarez W, Asaro F and Michel H V 1980 *Science* **208** 1095
Alvarez W and Muller R A 1984 *Nature* **308** 718
Asher D J and Clube S V M 1993 *Quantum J. R. Astron. Soc.* **34** 481
Asher D J and Steel D I 1993 *Mon. Not. R. Astron. Soc.* **263** 179
Asher D J, Clube S V M and Steel D I 1993 *Mon. Not. R. Astron. Soc.* **264** 93
Asher D J, Bailey, M E, Hahn, G and Steel, D I 1994 *Mon. Not. R. Astron. Soc.* **267** 26
Asphaug E and Benz W 1996 *Nature* **370** 120
Bahcall J N and Bahcall S 1985 *Nature* **316** 706

Bailey M E 1983 *Mon. Not. R. Astron. Soc.* **204** 603
——1986 *Mon. Not. R. Astron. Soc.* **218** 1
——1987 *Icarus* **69** 70
——1991 *Molecular Clouds* ed R A James and T Millar (Cambridge: Cambrdge University Press) p 273
——1994 *IAU Symp. 160: Asteroids, Comets, Meteors 1993* ed A Milani *et al* (Dordrecht: Kluwer) p 443
Bailey M E, Chambers J E and Hahn G 1992 *Astron. Astrophys.* **257** 315
Bailey M E, Clube S V M, Hahn G, Napier W M and Valsecchi G 1994 *Hazards due to Comets and Asteroids* (Tucson, AR: University of Arizona) p 479
Bailey M E, Clube S V M and Napier W M 1990 *The Origin of Comets* (Oxford: Pergamon)
Bailey M E and Emel'yanenko V V 1996 *Mon. Not. R. Astron. Soc.* **278** 1087
Bailey M E and Stagg C R 1988 *Mon. Not. R. Astron. Soc.* **235** 1
——1990 *Icarus* **86** 2
Baillie M G L 1994 *The Holocene* **4** 212
Baldwin R B 1985 *Icarus* **61** 63
Barnett T P, Dumenill L, Schlese U and Roeckner E 1988 *Science* **239** 504
Bertiau F C 1956 *Astrophys. J.* **128** 533
Biermann L 1978 *Astronomical Papers Dedicated to Bengt Strömgren* (Copenhagen: Copenhagen University Press) p 327
Biermann L and Michel K W 1978 *Moon and Planets* **18** 447
Blaauw A 1991 In *NATO Advanced Study Institute on Physics of Star Formation and Early Stellar Evolution* ed N Kylafis and Ch J Lada (Dordrecht: Kluwer) p 125
Blitz L and Shu F 1980 *Astrophys. J.* **238** 148
Bond G and Lotti R 1995 *Science* **267** 27
Borovicka J and Spurny P 1996 *Icarus* **121** 484
Bradley J P, Brownlee D E and Veblen D R 1983 *Nature* **301** 473
Byl J 1986 *Earth, Moon, Planets* **36** 263
Carusi A and Valsecchi G 1987 *Interplanetary Matter* ed Z Ceplecha and P Pecina, p 21
Ceplecha Z 1992 *Astron. Astrophys.* **263** 361
——1994 *Astron. Astrophys.* **286** 967
Chapman C R and Morrison D 1994 *Nature* **367** 33
Chen J and Jewitt D 1994 *Icarus* **108** 265
Chyba C F 1993 *Nature* **363** 701
Chyba C, Thomas, P and Zahnle K 1993 *Nature* **361** 40
Clube S V M 1978 *Vistas Astron.* **22** 77
——1988 *Dust in the Universe* ed M E Bailey and D A Williams (Cambridge: Cambridge University Press) p 331
——1995 *Vistas Astron.* **39** 673
Clube S V M and Napier W M 1982 *Quantum J. R. Astron. Soc.* **23** 45
——1984a *Mon. Not. R. Astron. Soc.* **211** 953
——1984b *Mon. Not. R. Astron. Soc.* **208** 575
——1986 *The Galaxy and the Solar System* ed R Smoluchowski, J N Bahcall and M S Matthews (Tucson, AR: University of Arizona) p 260
——1990 *The Cosmic Winter* (Oxford: Blackwell)
——1996 *Quantum J. R. Astron. Soc.* in press
Clube S V M, Hoyle F, Napier W M and Wickramasinghe N C 1996 submitted
Covey C, Schneider S H and Thompson S L 1987 *Nature* **308** 21
Davis D R and Farinella P 1996 *Icarus* in press
De Laubenfels M W 1956 *J. Paleont.* **30** 207
Delsemme A H and Patmiou M 1986 *20th ESLAB Symp. on the Exploration of Halley's Comet* ESA SP-250 vol II (Noordwijk: ESA Publications, ESTEC) p 409
Donn B 1976 *The Study of Comets: Part 2 (IAU Coll. 25)* (Washington, DC: NASA SP-393) p 663
Dorman J, Evans S, Nakamura Y and Latham G 1978 *Proc. Lunar Planet. Sci. Conf.* **9** 3615
Duncan M, Quinn T and Tremaine S 1988 *Astrophys. J. Lett.* **328** L69
Duncan M J, Levison H F and Budd S M 1995 *Astron. J.* **110** 3073
Edgeworth K E 1943 *J. Brit. Astron. Assoc.* **53** 181
——1949 *Mon. Not. R. Astron. Soc.* **109** 600
Eggen O J 1965 *Stars and Stellar Systems* vol V (Chicago, IL: University of Chicago) p 111
Ehrlich P R *et al* 1983 *Science* **222** 1293
Emel'yanenko V V and Bailey M E 1996 in press

Epstein S and Krishnamurphy R V 1990 *Phil. Trans. R. Soc.* A **330** 427
Everhart E 1967 *Astron. J.* **72** 1002
——1972 *Astrophys. Lett.* **10** 131
Farinella P and Davis D R 1996 *Science* **273** 938
Farley K A 1995 *Nature* **376** 153
Farley K A and Patterson D B 1995 *Nature* **378** 600
Fernandez J A 1978 *Icarus* **34** 173
——1992 *Chaos, Resonance and Collective Phenomena in the Solar System (IAU Symp. 152)* (Dordrecht: Kluwer) p 239
——1982 *Astron. J.* **87** 1318
Fernandez J A and Gallardo T 1994 *Astron. Astrophys.* **281** 911
Fernandez J A and Ip W-H 1991 *Comets in the Post-Halley Era* ed R L Newburn *et al* p 487
Festou M C, Rickman H and West R M 1996 *Astron. Astrophys. Rev.* in press
Fischer A G and Arthur M A 1977 *Soc. Econ. Paleont. Mineral. Spec. Publ.* **25** 19
Flannery B P and Krook M 1978 *Astrophys. J.* **223** 447
Fulle M 1992 *Nature* **359** 42
Gallant R L 1964 *Bombarded Earth* (London: John Baker)
Gerola H and Schwartz R A 1976 *Astrophys. J.* **206** 452
Gilmore G and Wyse R F G 1987 *The Galaxy (NATO Advanced Science Institute, Series C 207)* (Dordrecht: Reidel) p 247
Gladman B and Duncan M 1990 *Astron. J.* **100** 1680
Greenberg M J 1988 *Dust in the Universe* ed M E Bailey and D A Williams (Cambridge: Cambridge University Press) p 121
Greenberg M J and d'Hendecourt L B 1985 *Ices in the Solar System* ed J Klinger, D Benest, A Dollfus and R Smoluchowski (Dordrecht: Reidel) p 185
Hahn G and Bailey M E 1990 *Nature* **348** 132
Hamid S E, Marsden B G and Whipple F L 1968 *Astron. J.* **73** 727
Hartung J 1993 *Icarus* **104** 280
Heisler J and Tremaine S D 1986 *Icarus* **65** 13
Hills J G 1981 *Astron. J.* **86** 1730
Hills J G and Goda M P 1993 *Astron. J* **105** 1114
Hogg D W, Quinlan G D and Tremaine S 1991 *Astron. J.* **101** 2274
Holman M J and Wisdom J 1993 *Astron. J.* **105** 1987
Holmes A *The Age of the Earth—An Introduction to Geological Ideas* (London: Benn) p 1927
Hoyle F 1960 *Quantum J. R. Astron. Soc.* **1** 28
——1984 *Earth, Moon and Planets* **31** 229
——1996 *The Global Warming Debate (The European Science and Environment Forum)* p 179
Hoyle F and Wickramasinghe N C 1978 *Astrophys. Space Sci.* **53** 523
——1985 *Living Comets* (Cardiff: Cardiff University Press)
Hughes D W and Daniels P A 1982 *Mon. Not. R. Astron. Soc.* **198** 573
Irvine W M *et al* 1996 *Nature* **383** 418
Jones J 1986 *Mon. Not. R. Astron. Soc.* **221** 257
Kary D M and Dones L 1996 *Icarus* **121** 207
Keynes J M 1947 *Newton Tercentenary Celebrations* (Cambridge: Cambridge University Press) p 27
Kozai Y 1962 *Astron. J.* **67** 591
Kuiper G P 1951 *Astrophysics* (New York: McGraw-Hill) p 357
Kresák L 1978a *Bull. Astron. Inst. Czech.* **29** 129
——1978b *Bull. Astron. Inst. Czech.* **29** 135
——1980 *Solid Particles in the Solar System (IAU Symp. 90)* (Dordrecht: Reidel) p 211
Kresák L and Kresákova 1987 *Interplanetary Matter* vol 2, ed Z Ceplecha and P Pecina (Prague: Czechoslovakian Academy of Science) p 265
Lecar M and Franklin F A 1973 *Icarus* **20** 422
Levison H and Duncan M 1990 *Astron. J.* **100** 1669
——1994 *Icarus* **108** 18
Luu J X and Jewitt D C 1988 *Astron. J.* **95** 1256
——1996 *Astron. J.* **111** 499
Lüst Rh 1984 *Astron. Astrophys.* **141** 94
MacLow M-M and Zahnle K 1994 *Astrophys. J. Lett.* **434** L33

Marochnik L S, Mukhin L M and Sagdeev R Z 1988 *Science* **242** 547
Marsden B 1990 *Asteroids, Comets, Meteors III* ed C-I Lagerkvist, H Eickman, B A Lindblad and M Lindgren (Uppsala: Uppsala University) p 393
Matese J J and Whitman P G 1989 *Icarus* **82** 389
Matese J J, Whitmire P G, Innanen K A and Valtonen 1995 *Icarus* **116** 255
McCrea W H 1981 *Proc. R. Soc.* A **375** 1
McFarland J 1996 *Vistas Astron.* **40** 343
McGlynn T A and Chapman R D 1989 *Astrophys. J. Lett.* **346** L105
McIntosh B A and Hajduk A 1983 *Mon. Not. R. Astron. Soc.* **205** 931
McLaren 1970 *J. Paleont.* **44** 801
Melosh H J and Shenk P 1993 *Nature* **365** 731
Mendis D A and Marconi M L 1986 *Earth, Moon and Planets* **36** 187
Mumma M J *et al* 1996 *Science* **272** 1310
Napier W M 1983 *Asteroids, Comets, Meteors* (Uppsala: Uppsala Observatory) p 391
——1987 *Proc. Tenth European Regional Meeting in Astronomy (Prague)* vol 2, p 13
——1990 *Dusty Objects in the Universe* ed E Bussoletti and A A Vittone (Dordrecht: Kluwer) p 103
——1993 *Meteoroids and their Parent Bodies* ed J Stohl and I P Williams (Bratislava: Astr. Inst. Slovak Acad. Sci.) p 123
Napier W M and Clube S V M 1979 *Nature* **282** 455
Napier W M and Humphries C M 1986 *Mon. Not. R. Astron. Soc.* **221** 105
Napier W M and Staniucha M 1982 *Mon. Not. R. Astron. Soc.* **198** 723
Nininger H H 1942 *Pop. Astron.* **50** 270
O'Brien S R, Mayewski P A, Mecker L D, Meese D A, Twickler M S and Whitton S I 1995 *Science* **270** 1962
Olano C A and Pöppel W G L 1987 *Astron. Astrophys.* **179** 202
Oort J H 1950 *Bull. Astron. Inst. Neth.* **11** 91
Öpik E J 1932 *Proc. Am. Acad. Arts Sci.* **67** 169
——1958 *Irish Astron. J.* **5** 34
——1963 *Astron. Astrophys.* **2** 219
Ovenden M W 1972 *Nature* **239** 508
Pittich E M and Rickman H 1994 *Astron. Astrophys.* **281** 579
Poveda A 1988 *Astrophys. Space Sci.* **142** 67
Poveda A and Allen C 1975 *Astrophys. J.* **200** 42
Quinn T, Tremaine S and Duncan M 1990 *Astrophys. J.* **355** 667
Rabinowitz D L 1993 *Astrophys. J.* **407** 412
Rabinowitz D L, Gehrels T, Scotti J V, McMillan R S, Perry M L, Wisnieski W, Larson S M, Howell E S and Mueller B E A 1993 *Nature* **363** 704
Rampino M R and Caldeira K 1992 *Cel. Mech. Dyn. Astron.* **54** 143
Raup D M and Sepkoski J J 1984 *Proc. Natl Acad. Sci., USA* **81** 801
Reddish V C 1971 *Nature* **232** 40
Robock A 1984 *Nature* **310** 667
Safronov V S 1969 *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets* (Jerusalem: Israel Program for Scientific Translations)
Sagdeev R Z, Elyasberg P E and Moroz V I 1988 *Nature* **331** 240
Schneider S H and Thompson S L 1987 *Nature* **334** 221
Scoville N Z and Sanders D B 1986 *The Galaxy and the Solar System* ed R Smoluchowski, J N Bahcall and M S Matthews (Tucson, AR: University of Arizona) p 69
Scotti J V 1994 *IAU Symp. 160: Asteroids, Comets, Meteors 1993* ed A Milani *et al* (Dordrecht: Kluwer) p 17
Scotti J V and Melosh H J 1993 *Nature* **365** 733
Scotti J V and Jedicke R 1996 *Dynamics, Ephemerides and Astrometry of the Solar System* ed S Ferraz-Mello *et al* (Dordrecht: Kluwer) p 389
Sekanina Z, Chodas P W and Yeomans D K 1994 *Astron. Astrophys.* **289** 607
Sekanina Z and Yeomans D K 1984 *Astron J* **89** 154
Sepkoski J J 1990 *Geol. Soc. Am., Spec. Pap.* **247** 33
Seyfert C K and Sirkin L A 1979 *Earth History and Plate Tectonics* (New York: Harper Row)
Singer S F and Stanley J E 1980 *IAU Symp.* **90** 329
Šolc M, Vanŷsek V and Kissel J 1987 *ESA SP-278* (Noordwijk: ESA) p 359
Sonett CP and Finney S A 1990 *Phil. Trans. R. Soc.* A **330** 413
Stagg C R and Bailey M E 1989 *Mon. Not. R. Astron. Soc.* **241** 507

Steel D I 1995 *Earth, Moon and Planets* **71** 279

Steel D I and Asher D J 1996 *Mon. Not. R. Astron. Soc.* **280** 806

Steel D I, Asher D J, Napier W M and Clube S V M 1994 *Hazards due to Comets and Asteroids* ed T Gehrels (Tucson, AR: University of Arizona) p 463

Stern A and Campins H 1996 *Nature* **382** 507

Stǒhl J 1983 *Asteroids, Comets, Meteors* (Uppsala: Uppsala Observatory) p 419

Stothers R 1988 *Observatory* **108** 1

Sykes M V and Walker R G 1992 *Icarus* **95** 180

Taylor A D 1995 *Icarus* **116** 154

Thomson D J 1990 *Phil. Trans. R. Soc.* A **330** 601

Torbett M V and Smoluchowski R 1990 *Nature* **345** 49

Turco R P, Toon O B, Ackerman T P, Pollack J P and Sagan C 1983 *Science* **222** 1283

Urey H C 1973 *Nature* **242** 32

Valsecchi G B and Manara A 1996 *Astron. Astrophys.* in press

Valsecchi G B, Morbidelli A, Gonci R, Farinella P, Froeschlè Ch and Froeschlè Cl 1995 *Icarus* **118** 169

Valtonen M J, Matese J J, Zheng J Q and Whitmire P G 1996 *Earth, Moon and Planets*

Van Flandern T 1993 *Dark Matter, Missing Planets and New Comets* (Berkeley, CA: North Atlantic Books)

Westley M S, Baragiola R A, Johnson R E and Baratta G A 1995 *Nature* **373** 405

Whipple F L 1967 *The Zodiacal Light and the Interplanetary Medium* NASA SP-150 p 409

Whipple F L 1996 *Earth, Moon and Planets* **72** 69

Whipple F L and Hamid S E 1952 *Helwan Obs. Bull.* **41** 1

Williams 1981 *Megacycles* (Stroudsburg: Hutchinson Ross)

Wolstencroft R J and Walker N C 1988 *Phil. Trans. R. Soc.* A **325** 423

Yabushita S 1983 *Astrophys. Space Sci.* **89** 159

——1988 *Mon. Not. R. Astron. Soc.* **231** 723

——1989 *Mon. Not. R. Astron. Soc.* **240** 69

Yabushita S and Hatta N 1994 *Earth, Moon and Planets* **65** 7

Yamamoto T 1996 *Science* **273** 921

Yeomans D K 1986 *Proc. 20th ESLAB Symp. on the Exploration of Halley's Comet* ESA SP-250, p 419

Zheng J Q, Valtonen M J, Mikkola S and Rickman H 1996 *IAU Symp. 172: Dynamics, Ephemerides and Astrometry of the Solar System* ed S Ferraz-Mello, B Morando and J-E Arlot (Dordrecht: Kluwer) p 209