THE ASTROPHYSICAL JOURNAL, 328: L69–L73, 1988 May 15

THE ORIGIN OF SHORT-PERIOD COMETS

MARTIN DUNCAN
Lick Observatory, University of California at Santa Cruz

AND

THOMAS QUINN AND SCOTT TREMAINE
Canadian Institute for Theoretical Astrophysics, University of Toronto
Received 1987 December 15; accepted 1988 February 23

ABSTRACT

We present the key results of an extensive series of numerical simulations of the evolution of comet orbits due to the gravitational perturbations of the giant planets. Our results show that the inclination distribution of comets with large perihelion ($q \lesssim 30$ AU) that evolve to observable comets (i.e., those with $q \lesssim 1.5$ AU) is approximately preserved. Thus, the short-period (SP) comets, which are mostly in prograde, low-inclination orbits, cannot arise from gravitational scattering of any spherical population of comets (such as the Oort cloud). However, the distribution of orbital elements of SP comets arising from a population of low-inclination Neptune-crossing comets is in excellent agreement with observations. We conclude that the SP comets arise from a cometary belt in the outer solar system. The spatial distribution of comets in the belt, and the mechanism by which belt comets acquire Neptune-crossing orbits, are still unclear.

Subject headings: comets — solar system: general

I. INTRODUCTION

There are several striking features in the distribution of orbital elements of comets with orbital period $P < 200$ yr (the short-period or SP comets):

1. The distribution is strongly peaked toward periods $\lesssim 15$ yr. There are 100 known comets with orbital period $P \leq 15$ yr but only 21 with 15 yr $< P < 200$ yr (Marsden 1983).

2. The SP comets are mostly on low-inclination prograde orbits. Out of 121 known SP comets, only four are on retrograde orbits; the inclinations i satisfy $\langle \cos i \rangle = 0.88$ ($\langle \cos i \rangle = 0$ for isotropic orbits and $= 1$ for prograde orbits in the ecliptic).

3. The arguments of perihelion ω (the angle between perihelion and ascending node) of the SP comets are strongly peaked near 0° and 180° (see Fig. 2, below).

It is usually believed that the SP comets originate in the Oort comet cloud. In an influential paper, Everhart (1972) argued from an extensive set of orbital integrations that repeated interactions with Jupiter can produce SP orbits from near-parabolic orbits, so long as the initial inclination is small and the initial perihelion distance is near the orbit of Jupiter. He showed that the distribution of orbital elements for SP comets formed by this mechanism agreed well with observations. There is some concern, however, that Everhart's mechanism may not produce the correct number of SP comets. Joss (1973) has argued that the efficiency of this process is too low (by a factor of order 10^3–10^4) to produce the observed number of SP comets from the known flux of near-parabolic comets (see Bailey 1986 for more recent references and Shoemaker and Wolfe 1982 for a different point of view).

In an ingenious Monte Carlo simulation, Everhart (1977) subsequently showed that a fraction of near-parabolic orbits with perihelion as large as Neptune's orbit will be gravitationally scattered by the outer planets into orbits that are Jupiter-crossing and that a fraction of these will eventually become SP comets. In conventional models of the Oort cloud, the efficiency of this process is too low to remove the flux discrepancy noted by Joss. However, Bailey (1986) has suggested that if a very massive inner Oort cloud (inner radius ≈ 5000 AU, mass $\approx 700\ M_{\oplus}$) is present, the flux of near-parabolic comets into Neptune-crossing orbits may be sufficient to supply the SP comets.

An alternative theory proposes that SP comets originate in a belt of low-inclination comets just beyond the orbit of Neptune, between about 35 and 50 AU (e.g., Fernández 1980; Fernández and Ip 1983). The belt could be a natural remnant of the outermost parts of the solar nebula (Kuiper 1951; Cameron 1962), perhaps responsible for perturbations to Neptune's orbit (Whipple 1964, 1972; Bailey 1983) and possibly producing a component of the infrared background at 100 μm detected by *IRAS* (Low *et al.* 1984). *IRAS* data also indicate the presence of flattened dust shells around other stars, extending from 20 to 100 AU (see Weissman 1986 for a review). If some of the belt comets are on Neptune-crossing orbits, subsequent scattering by the outer planets converts some of these into observable SP comets, in the manner described by Everhart (1977).

In this *Letter* we present the key results of a set of extensive numerical simulations of gravitational scattering of comets by the giant planets, designed to determine whether the most likely source of the SP comets is the Oort cloud or the Kuiper belt. A more detailed discussion of our simulations is in preparation (Duncan, Quinn, and Tremaine 1988).

II. METHOD

We have used the regularized integration scheme described in Duncan, Quinn, and Tremaine (1987) to follow the orbital evolution of comets subject to the perturbations of the four giant planets. In the simulations described below, each orbit is integrated directly until the comet either is ejected or becomes visible to Earth-based observers, which we usually assume occurs when its perihelion $q < 1.5$ AU. A numerical obstacle that we must overcome is that orbital evolution can be a slow process: evolution from a Neptune-crossing orbit to a visible orbit typically takes millions of orbits. However, since gravitational scattering is a diffusion process, increasing the mass of all the giant planets by a fixed factor μ should change the rate of evolution but not the statistical properties of the final distribution of orbits. (We have confirmed this argument by running several simulations with two values of μ.) The simulations reported here have factors μ ranging from 1 to 40. Even with this shortcut, the simulations reported here consumed several months of CPU time on a dedicated Sun-3-microcomputer with FPA board.

III. RESULTS

We began with simulations designed to check Everhart's (1972) conclusion that the SP comets can evolve from comets on parabolic orbits with an isotropic distribution of inclinations. Following Everhart, the initial perihelion distribution of the parabolic comets was taken to be uniform in the range 4 AU $< q <$ 6 AU. To permit direct comparison with Everhart, we assumed that the comets became visible when $q <$ 2.5 AU, rather than using our usual criterion $q <$ 1.5 AU. Since these comets are strongly influenced by Jupiter, their orbits evolve sufficiently rapidly that we did not require any mass enhancement factor, i.e., we set $\mu = 1$. We followed an ensemble of 5000 comets until each comet was either ejected or became visible according to the above definition. We found that 28 comets became visible, after a median evolution time of 2.30×10^6 yr and a median number 713 of perihelion passages. Of these, 12 were SP comets ($P <$ 200 yr). Figure 1 (*left*) shows the distribution of initial inclinations, i_i, versus the final inclinations, i_f, at the time each comet first becomes visible. Note that the inclination is approximately preserved, even though the perihelion has evolved by about a factor of 2, and visible comets are produced with a wide range of inclinations. Both the inclination and semimajor axis distribution of the visible SP comets ($q <$ 2.5 AU, $P <$ 200 yr) are inconsistent with the observed SP distributions. The average value of cos (i_f) in our simulations was 0.29, compared with 0.88 in the observations; nine of the 12 comets had $P >$ 15 yr in the simulations compared to only 21 of 121 observed comets.

Our conclusions disagree with Everhart (1972), possibly because he did not adequately sample the high-inclination orbits. The inconsistency between the observed and model inclination and period distributions leads to the conclusion that the SP comets cannot arise through gravitational scattering of comets on isotropically distributed near-parabolic orbits with perihelion near Jupiter's orbit. An alternative possibility (Everhart 1977; Bailey 1986) is that Oort cloud comets with larger perihelion, up to q = 30 AU, are scattered into the visibility zone $q <$ 1.5 AU by a combination of the giant planets. To test this hypothesis, we examined the evolution of an ensemble of orbits with isotropically distributed inclinations and perihelia uniformly distributed in the range 20 AU $< q <$ 30 AU. In this case the initial evolution is slow because the Uranus and Neptune masses are small; to improve efficiency we enhanced the masses of all the giant planets by a factor $\mu = 40$ and began the integrations with semimajor axes a = 50 AU (for $a \gtrsim$ 50 AU the perihelion and inclination are approximately conserved in planetary scattering). In Figure 1 (*right*) we plot the inclinations i_i and i_f for those comets that evolved into visible SP orbits. (In fact the initial inclination distribution shown in this figure was only isotropic over the range $0° \leq i \leq 90°$. Retrograde orbits only evolve into visible orbits very rarely; they are usually quite stable and hence are expensive to compute since in principle they must be followed for the lifetime of the solar system. If retrograde orbits were included the conclusions below would be strengthened.) Although the distribution in semimajor axis is similar to what is observed, the model inclination distribution is still inconsistent with the observed inclination distribution for SP comets: there are far too many retrograde orbits.

It appears to be impossible for the action of planetary perturbations on an initially isotropic distribution of orbits to produce a distribution of visible comets with the observed inclination distribution. We conclude that SP comets do not arise in the Oort cloud, and we turn to the possibility that the source of the visible SP comets is a belt of low-inclination comets located near Neptune's orbit. Having such a source does not preclude an additional source in the Oort cloud, but as noted by Joss (1973), the Oort cloud is expected to contribute a negligible fraction of the observed sample.

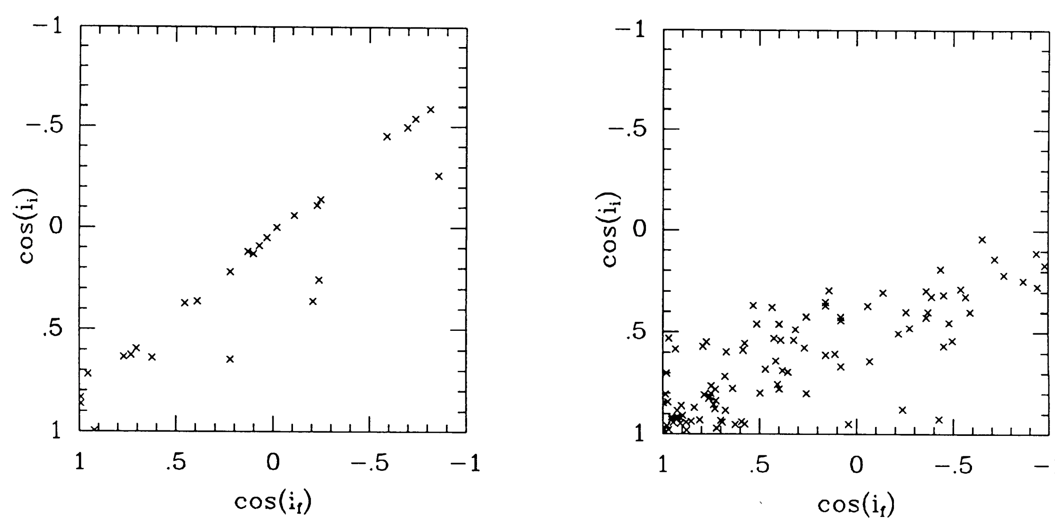

FIG. 1.—(*left*) Initial inclination is plotted against final inclination for the comets that evolve to $q <$ 2.5 AU from an initially isotropic distribution of 5000 parabolic comets with perihelion near Jupiter's orbit (4 AU $< q <$ 6 AU). (*right*) Initial vs. final inclination for the comets that evolve to SP orbits with $q <$ 1.5 AU from an initially isotropic, prograde [i.e., uniform in cos (i) between 0 and 1] distribution of 5000 Neptune-crossing comets (20 AU $< q <$ 30 AU) with semimajor axis 50 AU. To increase computational speed, the masses of all the giant planets have been increased in the right-hand panel by a factor $\mu = 40$.

We assume that some mechanism (to be discussed below) perturbs comets in the belt into Neptune-crossing orbits and follow their orbital evolution under the influence of the giant planets until they escape or become visible. The initial inclinations are chosen to be uniformly distributed in cos (i) between $i = 0$ and $i = 18°$, the initial perihelia are uniformly distributed between 20 and 30 AU, and the initial aphelia are all 50 AU. Our results are insensitive to the initial conditions so long as the comets begin on low-inclination Neptune-crossing orbits.

Figure 2 shows the distributions of semimajor axis a, inclination i, argument of perihelion ω, and aphelion for the comets when they first become visible ($q < 1.5$ AU). Also shown are the observed distributions for the comets in Marsden's (1983) catalog with $P < 200$ yr. Despite some minor differences, the agreement is impressive (see Fernández and Ip 1983 for a similar diagram). The two semimajor axis distributions both show strong peaks near $a = 3$ AU, with a shoulder near Jupiter's semimajor axis (5 AU). The aphelion distributions show a strong peak near Jupiter's semimajor axis. The inclination distributions are similar, although the simulation shows a somewhat broader distribution. The distributions of ω both show minima at $\omega = 90°$ and 270°, although the peaks are somewhat more pronounced in the data than the simulation. About 8% of the visible SP comets in our simulation are retrograde, in crude agreement with the observed value of 3%, and the retrograde comets tend to have larger periods (over 50% have $P > 15$ yr compared to 25% of all visible SP comets in the simulations). Thus, even comets like Halley's ($i = 162°$, $P = 76$ yr) may originate in a belt.

The small differences between the data and the simulation probably arise from our crude criterion for visibility ($q < 1.5$ AU) and from the enhanced importance of close encounters due to the increased planetary masses.

IV. DISCUSSION

A number of authors (e.g., Kuiper 1951; Fernández 1980) have pointed out that if the solid material in the outer planets is dispersed into a disk, and the resulting surface density distribution is extrapolated outward, then at least several Earth masses of material are expected to be present between 35 and 50 AU, with possibly much more material beyond 50 AU. This material provides a promising reservoir that may supply the

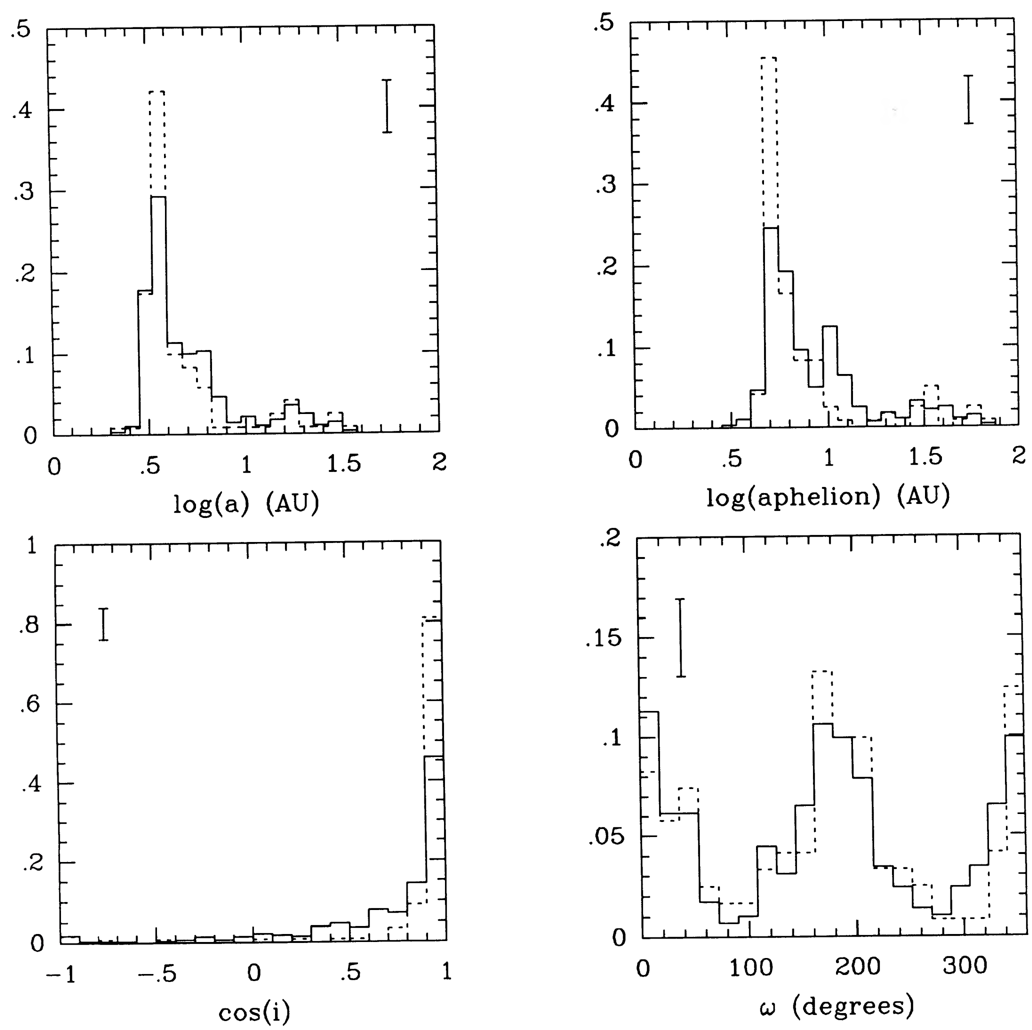

FIG. 2.—The solid histograms show the distributions of semimajor axis a, inclination i, argument of perihelion ω, and aphelion for comets that evolve to SP orbits with $q < 1.5$ AU in our simulations. The initial inclinations are uniformly distributed in cos (i) for i between 0° and 18°, the initial perihelia are uniformly distributed between 20 and 30 AU, and the initial semimajor axis is $a = 50$ AU. The dashed histograms show the distributions in the same four orbital elements for the comets in Marsden's (1983) catalog. Some typical statistical error bars are plotted. The histograms are based on 281 simulated comets and 121 observed comets.

SP comets. The mass of the most unstable perturbations in this belt is of order $m_0 = \Sigma(\lambda_c/2)^2$, where Σ is the surface density and $\lambda_c = 4\pi^2\Sigma r^3/M_\odot$ is the most unstable wavelength (Goldreich and Ward 1973); if 1 $M_\oplus$ is spread uniformly between 35 and 50 AU, then at $r = 40$ AU we find $m_0 = 1 \times 10^{18}$ g, not far from a typical comet mass.

We may use our results to estimate the minimum mass of the comet reservoir required to supply the SP comets for the lifetime of the solar system. In our simulations the fraction of Neptune crossers that become visible is ~ 0.17, somewhat higher than Fernández's (1980) estimate of $(0.5)^4 = 0.06$. Adopting a production rate of 10^{-2} yr^{-1} for SP comets (Fernández 1985) and a mean mass per comet of $10^{17.5}$ g we conclude that at least ~ 0.02 $M_\oplus$ of Neptune-crossing comets are required to maintain a steady supply of SP comets over the age of the solar system. Thus even a small fraction of an Earth mass of material is ample to supply the SP comets. This is consistent with crude upper limits to the mass of the comet belt of ~ 5 $M_\oplus$ (Anderson and Standish 1986) and ~ 1 $M_\oplus$ (Hamid, Marsden, and Whipple 1968; Yeomans 1986) obtained from spacecraft and comet trajectories.

The nature of the mechanism that delivers the belt comets into Neptune-crossing orbits remains unclear. Gravitational scattering by massive ($\gtrsim 10^{-4}$ $M_\oplus$) belt members could produce Neptune crossers at an adequate rate from comets on near-circular obits outside Neptune (Fernández 1980); however, for any plausible albedo the hypothetical massive belt members should already have been detected in photographic searches for slow-moving objects (see Luu and Jewitt 1988 for references).

An alternative possibility is that Neptune is the only important perturber of the belt comets, and that Neptune gradually feeds belt comets into Uranus-crossing orbits by repeated gravitational scattering over periods of several Gyr. This hypothesis can be tested by comparing the cratering history of solar system bodies (Shoemaker and Wolfe 1982) to the expected decay of a primordial population of belt comets. Computations by Fernández and Ip (1983), based on an approximate dynamical model, suggest that the cratering rate due to primordial Neptune-crossers declines by over two orders of magnitude in the past 3 Gyr, while the observations suggest a nearly constant cratering rate over this interval. Our more limited statistics suggest a similar conclusion: out of 1704 comets in our simulation, the last one survives for only 2.4 Gyr. However, in both cases the primordial orbits were all Neptune-crossers; low-eccentricity orbits near Neptune probably contribute a more steady flux.

The object Chiron orbiting between Saturn and Uranus (Kowal 1979) may well be a bright member of the parent population of the SP comets. In order to supply 10^{-2} SP comets per year from a population of low-eccentricity Neptune-crossers, our simulations suggest that there must be of order 10^5 comets inside the orbit of Uranus in the steady state. The mean eccentricity of these comets is 0.31, similar to the eccentricity of Chiron, 0.38. Comets have a broad luminosity function, which can be written in the form $\log N(<B) = \text{const} + kB$, where $N(<B)$ is the number of comets brighter than absolute magnitude B and $0.4 \lesssim k \lesssim 0.6$. If this luminosity function can be extrapolated, then in a sample of 10^5 comets brighter than $B = 16$ the brightest comet would have $3.5 \lesssim B \lesssim 7.7$; hence the existence of an object as bright as Chiron ($B = 7.0$) inside the orbit of Uranus is not surprising.

To summarize, the SP comets cannot be produced by planetary scattering of comets from the Oort cloud, or any other isotropic parent population. A comet belt (the "Kuiper belt") containing a fraction of an Earth mass and located in the outer parts of the solar system is plausible on cosmogonic grounds and appears to offer the most promising source for the SP comets, although the mechanism by which the comets are supplied to planet-crossing orbits remains unclear.

There are a number of possible approaches to further investigation of the Kuiper belt:

1. More detailed modeling of the dynamical influence of the belt on the outer planets, comets, and the *Pioneer* and *Voyager* spacecraft may yield improved upper limits to or actual detection of the belt mass.
2. Our models can be used to predict cratering rates on satellites of the outer planets for comparison with *Voyager* observations.
3. Occultations of bright stars by belt comets may be detectable by systematic searches (e.g., Bailey 1976).
4. Direct optical detection of belt members may be possible if belt members of 100 km radius or larger are common; for example, Chiron would have apparent visual magnitude 22 at 40 AU distance, and if we normalize using the expected number of comets in the belt ($\gtrsim 10^9$) and the expected fraction of Chiron-sized objects ($\gtrsim 10^{-5}$ for $k < 0.55$, where k is defined in the next-to-last paragraph) there would be at least 10^4 Chiron-sized objects in the belt.

We are grateful to Peter Goldreich, David Jewitt, and Gene Shoemaker for discussions and advice. This research was made possible by a Special Research Grant from the Connaught Fund at the University of Toronto and was supported by funding from NSERC.

REFERENCES

Anderson, J. D., and Standish, E. M. 1986, in *The Galaxy and the Solar System*, ed. R. Smoluchowski, J. N. Bahcall, and M. S. Matthews (Tucson: University of Arizona), p. 286.
Bailey, M. E. 1976, *Nature*, **259**, 290.
———. 1983, *Nature*, **302**, 399.
———. 1986, *Nature*, **324**, 350.
Cameron, A. G. W. 1962, *Icarus*, **1**, 13.
Duncan, M., Quinn, T., and Tremaine, S. 1987, *A.J.*, **94**, 1330.
———. 1988, in preparation.
Everhart, E. 1972, *Ap. Letters*, **10**, 131.
———. 1977, in *Comets-Asteroids-Meteorites*, ed. A. H. Delsemme (Toledo: University of Toledo), 99.
Fernández, J. A. 1980, *M.N.R.A.S.*, **192**, 481.
———. 1985, *Icarus*, **64**, 308.
Fernández, J. A., and Ip, W.-H. 1983, *Icarus*, **54**, 377.
Goldreich, P., and Ward, W. R. 1973, *Ap. J.*, **183**, 1051.
Hamid, S. E., Marsden, B. G., and Whipple, F. L. 1968, *A.J.*, **73**, 727.
Joss, P. C. 1973, *Astr. Ap.*, **25**, 271.
Kowal, C. T. 1979, in *Asteroids*, ed. T. Gehrels (Tucson: University of Arizona), 436.
Kuiper, G. P. 1951, in *Astrophysics*, ed. J. A. Hynek (New York: McGraw-Hill), p. 357.
Low, F. J., *et al.* 1984, *Ap. J.* (*Letters*), **278**, L19.
Luu, J. X., and Jewitt, D. 1988, *A.J.*, in press.

Marsden, B. G. 1983, *Catalog of Cometary Orbits* (Hillside: Enslaw).
Shoemaker, E. M., and Wolfe, R. F. 1982, in *Satellites of Jupiter*, ed. D. Morrison (Tucson: University of Arizona), p. 277.
Weissman, P. R. 1986, in *The Galaxy and the Solar System*, ed. R. Smoluchowski, J. N. Bahcall, and M. S. Matthews (Tucson: University of Arizona), p. 204.
Whipple, F. L. 1964, *Proc. Nat. Acad. Sci.*, **51**, 711.
———. 1972, in *IAU Symposium 45, The Motion, Evolution of Orbits, and Origins of Comets*, ed. G. A. Chebotarev, E. I. Kazimirchak-Polanskaya, and B. G. Marsden (Dordrecht: Reidel), p. 401.
Yeomans, D. K. 1986, in *Proc. 20th ESLAB Symposium on the Exploration of Halley's Comet* (ESA SP-250), p. 419.

MARTIN DUNCAN: Lick Observatory, University of California, Santa Cruz, CA 950064

THOMAS QUINN and SCOTT TREMAINE: Canadian Institute for Theoretical Astrophysics, McLennan Labs, University of Toronto, Toronto, Canada M5S 1A1