The Origin of Comets

Persistent Puzzles Through Time

Mark E. Bailey
Armagh Observatory and Planetarium

Exploring the Cosmos since 1790

meb@arm.ac.uk
http://star.arm.ac.uk/
http://climate.arm.ac.uk/

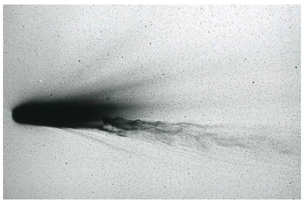
Some Famous Comets — I

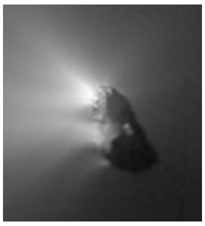
Comet C/1858 L1 (Donati)

- Discovered 2nd June 1858.
- Brightened through July and August.
- Easy naked-eye object during September and October that year.

Described by many as "The most beautiful comet of all time!"

Some Famous Comets — II





Comet 1P/Halley)

- Perhaps the most famous periodic comet.
- Returns every 75–76 years.
- ROE/UKST image (top) shows great tail 'disconnection' event of 1986 March 9.
- Nucleus imaged by ESA Giotto spacecraft 1986 March 14 (H.U. Keller)
  - size ∼15.3 ×7.2 × 7.2 km
  - average albedo ∼0.04
  - only 10–20% of surface 'active'

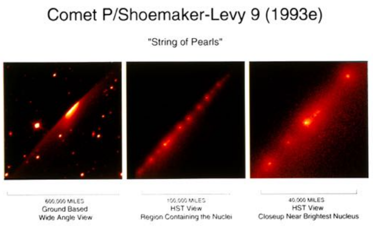
Some Famous Comets — III

- Discovered 1995 July 23 by Alan Hale and Thomas Bopp.
- A 'great comet', the best many of us will remember.
- Visible for several months during Spring 1997.
- Image signed by Thomas Bopp 1997 June 20, taken on 1997 March 28.

Comet C/1995 O1 (Hale-Bopp)

Some Famous Comets — IV: Shoemaker-Levy 9

- Comet D/1993 F2 Discovered 1993 March 25.
- Previously passed within Roche limit of Jupiter on 1992 July 8; broke into fragments.
- These fragments (the observed SL 9 comet) impacted on Jupiter from 1994 July 16–22.
- Impacts and impact scars visible from Earth.



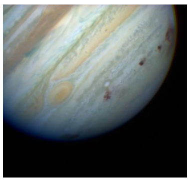
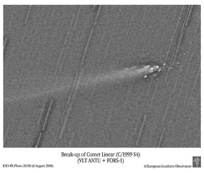


Image credits: H. Weaver & T. Smith; NASA/ESA

Various Cometary End-States

Disintegration



Break-up of Comet C/1999 S4 (LINEAR), 2000 August. Image credit ESO.

Outgassing

Comet C/1996 B2 (Hyakutake), 1996 March. Image credit D. Diereck.

Sun-Grazer

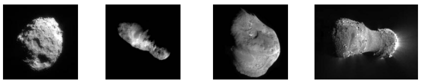
Comet C/1965 S1 (Ikeya-Seki), 1965 October. Image credit A. McClure.

Unresolved Questions

What are comets, and why so diverse?
How are they formed, and where?
Where do they primarily come from now?
What effects do they have on Earth (and Sun)?
How do they die and where do they go?

- dynamical ejection from solar system;
- collision with planets, or with Sun;
- evolution to inert end-state: e.g. by outgassing or formation of inert crust;
- physical decay and disintegration: e.g. loss of volatiles and dust, splitting, breakup etc.

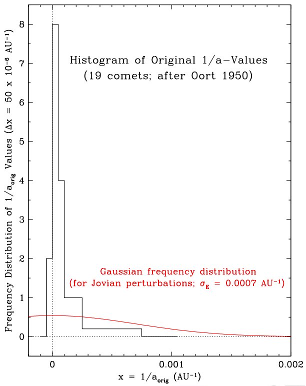
Comet:	81P/Wild 2	19P/Borrelly	9P/Tempel 1	103P/Hartley 2
Size:	∼4.2 km	∼5.0 km	∼6.1 km	∼1.4 km



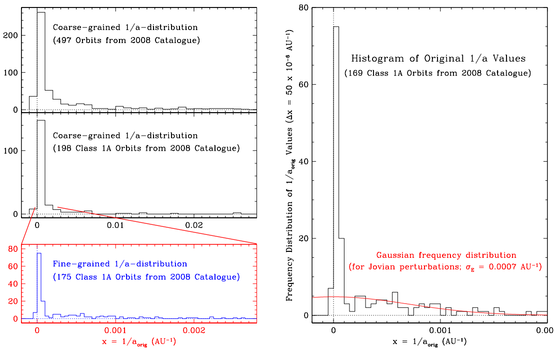
Credits: NASA/JPL/Stardust/Deep Space 1/Deep Impact

Birth of a Theory: The 1950 Oort Cloud

1. Oort considers the original $1/a$-values of the 19 most accurate orbits; i.e. those with mean errors $< 10^{-4}$ AU^{-1}.
2. Enables fine-grained binning of $1/a$-distribution for first time.
3. More than half had 'original' $1/a$-values $< 50 \times 10^{-6}$ AU^{-1}; and none had $1/a > 750 \times 10^{-6}$ AU^{-1}.
4. Note the extreme narrowness of the sharp peak in the distribution of 'observed' original $1/a$-values.



Comparison with Modern Data



Argument for Oort Cloud — c.1950

1. Sharp spike in observed $1/a$-distribution rules out interstellar capture (Van Woerkom 1948); and Lyttleton's 'capture theory' (1948) seriously deficient . . . suggests comets have primordial solar system origin, and the observed comets are coming into inner planetary region for the first time.
2. If comets are primordial, there must be a 'comet store' — the 'home' of the comet — somewhere beyond the zone of visibility, where comets can survive. Logically, this must contain comets in orbits of large perihelion distance.
3. Oort then addresses how to get comets from safe storage into inner solar system:
   - Planetary perturbations? — NO: they broaden the $1/a$-distribution too much (van Woerkom), contradicting observations.
   - Resistance of dense interstellar medium? — NO: it is implausible, and such a medium would primarily affect the comets' aphelion distances, again contradicting observed $1/a$-distribution.
   - Stellar perturbations? — YES: cometary orbits extend up to halfway to the nearest star; they must be affected by passing stars (cf. Öpik 1932).

Argument for Oort Cloud — in Modern Terms

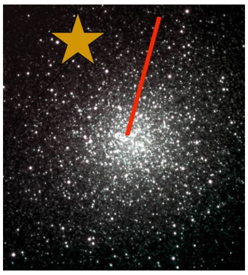
1. Observations $\Longrightarrow$ We see ~ 1 'new' comet ($q < 5$ AU, $H_{10} < 7$) discovered per year.
   - Semi-major axes $a > 2 \times 10^4$ AU: i.e. near parabolic limit; orbital periods $P \simeq$ 3–30 Myr — short compared to age of solar system.
   - These so-called 'new' comets strongly perturbed by Jupiter, so that roughly half ejected, the remainder 'captured'.
   - 'Captured' comets return, to be ejected or lost to short-period orbits and eventual decay.
2. Conclude: All observed comets are ultimately lost; and the 'loss cone' affects all orbits with $q \lesssim 15$ AU. The loss timescale $\ll$ age of solar system.
3. $\Longrightarrow$ comets are either a transient phenomenon, or there is a long-lived reservoir to replenish those that are lost.
4. Oort adopts primordial 'steady-state' hypothesis.

Further Details

1. 'New' comets are only lost if q lies within loss cone, i.e. $q \lesssim 15$ AU; $\Longrightarrow$ Oort's reservoir must contain long-period comets of large q.
2. For long-period orbits, planets change the orbital energy, i.e. change $1/a$, keeping q nearly constant; stars change the angular momentum, i.e. change q, keeping $1/a$ constant.
3. The change in q is about the size of the loss cone, provided the orbit is large enough.
   - Δq per revolution $\propto a^{7/2}$, i.e. depends sensitively on a.
   - $\Longrightarrow$ the reservoir must contain orbits of very long period ($a > 2 \times 10^4$ AU, $P > 5$ Myr) — just like the observations.
   - Leads to Oort's idea of a nearly spherical cloud of comets with orbits extending up to halfway to nearest star.
4. The cloud is 'gardened' by various external perturbations.
   - including stellar, molecular cloud and large-scale systematic effects of Galactic tide.

View of Oort Cloud

1. Like a globular star cluster, such as M13. . .
   Imagine Sun at centre
   - the stars become 'comets'
   - the shape (like a flattened rugby ball) is about right
   - the strong concentration of comets towards the centre is about right
   - the overall dynamics is similar
2. Can calculate 'families' of Oort cloud models, in the same way as for star clusters and galaxies
3. External perturbations (e.g. stars) change cometary orbits



The 'loss cone' behaves just like the loss cone around a black hole in a galactic nucleus

Structure of Oort Cloud

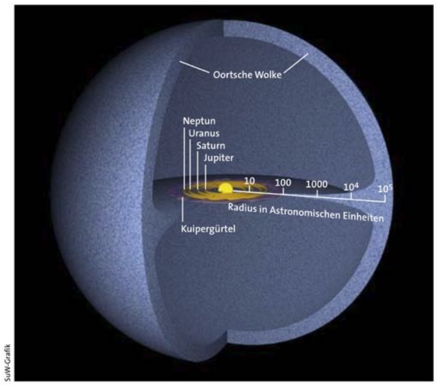
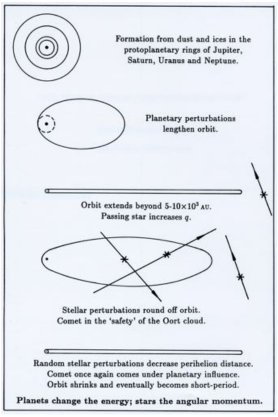


Image © Sterne und Weltraum (2011 Feb., p.20)

Standard (1950) Model

1. Assume: (1) spherical symmetry; outer radius $R_0 \simeq$200,000 AU; (2) random velocities, 'gardened' by stellar perturbations; (3) hydrostatic equilibrium (cloud neither expanding nor contracting); and (4) a simple energy distribution, e.g. a power-law distribution of orbital energies per unit mass $E = -GM_\odot/2a$.
2. If $f(E)\,dE \propto |E|^{-\gamma}dE$, then the number density $n(r)$ is roughly proportional to $r^{\gamma-4}$.
3. Oort's (1950) model has $\gamma = 5/2$, corresponding to velocity space being uniformly filled at r up to a value $V_{\rm max}$ equal to the free-fall speed from R_0 to r. This implies $n(r) \underset{\sim}{\propto} r^{-3/2}$, i.e. most of the mass near the outer edge.
4. Other models have smaller γ (e.g. $\gamma \sim 0$), and much sharper inward density increases. The structure is much more like a dense star cluster, with a strong concentration of mass towards the centre, not a shell.
5. Leads to the concept of an inner Oort cloud, i.e. a Dense Inner Core: a region inaccessible to observation but possibly containing most of the cometary mass, and relatively safe from external perturbations.

Oort Cloud Formation and Evolution Under Planetary Perturbations



Formation from dust and ices in the protoplanetary rings of Jupiter, Saturn, Uranus and Neptune.

Planetary perturbations lengthen orbit.

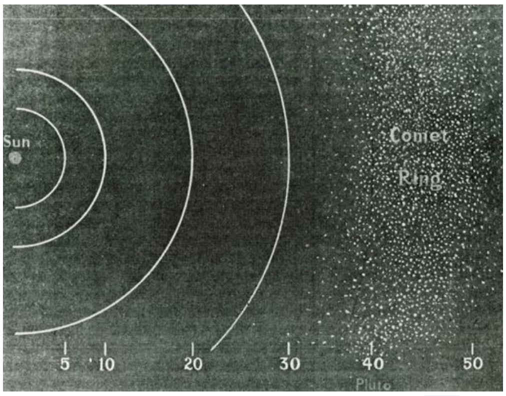
Orbit extends beyond 5-10×10³ AU. Passing star increases q.

Stellar perturbations round off orbit. Comet in the 'safety' of the Oort cloud.

Random stellar perturbations decrease perihelion distance. Comet once again comes under planetary influence. Orbit shrinks and eventually becomes short-period.

Planets change the energy; stars the angular momentum.

Whipple's (1964) Comet Ring: The Prototype EKB



Post-1990 Consensus

1. Comets are primordial solar-system bodies.
2. Formed by coagulation of originally interstellar dust grains during and after formation of protoplanetary disc.
3. Coagulation proceeds rapidly in inner solar system, more slowly farther out (cf. Kant 1755); to produce the protoplanetary building blocks, namely: cometesimals and planetesimals.
4. Late stages of planet formation involve (1) planetary and proto-planet collisions; (2) planetary migration (under mutual gravitational perturbations and evolution of protoplanetary disc); and (3) late-stage bombardment of planetary surfaces by comets and asteroids.
5. Most work on origin of comets now focuses on (1) dynamical evolution of short-period comets; (2) simulations of origin and evolution of the Solar System and formation of Oort cloud; (3) origin of Centaurs, in the Jupiter-Neptune region and beyond; and (4) the structure and evolution of the trans-Neptunian region, i.e. the Edgeworth-Kuiper belt and beyond.

Steps to Making Comets – I: Pre-Stellar Phase

1. Form dust grains in atmospheres of cool giant stars; eject to interstellar medium (ISM) via stellar winds.
2. Cook in ISM for 10–1,000 Myr: complex cycling of grains through hot diffuse ISM, cool molecular clouds (MCs) and cold MC Cores.
    - In the clouds, grains accrete a frosting of interstellar volatiles; in the hot ISM, ice is sputtered and UV photo-processed; and grains are ground down by collisions and evaporation.
3. Produce interstellar dust with a complex chemistry and broad size distribution; some grains have diameters up to microns or more.
4. Ices on and within the interstellar dust aggregates contain clues to the grains' previous history and to the processes that accompanied their 'final' pre-solar accumulation as part of the Sun's parent molecular cloud.
    - Cometary dust has a rich Cosmic Chemical Memory; cometary dust samples pre-solar history of solar-system material.

Steps to Making Comets – II: Protosolar Disc Phase

1. Form the solar nebula from a rotating protosolar molecular cloud. It cools and collapses to produce a dense gas-and-dust disc.
2. Typical cloud parameters: Temperature $T \approx 10\,\mathrm{K}$; Radius $R \approx 0.1\,\mathrm{pc}$; Mass $M \approx 1$–$2\,M_\odot$. Initial disc radius R_d small compared to R, but large compared to current planetary system. For reasonable parameters, $R_d \approx$ few $\times$ 100 AU.
3. Grains grow during nebular collapse and during disc evolution, acquiring a 'frosting' of ices from condensing volatiles in the MC core and protoplanetary disc.
4. In the inner few AU of nebula, dust destroyed by collisions or by heating from the newly formed proto-Sun; dust farther out retains its Cosmic Chemical Memory.

By time Sun forms, expect grains with a complex 'hierarchical' structure, with evidence of both hot (pre-stellar) phases and cold (MC) phases of evolution.

Models of Interstellar Grain Aggregates

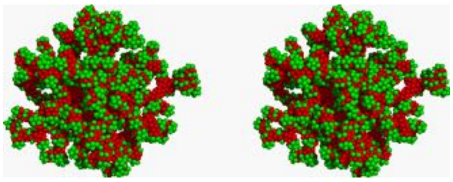


Image credits: Top: © E.L. Wright (UCLA); Bottom J.M. Greenberg (Leiden)

Steps to Making Comets – III: Protoplanetary Disc Phase

1. Condensed ice composition: expect ices such as water, carbon monoxide, carbon dioxide, methanol, hydrogen cyanide, ammonia, methane etc.
2. Disc surface density Σ_d(solids) at 10 AU approximately 10 kg m^{-2}; radial variation roughly a power law, i.e. $\Sigma_d \propto r^{-3/2}$. Gas-to-Dust ratio roughly 50 initially.
    - Surface density corresponds to a traditional 'minimum mass' protoplanetary disc within planetary region; total mass of solids within $\sim$300 AU could range up to several 100 Earth masses.
3. Initial grain growth proceeds rapidly in presence of gas through turbulence-driven coagulation. Large grains initially drift inwards due to gas drag and accrete smaller 'background' grains.
4. Grain radius versus time: $a(t) \approx 0.3\,(100\,\mathrm{AU}/r)^3\,(t/1\,\mathrm{Myr})$ m. Thus, 'boulders', i.e. bodies with sizes up to $\approx$10–100 m, may form within 30 AU in a gas-clearance time-scale $\lesssim$ a few Myr, but probably much smaller objects — 'dirty snowballs' — farther out, i.e. sizes $\approx$ 1–10 m.

Steps to Making Comets – IV: Growth Without Gas

1. Dissipation of gas disc $\Longrightarrow$ further grain growth in absence of gas.
2. Two main channels: (1) 'standard' planetesimal picture and variants therein (widely accepted); (2) local gravitational instability picture (much less widely accepted).
3. Consider (1) 'standard' planetesimal picture:
    - $\Longrightarrow$ continued collisions and growth of 'boulders'/'snowballs' to bodies up to several tens of km in protoplanetary zone.
    - Produces comets with collisionally evolved structure on scale of 'boulders' (i.e. $\sim$10–100 m), and looser 'rubble-pile' structure on larger ($\gtrsim$ km) scales; $\Longrightarrow$ comets collisionally evolved.
    - Gravitational stirring by the largest bodies leads to continued growth, ultimately to make large planetesimals and planets.
4. Problems: time-scale to produce Uranus and Neptune too long; leads to need for migration models; Comets are planetary 'left-overs' formed in or close to outer planetary region, so total cometary mass should be not much greater than that of solids in the planets; role of planetary migration; roles of EKB and Oort cloud.

Steps to Making Comets IV (cont.): (2) Grav. Instability

1. Formation of dynamically cold, quiescent disc of dirty snowballs in outer solar system; random velocities decrease due to collisions.
2. $\Longrightarrow$ conditions for local gravitational instability: the dirty-snowball disc fragments into subdiscs with characteristic sizes depending only on Σ_d and r. Detailed analysis shows that the first unstable mode, λ_c, has a wavelength $\lambda_c \simeq 4\pi^2 G\Sigma_d/\Omega^2$, where $\Omega = (GM_\odot/r^3)^{1/2}$. The most unstable modes have wavelengths about half this, i.e. $\lambda_p \simeq \lambda_c/2$.
3. Subdiscs evolve like mini protoplanetary discs: to produce central objects (often multiple systems) with masses comparable to the mass m_p of the subdiscs, i.e. $m_p \simeq \pi\,(\lambda_c/8)^2\,\Sigma_s = \pi^5\Sigma_s^3 r^6/4M_\odot^2 \propto r^{3/2}$, i.e. $m_p \gtrsim 10^{18}$ kg for $r \gtrsim 50$ AU.
4. First-formed objects have masses comparable to observed outer solar-system objects; and — once formed — collisions become rare.
5. The model predicts that comets are: (1) mostly made in outer solar system during evolution of subdiscs probably to produce multiple central objects; (2) the products of 'gentle' accumulation of 'boulders' or smaller 'snowball'-size components; and (3) therefore largely collisionally unevolved.

Steps to Making Comets – V: Summary

1. Comets produced by hierarchical accretion in outer planetary system; final sizes range from a few km up to a few 100 km.
2. Cosmic Chemical Memory: interstellar and interplanetary dust aggregates contain ices that give clues to each of the distinct phases of grain growth in presence of gas, i.e. (1) interstellar gas and MC phases; (2) protostellar cloud and collapse phases; and (3) early disc evolution in presence of gas.
3. Evolution in absence of gas much more uncertain; but 'boulders' and/or 'snowballs' must somehow grow into kilometre-size (and larger) comet nuclei.
    - In planetesimal picture, comets collisionally evolved; mostly formed in protoplanetary region and may have rubble-pile structure with more compact elements on scale of 'boulders' (10–100 m).
    - In gravitational instability picture, comets collisionally unevolved and of low-strength; most formed beyond planetary region and may have substructure on scale of 'snowballs' ($\lesssim 10$ m or less).

Persistent Puzzles of Comets

1. What are comets . . . Are they all the same . . . Or are some comets (e.g. long-period) different from others (e.g. short-period)?
2. How are comets formed, and where? For example, are comets formed in or beyond the protoplanetary region; or in the protostellar molecular cloud; or beyond, in interstellar space?
3. What is (or are) the proximate source(s) of observed comets?
4. What is the structure and evolution of the 'observed' Oort cloud; how was it formed; and does it contain a massive dense inner core?
5. What is the role of the newly discovered, large, outer solar system bodies: Centaurs, Edgeworth-Kuiper belt objects, trans-Neptunian objects etc?
6. What is the cometary mass function and average cometary mass? . . . and how many comets are there, and what is their total mass?
7. Are comets fragile or strong; what are their end-states; and what is the impact of cometary debris on the planets, Earth and Sun?

New Discoveries: Pluto and Edgeworth-Kuiper Belt

1. Discovery of Pluto (1930 February 18), announced March 13.
2. Consistent with earlier speculations (e.g. Lowell) about a 'Planet X' beyond Neptune; or that small objects might exist in the region beyond Neptune (e.g. Campbell 1916, Aitken 1926, Leuschner 1927, Leonard 1930).
3. Stimulates work by Edgeworth (1938, 1943); and later by Kuiper (1951), Whipple (1964), Fernández (1980), Duncan et al. (1988), Quinn et al. (1990) and others, focusing on JF short-period comets.
4. Searches by Kowal (1976–1985), Luu & Jewitt (1988), Levison & Duncan (1989), Tyson et al. (1992), eventually successful. Discovery of 'QB1', i.e. minor planet (15760) 1992 QB$_1$, the prototype 'cubewano' and the first 'Kuiper Belt' object.
5. Pluto = minor planet (134340) now among several large trans-Neptunian objects (TNOs) classified as 'dwarf planets'; Pluto: the 'king' of the Kuiper belt.

Pluto-Charon System and Trans-Neptunian Disc

1. The classical 'Kuiper Belt': predicted by Irish scientist K.E. Edgeworth and others around middle of 20th century.
2. Extends beyond Neptune: a vast belt or disc of icy planetesimals in low-inclination orbits: the trans-Neptunian disc.
3. There are $\approx 10^5$ trans-Neptunian objects (TNOs) with diameters greater than 100 km. Many more ($\approx 10^9$), it is believed, of 'ordinary comet' size.

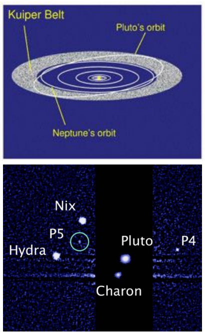


Image credits: Johns Hopkins University; NASA/ESA/HST/M. Showalter

Artist's Impression of Some Large TNOs/Dwarf Planets

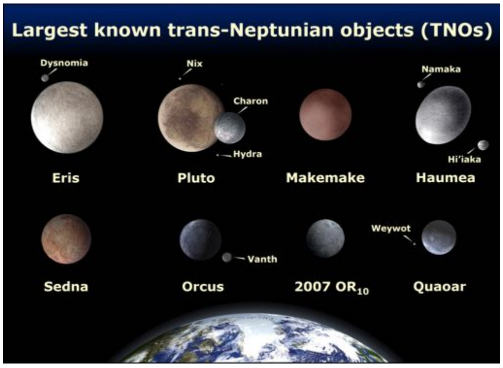


Image Credit: Wikipedia; based on 2006 press release by NASA/ESA/HST.

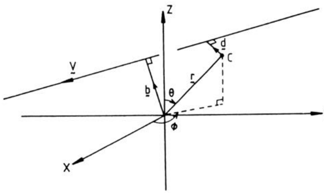
Evolution of Oort Cloud: Contrasting Views

1. 1950 Model: quasi-steady state; comets in long-term 'deep freeze' of Oort cloud for age of solar system; stars dominate the evolution; no dense inner core.
2. Modern view:
    - On short timescales: $t \lesssim 10$ Myr: Changes in perihelion distance dominated by Galactic tide and passing stars; leads to a quasi-steady long-period comet flux; changes in orbital energy dominated by stellar perturbations.
    - On medium timescales: $10 \lesssim t \lesssim 500$ Myr: periodic new-comet flux due to Sun's orbit about Galactic plan; rare, close stellar passages more important for randomizing orbits; changes in orbital energy still dominated by stars.
    - On long timescales: $500 \lesssim t \lesssim 4000$ Myr: rare, close molecular cloud encounters disrupt outer cloud, dominating changes in orbital energy beyond c.10,000 AU; rare, close stellar encounters stir up inner core.

    Together, these major upheavals replenish the transition zone between inner and outer Oort cloud and stir up orbits in Dense Inner Core.

New Discoveries: Survival Problem for Oort Cloud

1. Consider a perturber of mass M passing Sun with velocity $\mathbf{V}$ and impact parameter $\mathbf{b}$ with respect to Sun and $\mathbf{d}$ with respect to a comet at heliocentric distance $\mathbf{r}$.



2. Then the relative velocity change of the comet with respect to the Sun is the difference of the two impulses, i.e.

$$\mathbf{\Delta v} = \frac{2GM}{dV}\hat{\mathbf{d}} - \frac{2GM}{bV}\hat{\mathbf{b}} = \frac{2GM}{bV}\left\{\left(\frac{b^2}{d^2} - 1\right)\hat{\mathbf{b}} - \frac{rb}{d^2}\left[\hat{\mathbf{r}} - (\hat{\mathbf{r}}.\hat{\mathbf{V}})\hat{\mathbf{V}}\right]\right\}$$

Half-Life, $t_{1/2}(a)$, for Survival: Oort Cloud Evolution

1. Two main types of external perturber: stars and molecular clouds.
    - Galactic tide also drives comets into inner solar system, but has little direct effect on Oort cloud's disruption.
2. Stars pass through and beyond the Oort cloud, causing gradual unbinding of cometary orbits; the 'stellar' half-life is $t_{1/2,*} \simeq 2 \times \left(2 \times 10^4\,\mathrm{AU}/a\right)$ Gyr.
3. Molecular clouds pass beyond the Oort cloud, but are much more massive than stars; the 'molecular cloud' half-life is $t_{1/2,c} \simeq 2 \times \left(2 \times 10^4\,\mathrm{AU}/a\right)^3$ Gyr

$\Longrightarrow$ 'standard Oort cloud dynamically unstable beyond $a \simeq 2 \times 10^4$ AU, over the age of the solar system (4.5 Gyr)

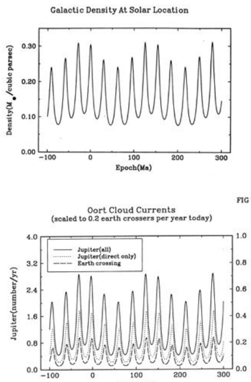
The Oort cloud is a leaky reservoir which must be replenished from within, possibly the trans-Neptunian region or a Dense Inner Core.

New Discoveries: Time-Variable Cometary Influx

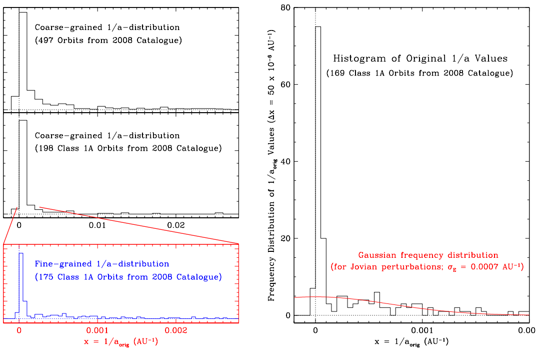
1. Galactic tide dominates quasi-steady new-comet flux from Oort cloud.
2. Comet flux roughly proportional to mass-density, $\rho_S(t)$ at Sun's location in Galaxy (see Figure, after J. Matese et al. 1995).
3. Δq per revolution depends on q, a, and Galactic latitude of perihelion, b, i.e.

   Δq per revolution $=$ $(10\pi^2\sqrt{2}\,\rho/M_\odot)\sin(2b)\,q^{1/2}a^{7/2}$

   $\Longrightarrow$ Galactic influence on comet influx (c.30 Myr cycles)



New Discoveries: Fading Problem – Recall 1/*a*-distribution



Halley-Type Comet (HTC) Capture Probability

Inclination-averaged mean capture probability from near-parabolic orbits to a HTC orbit:

- Decreases sharply with increasing q.
- Non-zero out to $q \simeq 15$ AU.
- Averages $\sim$0.01 for $q \lesssim 5$ AU.

The new-comet flux ($\sim$1 per year) and mean dynamical lifetime as a HTC ($\sim$0.3 Myr), and the capture probability, p_c determines the predicted number of HTCs.

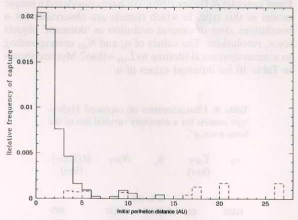


Figure 4. The relative frequency of capture from an isotropic near-parabolic source into Halley-type orbits (solid line) and Jupiter-family orbits (dashed line). Note that whereas Halley-type comets primarily originate from orbits of small perihelion distance, Jupiter-family comets primarily come from initial orbits in the outer Solar system.

Fading Problem: Where Are the 'Dead' Comets?

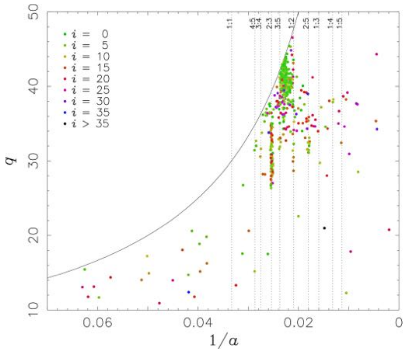
1. Observed new-comet flux: Approximately 1 comet per year brighter than $H_{10} = 7$ (corresponds to diameter $d \gtrsim 5$ km) with $q < 5$ AU, i.e. with perihelion distance within Jupiter's orbit.
2. Capture probability to 'Halley-type comet' (HTC), i.e. capture probability to $P \lesssim 200$ yr: $p_c \simeq 0.01$ per new comet; the rest get ejected.
3. Mean dynamical lifetime as a Halley-type comet: $t_{\rm dyn} \simeq 3 \times 10^5$ yr.
4. $\Longrightarrow$ steady-state number of HTCs, $N_{\rm HTC}$, given by $N_{\rm HTC} \simeq 1 \times 0.01 \times 300,000 \approx 3000$.
5. 30–100 times more than observed: where are the dead comets?!
   - Perhaps they are 'dark' HT asteroids; 'boulders'; or 'dust'?
   - In any case, comets must have short lifetimes in visible region.

Origin of Jupiter-Family Comets? — Not Kuiper Belt!

1. Jupiter-family (JF) short-period comets (SPCs) mostly have low inclinations.
2. Suggests a source in a flattened, low-eccentricity, disc-like distribution (Duncan, Quinn, & Tremaine, 1988): the 'classical' Kuiper belt.
   - Comets must have perihelia near Neptune, i.e. $q \approx 30$ AU, in order to be efficiently captured and 'handed down' to the Jupiter family.
   - Simulations require at least 4×10^9 such comets in the comet belt, if this is the dominant source of JFCs.
   - The dynamical lifetime of JFCs is $\approx 3 \times 10^5$ yr; their active lifetime is much shorter, i.e. $\approx 1.2 \times 10^4$ yr (otherwise inclinations increase).
3. Two main problems: (1) the required source orbits are not observed; and (2) evolution of an initial distribution of low-inclination Neptune-crossing orbits inevitably produces a 'Scattered Disc' containing a similar number of comets in much more eccentric low-inclination orbits. These are much more readily captured into JFC orbits.

$\Longrightarrow$ JFCs primarily not from Kuiper belt, but from Scattered Disc.

Distribution of Observed High-Accuracy TNOs



High-accuracy TNO orbits showing mean-motion resonances with Neptune and lack of non-resonant objects with q near Neptune and $a \lesssim 50$ AU. Image credit: David Asher.

New Number Problems: Scattered Disc and Oort Cloud

1. Simulations require number of comets in Scattered Disc to be $\lesssim 10^9$.

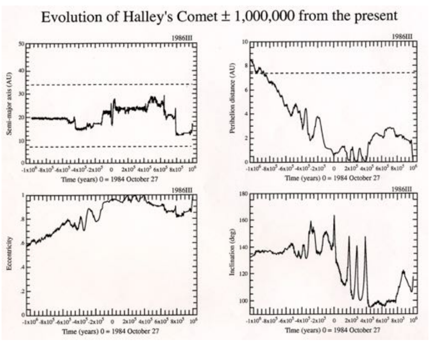
   But best observational estimates of number in Scattered Disc are $\approx$1–2$\,\times 10^8$, albeit with large uncertainty.
2. Simulations require number of comets in Oort Cloud to be approximately ten times the number in Scattered Disc, i.e. $\lesssim 10^{10}$.

   But best observational estimates of number in Oort Cloud $\gtrsim 2 \times 10^{11}$.
3. Two recent 'solutions' (NB: ideas go back many years):
   - Observed JFCs not in steady state; or large Scattered Disc Objects tidally break up into many fragments during dynamical evolution towards JFCs (Volk & Malhotra 2008).
   - Observed Oort cloud not primordial to solar system, but instead comprises largely captured comets, e.g. ejected from the Scattered Discs of other stars making up our Sun's parent star cluster (Levison, Duncan, Brasser & Kaufmann 2010).

New Discoveries: Complex Dynamical Evolution

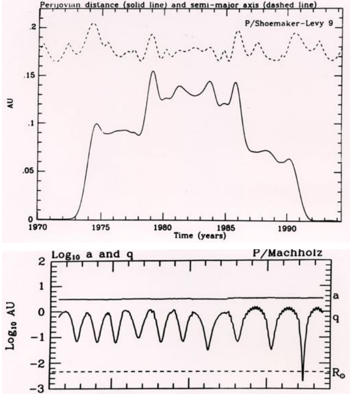
e.g. 1P/Halley:

1. Resonances: mean-motion and secular.
2. Kozai Cycles: Correlated large changes of eccentricity and inclination.
3. Sungrazing: a ubiquitous cometary end-state.



New Discoveries: Kozai Cycles and Resonances

1. Kozai Cycles, e.g. Comet S-L 9, 96P/Machholz, produce correlated large changes of eccentricity and inclination; leads to Jupiter-grazing and Sun-grazing.
2. A very general dynamical process.
3. Also seen in evolution of Oort cloud, exoplanet and multiple-star systems, and galaxy satellites.



New Discoveries: Complex Secular Resonances

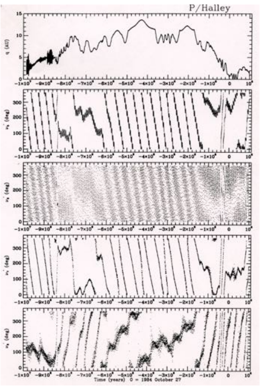
e.g. 1P/Halley Secular Resonances:

1. Note enormous secular evolution of perihelion distance.

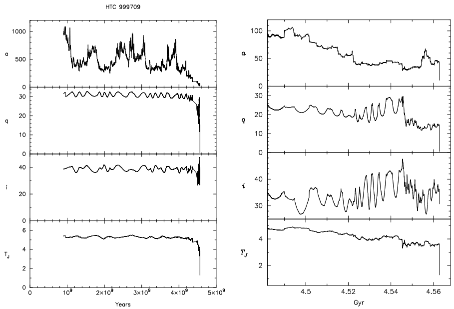
   Associated with critical argument $\nu_p = (\Omega - \omega) - (\Omega_p + \omega_p)$

   When this $\sim$constant, line of apses of comet's orbit locks on to the rate of precession of one of the giant planets (J, S, U, N).

   Figure shows effects on q of such resonances with Jupiter (2nd panel) and Neptune (5th panel). $-10 < t < 1$ Myr.
2. This kind of evolution totally unexpected: quite different from pure 'random walk'.



Halley-Type Comets From Inner Oort Cloud



Example of Halley-type comet from inner Oort cloud, involving gradual dynamical transfer from outer solar system ($a > 10^3$ AU and initial q near Neptune) through weak perturbations. ≈10% of HTCs originate this way. Image credit: Emel'yanenko et al. 2007, MNRAS, 381, 779–789.

New Puzzles: How Formed and Whence Come Comets?

1. How are comets formed? And where?

   In or beyond the protoplanetary region; in the Sun's parent molecular cloud; in circumstellar discs around other stars; or interstellar space?
   - Standard View: by coagulation and/or subsequent gravitational instability of small bodies (ice-covered dust/boulders) in and/or just beyond the solar protoplanetary disc.
2. What are the proximate source(s) of observed comets?
   - Standard View: Principally Oort Cloud for long-period comets and Halley-type SPCs; the Scattered Disc for Centaurs and JF SPCs.
   - However, Emel'yanenko, Asher & Bailey argue for an Oort Cloud source for long-period comets, Halley-type SPCs, AND 'Centaurs' and ≈half the JFCs. The observed near-Neptune high-eccentricity orbits can explain the remaining Centaurs and the other ≈half of JFCs.

   NB: Here, 'Centaurs' are objects with $5 < q < 28$ AU and $a < 1000$ AU and any inclination; and near-Neptune high-eccentricity orbits have $28 < q < 35.5$ AU and $60 < a < 1000$ AU and also any inclination.

Nature of Comets: One, Two or More Physical Types?

1. Comets very diverse. But are they essentially the same objects, or are there two or more subtypes, e.g. correlating with different orbital periods, initial mass or dynamical class?
2. Standard View: Includes at least two subtypes, principally (1) Centaurs and JFCs (which mostly come from the Scattered Disc; itself mostly from the near-Neptune part of the protoplanetary disc); and (2) LPCs and HTCs (mostly from the Oort Cloud; itself mostly from the Jupiter-Saturn-Uranus part of the protoplanetary disc).
   - JFCs have long active lifetimes in the visible region ($q < \lesssim 2.5$ AU), greater than $\approx 10^3$ revolutions; LPCs and HTCs have short active lifetimes, less than ≈200 revolutions. ($\Longrightarrow$ another 'fading' problem!)
3. Alternatively: All comets essentially the same (apart from mass). Formed by gentle accretion in outer regions of a heterogeneous protoplanetary disc, and ejected to produce the Oort cloud and its more flattened dense inner core.
   - All comets fragile, with short active lifetimes less than ≈ 200 revolutions in visible region. No clearcut physical distinction between HTCs and most JFCs, apart from their 'age' and typical mass.

Further Questions

1. What is structure and evolution of the 'observed' Oort cloud. How was it formed; and does it contain a massive dense inner core?
2. What is the cometary mass function and the average cometary mass? How many comets are there, and what is their total mass?
   - Is this consistent with standard 'low mass' protoplanetary disc models?
3. What is role played by newly discovered, large, outer solar system bodies: Centaurs, Edgeworth-Kuiper belt objects, trans-Neptunian objects etc?
4. Are comets fragile or strong? What are their end-states; and what is the impact of cometary debris on the Earth, other planets, and Sun?

General Conclusions

1. 'Comets' — even ordinary ones —can sometimes become the most prominent objects in sky; their study goes back thousands of years.
2. Comets touch on many areas of astronomy, not least solar-system science; they have had a significant impact on the Earth and on the development of scientific ideas.
3. Earth is now viewed as an 'open' system in touch with its near-space environment: a paradigm shift as significant as Copernicanism.
4. Solar system 'very leaky' — with important implications for the amount of dust, small bodies and planets within molecular clouds and the interstellar medium. For example, what about comet clouds around other stars?
5. Emerging Modern Picture of Comets: a balance between the historic catastrophist and subsequent uniformitarian views; i.e. comets as destroyers of life AND as bodies that bring the necessities of life (e.g. water, organics, perhaps seeds of life itself) to Earth.

Cometary Impacts Through Time?

1. Ancient history suggests 'the sky' may have been significantly different in proto-historic times (e.g. more 'active', more interplanetary debris, brighter zodiacal light etc.). How can that be?
2. Cometary masses range up to the size of dwarf planets. what are the effects of occasional 'giants' on Earth (and Sun)? What is the average mass of a comet?
3. Are all comets essentially the same; or are there two or more different classes, e.g. depending on origin and/or dynamical characteristics?
4. Total mass of Oort cloud may be very large ($\approx 10^2\,M_\oplus\ \mathrm{pc}^{-3}$). Does this lead to serious difficulties for the 'standard' primordial solar system picture?
5. 'Fading problem' still not understood, but effectively determines the predicted $1/a$-distribution. What happens to the cometary debris?
6. Meteoroid streams initially very fine-grained. This implies strong time-dependence in accretion of dust and small bodies on Earth.

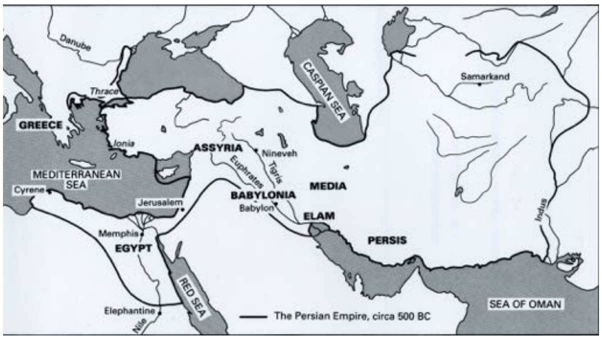
Comets in Astronomy and History

Competition between two main factors:

1. General advances in science and understanding that began in the Renaissance, i.e. the few centuries up to the Industrial Revolution;
    - Provide a backdrop against with to 'read' the literature on comets in the 17th–19th century and earlier;
    - End of the 18th century: a kind of 'watershed' between an older pre-scientific view of the natural world, and the modern 'scientific' view'.
2. A more or less continuous strand of interest in comets and cometary debris, from the earliest times right up to the present day, viz:
    - The physical and societal impact of comets;
    - Comets as agents of destruction (catastrophism) versus celestial bodies that convey(ed) ingredients necessary to sustain 'Life' on Earth;
    - The rejection and rediscovery of cometary catastrophism.

$\Longrightarrow$ New paradigm: Earth in touch with its near-space environment.

Cradle of Civilization I : Near-Eastern View

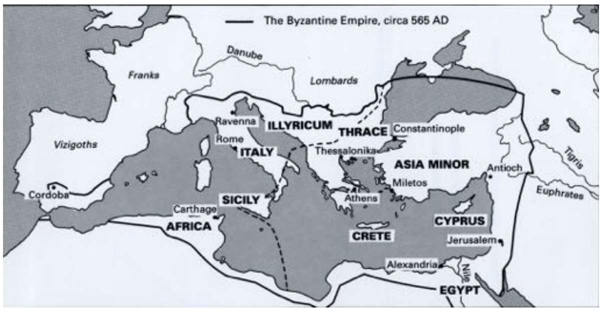


Foundations of Astrology

Four broad phases can be identified:

1. Judicial Astrology (≈3000–1000 BC)
    - Events in sky self-evidently influence events on Earth.
    - Celestial 'order' transmitted to Earth by sky-gods or deities.
    - $\Longrightarrow$ a powerful 'motive' to observe the sky and interpret the celestial 'omens'.
        - The sky gods are 'announcing' events on Earth, for example through the appearance of a bright comet or meteor, or by the fall of a meteorite or thunderbolt hurled by the sky-god Jupiter etc.
2. Zodiacal Astrology (≈1000–400 BC)
    - A slow transformation from Judicial Astrology to an increasing focus on the important part of the sky associated with the principal sky-gods, i.e. the Zodiac.
    - The sky divided into sections, each with a different perceived 'influence' on people or events on Earth.

Cradle of Civilization II: Mediterranean View



Early Greek Ideas: Anaximander's Jets of Fire

1. Earth seen as a short, squat cylinder three times as wide as long, surrounded by air and floating freely at the centre of the observable Universe in an infinite space.
2. Sun, planets, stars are enclosed circular hoops of fire below the Sun and Moon. They only become visible due to holes in their enclosing hoops that allow the fiery substance to leak out and become visible. There seems to be no rational explanation for this surreal view about the sky.

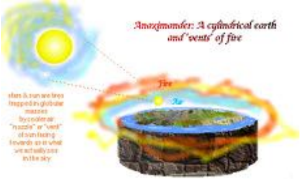


Image credit: Tony Mendes

Later Developments: Horoscopic Astrology to Science

3. Horoscopic Astrology (≈400 BC to ≈1600 AD)
    - Based on the entirely false premise that wandering stars ('planets') exert a distant controlling influence on human affairs.
        - Provides an early example of a powerful, but 'magical' scientific concept, i.e. 'action at a distance'.
    - Motivates careful observations of the planets; their paths against the fixed stars; their periods of revolution etc; all linked to predictions.
        - Demonstrates growing understanding and an increasingly 'scientific' approach to observatons of the natural world;
    - Nevertheless, the focus on unimportant chance alignments of planets and stars, planetary conjunctions etc. (e.g. 'Star of Bethlehem'), and on the 'random' appearance of an occasional bright comet etc. ultimately proves to be a cul-de-sac for science.
    - Despite this, the idea of horoscopic astrology has proved remarkably hard to shift: it's still believed by upwards of 25% of the population!
4. Scientific Astronomy (≈1600 AD to present)

Romans and Etruscans: Seeing Comets/Meteors as Omens

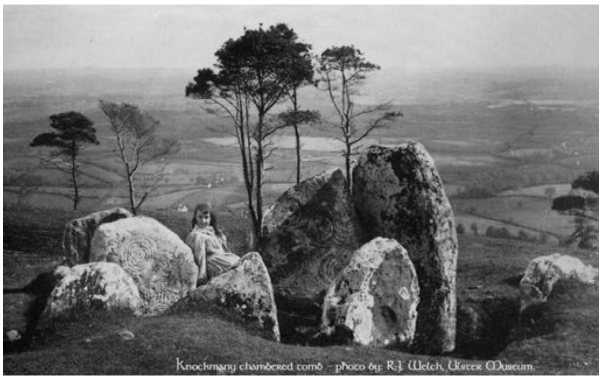
1. Accurate astronomical observations are the key to predictions, and early Babylonian astronomers expanded their knowledge of planetary recurrence cycles to include meteors — leading to the omen literature.
2. Resulting prophecies always took the form "If [astronomical observation] then [terrestrial effect]", e.g. Bjorkman (*Meteoritics*, 8, 91, 1973): *"If a shooting star flashes as bright as a light or as a torch from east to west and disappears on the horizon, then the army of the enemy will be slain in its onslaught"*

What could have motivated these ideas? Seneca (c.4 BC – 65 AD) gives some insight. Referring to the difference between 'us' Romans and the former Etruscans, he remarks, *"... Whereas we believe that lightning is caused by clouds colliding, they believe that clouds collide in order to create lightning. Since they attribute everything to the gods, they are led to believe not that events have a meaning because they have happened, but that they happen in order to express a meaning."*

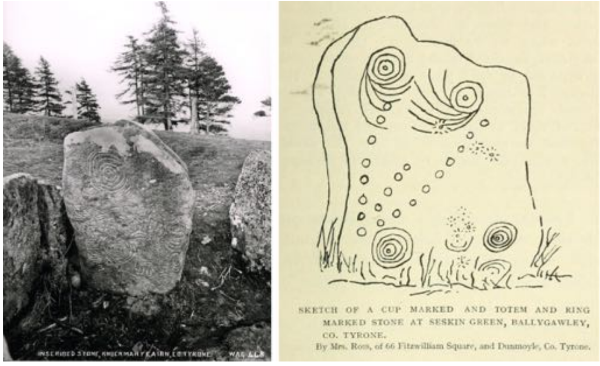
Cradle of Civilization III: Atlantic View / Megalithic Science

Rock Art at Knockmany Chambered Tomb, Co. Tyrone

Knockmany in Nineteenth Century

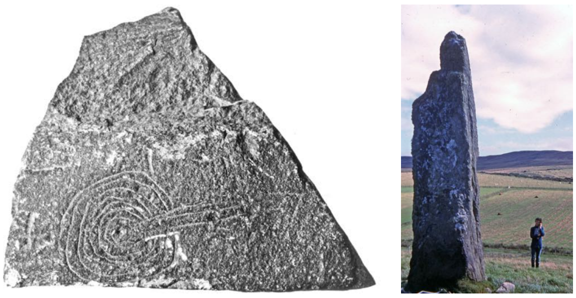


Knockmany and SessKilgreen: Early 20th Century



Images of Knockmany in the 1920s and SessKilgreen around 1910

Rock Art and Megalith in Scotland

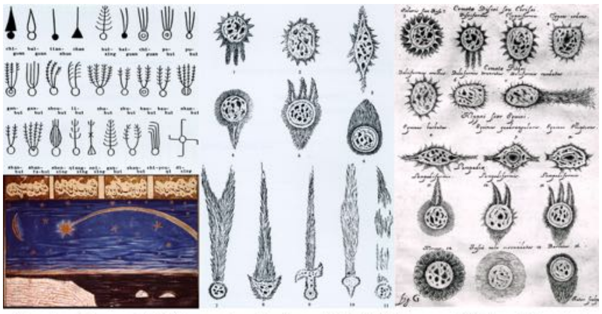


Megalithic markings on a rock from Traprain Law, East Lothian, Scotland; and the huge megalith at Beacharr, Argyle Peninsula, Scotland

Commonly Occurring Motifs in British Rock Art

Frequently occurring designs in British rock art. Left: In Argyll, Scotland (after Ronald Morris,1977). Right: In Northumberland, England (after Stan Beckensall, 1983)

Chinese/Greek/Roman Classification of Comets

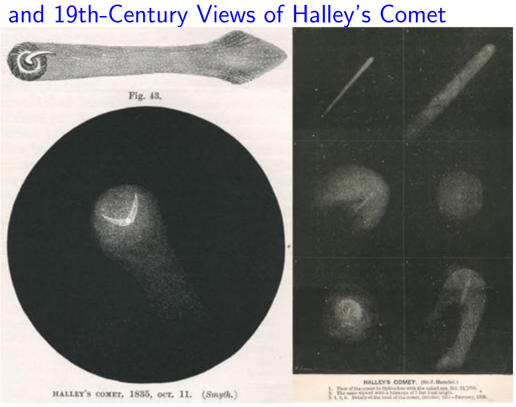


Cometary forms. Left: Chinese classification c.168 BC (Xi Ze-song, 1984) and Comet of 1577 (M. Dizer, Kandili Observatory). Centre and Right: Greek and Roman classification schemes for comets (Hevelius 1668)

Comet Images from Fifteenth to Nineteenth Century

Halley's Comet (674; from Nurenburg Chronicle 1497); Comet Holmes (1892 November 9/16); Donati's Comet 1858 and Great Comet 1861; and Comet 1527

17th and 19th-Century Views of Halley's Comet



Halley's Comet in1683 (Hevelius) and in 1835/36 (Smyth/Herschel)

Modern Clues: Yes, Objects Can Collide With Earth!

Left: The c.10 Mt Tunguska event in Siberia on 1908 June 30 (Kulik), compared with the tree-fall pattern superimposed over London (J. Tate). Right: Sikhote-Alin meteorite (Courtesy Russian Academy of Sciences).

Effects of Impacts: Great and Small

Impacts can produce effects ranging from mass-extinctions of life (e.g. K/T boundary c.65 Myr ago) to just local damage (e.g. Sikhote-Alin meteorite, 1947). They can also lead to new mythology and 'superstition' (e.g. erection at Tunguska ground-zero of totem pole dedicated to Agby: the Siberian god who brings fire to the forest).

Short-Term Implications

Ancient societies appear to be obsessed by the sky:

- e.g. early astronomical interest in 'the sky'; evidence of megalithic monuments/prehistoric 'rock art'.
- Neugebauer: "... ancient 'astrology' can be much better compared with weather prediction from phenomena observed in the sky than with astrology in the modern sense of the word." Suggests knowledge of the direct link between sky and Earth.
- Consistent with more "activity" in the sky in the distant past.

Suggests that some solar-system phenomena may change on much shorter time-scales than we normally consider possible.

Beaghmore Stone Circles, Co. Tyrone, N. Ireland

Image courtesy of and copyright NIEA

Ancient Greek Mysteries Suggest A More Active Night Sky

Ancient Greek "mysteries": Problem of Milky Way ... Zodiacal Light?

- Anaximander: describes stars as like lighted jets of gas spurting out of a punctured hoop of fire.
- Aristotle: believes the Milky Way to lie in the sublunary zone, a hot accumulation of the disintegration products of many comets.
- Anaximander, Parmenides, Leucipus: the 'stars' lie below the Sun and the Moon.
- Metrodorus and Oenopides of Chios: the Milky Way is the former path of the Sun.
- Anaximander and Democritus: the Milky Way lies in the shadow of the Earth.

Image of Milky Way (A. White); Leonid meteor storm; and zodiacal light.

'Why Astronomy?!'

There are three main strands of interest:

1. The broadly cosmological, 'quasi-religious' strand, going back thousands of years — the quest to understand our 'Origins', Man's place in the Universe, Origin of Religion, etc.;
2. The 'practical' strand, i.e. the commercial and economic 'spin-off' from astronomy, including education and the arts — e.g. the calendar; navigation; celestial mechanics; Earth observation; image processing; the 'inspiration' of astronomy and its technical 'spin-off' — including space exploration and national defence, i.e. 'Spaceguard';
3. The strand of pure science or 'Astrophysics' — the project to understand the nature, contents and interactions of all the objects in the entire Universe ...

We live during a rare time: a 'Golden Age' of astronomy, where these three strands have come together as if in conjunction, positively reinforcing each other. ⟹ unprecendented advances in both observation and theory, and with observations almost always leading the latter!

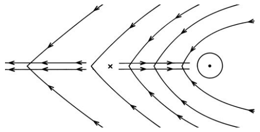
Acknowledgements

Astronomy at Armagh Observatory is funded by the NI Department for Communities, the UK Science and Technology Facilities Council and other grant-awarding organizations

---

Lyttleton's (1948) Accretion Theory

1. A novel variant of the interstellar hypothesis; the first to address both where comets come from and how they are formed.
2. Consider motion of Sun through a dense dust cloud of density $\rho_{\rm dust}$. Collisions of dust grains on axis of symmetry dissipate energy and cause some grains to be captured — these coalesce to become proto-comets.



3. Get inflow within a stagnation radius r_0, approximately the accretion radius $R_A = GM_\odot/V^2$. For $V = 5\,{\rm km\ s^{-1}}$, $R_A \simeq 35\,{\rm AU}$.
4. In a steady-state, the stream mass per unit length is $\mu \simeq 2\pi\rho_{\rm dust}R_A^2$ and the stream velocity V_s is roughly the free-fall speed from R_A. Thus, any new comets have initial semi-major axes $a \lesssim R_A/2$.

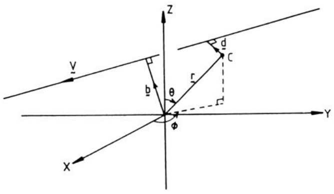
Problems with Lyttleton's Theory

1. Dust clouds do not exist on their own. The interstellar dust is dominated (by a mass fraction of at least a factor of 50) by hydrogen gas. Effects of gas must be included; this was never done.
2. The supposed proto-comets are far too small. Even if an accretion stream could be set up, only very short segments of length $d(r) \lesssim 2\sqrt{\mu r^3/M_\odot}$ at heliocentric distance r could successfully contract against the tidal field of the Sun. This leads to $m_c \lesssim 10^8 (10\,{\rm km\ s^{-1}}/V)^9 (\rho_{\rm dust}/10^{-22}\,{\rm kg\ m^{-3}})^{3/2}\,{\rm kg}$.
3. Inital orbits too short period and too anisotropic. Lyttleton argues for a long period of randomisation of orbits following the last accretion episode, but then the predicted $1/a$-distribution quite wrong (diffusion theory).
4. The supposed proto-comets are on initial orbits directed towards Sun (or solar-system barycentre). All the initial comets will fall onto the Sun, unless inhomogeneities or planetary perturbations are invoked to deflect the stream.

In summary, despite strong advocacy of theory by Lyttleton for next 30 years: "The theory is disproved: an honourable fate for a good theory"!

Survival Problem: Physics of External Perturbations

1. Consider a perturber of mass M passing Sun with velocity $\mathbf{V}$ and impact parameter $\mathbf{b}$ with respect to Sun and $\mathbf{d}$ with respect to a comet at heliocentric distance $\mathbf{r}$.



2. Then the relative velocity change of the comet with respect to the Sun is the difference of the two impulses, i.e.

$$\mathbf{\Delta v} = \frac{2GM}{dV}\hat{\mathbf{d}} - \frac{2GM}{bV}\hat{\mathbf{b}} = \frac{2GM}{bV}\left\{\left(\frac{b^2}{d^2}-1\right)\hat{\mathbf{b}} - \frac{rb}{d^2}\left[\hat{\mathbf{r}} - (\hat{\mathbf{r}}.\hat{\mathbf{V}})\hat{\mathbf{V}}\right]\right\}$$

Heuristic Results

1. Mean relative velocity change in a single encounter is approximately:

$$\Delta v = \frac{2GM}{bV}\begin{cases} \sqrt{2} & b < \sqrt{7/12}\,a \\ \sqrt{7/6}\,a/b & b > \sqrt{7/12}\,a \end{cases}$$

2. On short timescales (e.g. $t \lesssim 30\,{\rm Myr}$), the closest stellar encounter expected during a given time interval t has impact parameter $b_{\rm min} \simeq (2\pi nVt)^{-1/2}$, where n is the number density of perturbers. For stars this usually implies $b \gtrsim a$, which leads to

$$\Delta v_{\rm max} \simeq 4\pi(7/6)^{1/2}\,G\rho a t$$

where $\rho = nM$ ($\approx 0.05\,M_\odot\ {\rm pc^{-3}}$ for stars) is the mass density of perturbers.

3. This leads to $\Delta v_{\rm max} \simeq 4.3\,(a/3\times10^4\,{\rm AU})(t/10\,{\rm Myr})\ {\rm m\ s^{-1}}$.
4. Finally, setting $t = P(a) \simeq 5.2\,(a/3\times10^4\,{\rm AU})^{3/2}\,{\rm Myr}$, the maximum change in perihelion distance during a single revolution can be shown to be of the order of

$$\Delta q \approx 5\,(a/3\times10^4\,{\rm AU})^{7/2}(q/1\,{\rm AU})^{1/2}\quad {\rm AU}$$

Mean Energy Transfer Rate

1. Change in orbital energy in a single encounter: let $\mathbf{\Delta v}$ be the relative velocity change of the comet with respect to the Sun, and let $\mathbf{v_0}$ be its orbital velocity, then

$$\Delta E = \mathbf{v_0}.\mathbf{\Delta v} + \frac{1}{2}(\mathbf{\Delta v})^2$$

Cometary orbital energies thus diffuse and systematically increase (i.e. become less tightly bound) owing to external perturbations.

2. Approximate result for point-mass perturbers: define $a_c = \sqrt{12/7}\,b_{\rm min}$, where $b_{\rm min} \simeq (2\pi nVt)^{-1/2}$ is the most probable minimum impact parameter for the perturbers of number density n, then the mean energy transfer rate can be shown to be approximately

$$\dot{\varepsilon}(t) = \frac{4\pi G^2M^2n}{V}\begin{cases} (a/a_c)^2 & a < a_c \\ 2\ln(a/a_c)+1 & a > a_c \end{cases}$$

Time-Scales For Survival

1. For stars and $t \simeq 4.5\,{\rm Gyr}$, we have $a \gg a_c$. This implies $\dot{\varepsilon}_* \approx {\rm const.}$.
2. For molecular clouds and $t \simeq 4.5\,{\rm Gyr}$, we have $a \ll a_c$, i.e. $\dot{\varepsilon}_c \propto a^2$.
3. The net result is:

$$\dot{\varepsilon} = \dot{\varepsilon}_* + \dot{\varepsilon}_c \simeq A_* + A_c a^2$$

where for typical parameters $A_* \simeq 10^{-13}\,{\rm m^2s^{-3}}$ and $A_c \approx 10^{-44}\,{\rm s^{-3}}$.

4. Solving the energy evolution equation for each type of perturber leads to the half-life due to stellar and molecular cloud perturbations, i.e.

$$t_{1/2,*} = \frac{1}{4.732}\frac{GM_\odot}{A_* a} \simeq 2\times10^9\left(\frac{2\times10^4\,{\rm AU}}{a}\right)\ {\rm yr}$$

and

$$t_{1/2,c} = \frac{1}{8.190}\frac{GM_\odot}{A_c a^3} \simeq 2\times10^9\left(\frac{2\times10^4\,{\rm AU}}{a}\right)^3\ {\rm yr}$$

5. Thus, due to both clouds and stars, the majority of comets with initial $a \gtrsim 2\times10^4\,{\rm AU}$ will be lost. This is the Oort cloud survival problem.

Summary of Survival Problem: Oort Cloud Evolution

1. Two main types of external perturber: stars and molecular clouds.
   - Galactic tide also drives comets into inner solar system, but has little direct effect on Oort cloud's disruption.
2. Stars pass through and beyond the Oort cloud, causing gradual unbinding of cometary orbits; the 'stellar' half-life is $t_{1/2,*} \simeq 2\times(2\times10^4\,{\rm AU}/a)$ Gyr.
3. Molecular clouds pass beyond the Oort cloud, but are much more massive than stars; the 'molecular cloud' half-life is $t_{1/2,c} \simeq 2\times(2\times10^4\,{\rm AU}/a)^3$ Gyr

$\Longrightarrow$ 'standard Oort cloud dynamically unstable beyond $a \simeq 2\times10^4\,{\rm AU}$, over the age of the solar system (4.5 Gyr)

The Oort cloud is a leaky reservoir which must be replenished from within, possibly the trans-Neptunian region or a Dense Inner Core.