Saas-Fee Advanced Course 35

D. Jewitt A. Morbidelli H. Rauer

# Trans-Neptunian Objects and Comets

Saas-Fee Advanced Course 35

Swiss Society for Astrophysics and Astronomy
Edited by K. Altwegg, W. Benz and N. Thomas

With 132 Figures, 18 in Color

Springer

David Jewitt
University of Hawaii
Institute for Astronomy
2680 Woodlawn Drive
Honolulu, HI 96822, USA
jewitt@hawaii.edu

Alessandro Morbidelli
Observatoire de la Côte
dŠAzur/CNRS
B.P. 4229
06304 Nice Cedex 4, France
morby@obs-nice.fr

Heike Rauer
DLR/Institut für Planetenforschung
Rutherfordstr. 2
12489 Berlin, Germany
and
TU Berlin/Zentrum
für Astronomie und Astrophysik
Hardenbergstr. 36
10623 Berlin, Germany
heike.rauer@dlr.de

*Volume Editors:*
Kathrin Altwegg
Willy Benz
Nicolas Thomas
Universität Bern
Physikalisches Institut
Sidlerstrasse 5
3012 Bern, Switzerland

This series is edited on behalf of the Swiss Society for Astrophysics and Astronomy:
Société Suisse d'Astrophysique et d'Astronomie
Observatoire de Genève, ch. des Maillettes 51, 1290 Sauverny, Switzerland

*Cover picture:* See chapter by D. Jewitt, Fig. 12.

Library of Congress Control Number: 2007934029

ISBN 978-3-540-71957-1 Springer Berlin Heidelberg New York



Springer is a part of Springer Science+Business Media
springer.com


Typesetting: by the authors and Integra using a Springer LaTeX macro package
Cover design: WMXDesign GmbH, Heidelberg

Printed on acid-free paper SPIN: 11808169 55/Integra 5 4 3 2 1 0

# Preface

The 35th Saas Fee Winter School was held on 13–18 March 2005 in the skiing village of Mürren in the Berner Oberland. In view of the excitement generated over the past 15 years by the discovery of the Kuiper Belt and Trans-Neptunian Objects and also by the ongoing Rosetta mission to comet Churyumov-Gerasimenko, it was decided to combine discussion of these primitive objects into one winter school under the title, "Trans-Neptunian Objects and Comets." The aim was to provide an overview of these objects, to discuss their relationships, and to identify directions for future research. The school attracted over 60 students from all over the world. We were fortunate that not merely were the students able to hear a set of outstanding lectures but were also able to enjoy marvellous weather in one of the most beautiful parts of Switzerland.

The organizers thank the lecturers, Dave Jewitt, Alessandro Morbidelli, and Heike Rauer, for the tremendous effort they made in preparing the lectures and the text for this volume. Stephan Graf, Annette Jäckel, and Jonathan Horner provided reviews, checked the text and references, and assisted in the production. We also thank Frau Staehli and the staff of the Hotel Eiger in Mürren for the warm welcome and their generosity. We also thank Ms. Kathrin Weyeneth and Ms. Edith Hertig from the Physikalisches Institut for their secretarial support for the school.

Financial assistance provided by the Swiss Society for Astrophysics and Astronomy and the European Space Agency is gratefully acknowledged.

*Kathrin Altwegg*
*Willy Benz*
*Nicolas Thomas*

# Contents

# List of Previous Saas-Fee Advanced Courses

!! 2005 Trans-Neptunian Objects and Comets
*D. Jewitt, A. Morbidelli, H. Rauer*

!! 2004 The Sun, Solar Analogs and the Climate
*J.D. Haigh, M. Lockwood, M.S. Giampapa*

!! 2003 Gravitation Lensing: Strong, Weak and Micro
*P. Schneider, C. Kochanek, J. Wambsganss*

!! 2002 The Cold Universe
*A.W. Blain, F. Combes, B.T. Draine*

!! 2001 Extrasolar Planets
*T. Guillot, P. Cassen, A. Quirrenbach*

!! 2000 High-Energy Spectroscopic Astrophysics
*S.M. Kahn, P. von Ballmoos, R.A. Sunyaev*

!! 1999 Physics of Star Formation in Galaxies
*F. Palla, H. Zinnecker*

!! 1998 Star Clusters
*B.W. Carney, W.E. Harris*

!! 1997 Computational Methods for Astrophysical Fluid Flow
*R.J. LeVeque, D. Mihalas, E.A. Dorfi, E. Müller*

!! 1996 Galaxies Interactions and Induced Star Formation
*R.C. Kennicutt, F. Schweizer, J.E. Barnes*

!! 1995 Stellar Remnants
*S.D. Kawaler, I. Novikov, G. Srinivasan*

!! 1994 Plasma Astrophysics
*J.G. Kirk, D.B. Melrose, E.R. Priest*

!! 1993 The Deep Universe
*A.R. Sandage, R.G. Kron, M.S. Longair*

!! 1992 Interacting Binaries
*S.N. Shore, M. Livio, E.J.P. van den Heuvel*

!! 1991 The Galactic Interstellar Medium
*W.B. Burton, B.G. Elmegreen, R. Genzel*

!! 1990 Active Galactic Nuclei
*R. Blandford, H. Netzer, L. Woltjer*

* 1989 The Milky Way as a Galaxy
*G. Gilmore, I. King, P. van der Kruit*

! 1988 Radiation in Moving Gaseous Media
*H. Frisch, R.P. Kudritzki, H.W. Yorke*

! 1987 Large Scale Structures in the Universe
*A.C. Fabian, M. Geller, A. Szalay*

! 1986 Nucleosynthesis and Chemical Evolution
*J. Audouze, C. Chiosi, S.E. Woosley*

! 1985 High Resolution in Astronomy
*R.S. Booth, J.W. Brault, A. Labeyrie*

! 1984 Planets, Their Origin, Interior and Atmosphere
*D. Gautier, W.B. Hubbard, H. Reeves*

! 1983 Astrophysical Processes in Upper Main Sequence Stars
*A.N. Cox, S. Vauclair, J.P. Zahn*

* 1982 Morphology and Dynamics of Galaxies
*J. Binney, J. Kormendy, S.D.M. White*

! 1981 Activity and Outer Atmospheres of the Sun and Stars
*F. Praderie, D.S. Spicer, G.L. Withbroe*

* 1980 Star Formation
*J. Appenzeller, J. Lequeux, J. Silk*

* 1979 Extragalactic High Energy Physics
*F. Pacini, C. Ryter, P.A. Strittmatter*

* 1978 Observational Cosmology
*J.E. Gunn, M.S. Longair, M.J. Rees*

* 1977 Advanced Stages in Stellar Evolution
*I. Iben Jr., A. Renzini, D.N. Schramm*

* 1976 Galaxies
*K. Freeman, R.C. Larson, B. Tinsley*

* 1975 Atomic and Molecular Processes in Astrophysics
*A. Dalgarno, F. Masnou-Seeuws, R.V.P. McWhirter*

* 1974 Magnetohydrodynamics
*L. Mestel, N.O. Weiss*

* 1973 Dynamical Structure and Evolution of Stellar Systems
*G. Contopoulos, M. Hénon, D. Lynden-Bell*

* 1972 Interstellar Matter
*N.C. Wickramasinghe, F.D. Kahn, P.G. Metzger*

* 1971 Theory of the Stellar Atmospheres
*D. Mihalas, B. Pagel, P. Souffrin*

---

* Out of print
! May be ordered from Geneva Observatory
Saas-Fee Courses
Geneva Observatory
CH-1290 Sauverny
Switzerland
!! May be ordered from Springer and/or are available online at springerlink.com.

# Kuiper Belt and Comets: An Observational Perspective

D. Jewitt

***Note to the Reader***
*These notes outline a series of lectures given at the Saas Fee Winter School held in Mürren, Switzerland, in March 2005. As I see it, the main aim of the Winter School is to communicate (especially) with young people in order to inflame their interests in science and to encourage them to see ways in which they can contribute and maybe do a better job than we have done so far. With this in mind, I have written up my lectures in a less than formal but hopefully informative and entertaining style, and I have taken a few detours to discuss subjects that I think are important but which are usually glossed-over in the scientific literature.*

## 1 Preamble

Almost exactly 400 years ago, planetary astronomy kick-started the era of modern science, with a series of remarkable discoveries by Galileo concerning the surfaces of the Moon and Sun, the phases of Venus, and the existence and motions of Jupiter's large satellites. By the early 20th century, the focus of astronomical attention had turned to objects at larger distances, and to questions of galactic structure and cosmological interest. At the start of the 21st century, the tide has turned again. The study of the Solar system, particularly of its newly discovered outer parts, is one of the hottest topics in modern astrophysics with great potential for revealing fundamental clues about the origin of planets and even the emergence of life. New technology has been crucial to each of these steps. Galileo's refractor gave a totally new view of the sky. A hundred years ago, photographic plates and large telescopes allowed the first spectroscopic observations of distant galaxies revealing, through Hubble's law, the third dimension of distance into the plane of the sky. In our own time, highly sensitive, wide-field electronic detectors have enabled the discovery and the exploration of the Kuiper Belt, while fast computers allow us to make numerical simulations with a degree of sophistication that was previously unimaginable.

D. Jewitt, *Kuiper Belt and Comets.* In: K. Altwegg et al., Trans-Neptunian Objects and Comets, Saas-Fee Advanced Courses, pp. 1–78 (2008)
DOI 10.1007/978-3-540-71958-8_1

As a result of all this, our view of the Solar system is in the middle of a great change. Our appreciation of the different types of objects (planets, asteroids, comets, etc) orbiting the Sun is changing in response to new observations. Our understanding of their evolutionary connections with each other and with the formation epoch is changing as we develop more and more elaborate schemes to synthesize the new data. Additionally, our perception of the Solar system in the bigger context of the galactic disk is changing, particularly as we detect planets encircling other stars (in systems that are, for the most part, dynamically not very like our own). All of this makes it a great time to review what we know about the Solar system in the context of the Saas Fee winter school series, one of very few Saas Fee lectures to be dedicated to the universe at $z \sim 0$.

This article parallels five lectures given in Mürren, Switzerland, in March 2005, as part of the Saas Fee Lecture Series entitled "Trans-Neptunian Objects and Comets." Some of these lectures were given "off the cuff," and I have tried to reconstruct them from memory and a few notes. The degree to which this succeeds is unknown and it does not matter: the participants, like this lecturer, have no doubt forgotten most of what was said while readers who were not in Saas Fee for the Lecture Series never knew. The style of the write-up is deliberately informal.

### 1.1 The Conduct of Research into the Subject

In this section, I want to take advantage of the open format of the Saas Fee lecture series to briefly discuss the conduct of modern science, particularly as it relates to the new study of the Solar system. Partly, this is for fun and for my own entertainment, but I also have a serious purpose: there are real misconceptions about what is happening (as opposed to what should happen), sometimes even in the minds of the best scientists. Most of us probably possess vaguely Popperian [124] notions about the conduct of science. Essentially, Popper argued that we advance in science by the falsification of hypotheses. Observations suggest hypotheses that make predictions, which can be confirmed or refuted by new observations, and so on. But not all of us work within this framework, and there are few clues as to the real methods or motivations of scientists in the stylized and frequently dry presentations that are demanded for publication in the refereed journals. It is the absence of discussion about the realities of the practice of science that has allowed false ideas to spread unchecked. The Saas Fee participants, especially those likely to become major figures in the future exploration of the Solar system, are the main targets of my remarks.

#### Observations

Observationally, the goal is to determine objective reality through careful studies that are unbiassed (or at least well calibrated), fully understood, independently reproducible and motivated by the desire to test a hypothesis. Several things must be said about this idealized goal.

– Real science is much more affected by chance discoveries than one would guess from the simple description of Popper's scheme, above. Sometimes, the biggest advance comes from simply looking, not from testing a hypothesis.
– The flip-side of this is that the human brain is rarely able to perceive or assimilate things that it does not expect to see, and so, fundamental discoveries made by chance are very rare (but disproportionately important). We are like ants in the city: comfortable with the dirt in front of us but unable to perceive the buildings above.
– Although it seems that it should be otherwise, taking good observations is incredibly hard. Too many things can go wrong; there are many sources of error both random and systematic, and it is often difficult or impossible to accurately quantify these uncertainties. As a result, observations that seem secure (or "statistically significant" as we say with a misleading air of detachment) are often wrong, leading us up blind alleys that can take years to escape.
– An equally serious problem is that it is easy to take the "wrong" measurement, by which I mean a measurement that has no great impact on our perception of the big picture. In fact, most observers, including this one, spend most of their time taking measurements that are unimportant. The simple reason is that we usually cannot see clearly enough to predetermine which measurements will be of the greatest value. Theories and models are supposed to help us here: usually they do not.

As observers, we are swimming in mud (Fig. 1): it is hard work, we cannot see where we are going but sometimes we bump into interesting things as we crawl our way along.

## Theories and Models

The purpose of theories and models is to use available data together with established physical laws to make observationally testable predictions. Predictions provide an objective and indispensable way to test the theories and models. Unfortunately, theory rarely works this way, because the systems under consideration are very complicated and a large number of processes interact in a way that is difficult to treat. Making observationally testable predictions is difficult because a given model, with changes to one or two of its many free parameters, can usually accommodate a wide range of outcomes, regardless of whether the model is correct. Making predictions that are falsifiable is the hard part of making models, which is why many modelers do not do it.

Here are some problems with theory and theorists.

**Fig. 1.** Observers, doing what they do best. Photo courtesy Talisman Creations

- The main problem for theorists and modelers is that the world is very complex, and most problems are observationally under-constrained. Analytical approaches offer real insight and understanding but are mostly confined to the study of highly simplified approximations. Numerical approaches provide a way to deal with the complexity, but at the expense of adding typically large numbers of under-determined model parameters and initial conditions.
- It has become common to present models that fit the available data but which offer no observationally testable predictions, leaving the reader to speculate about what predictions the model might make if only the authors had written them down. The reason for this is clear enough: making observationally testable predictions is difficult (and scary too: you could be wrong!). But without predictions the models have no scientific value. Some have argued that the mere fact that a model can fit many and varied observations in a self-consistent way is evidence in itself for the correctness of the model. Nonsense!
- The meaning of the word "predict" is also under attack. Sometimes, the authors say that their model "predicts" some quantity or property, but closer inspection shows that the thing has already been measured. One cannot predict something which is already known! What the modelers mean is that they can fit the data, not predict new data. There is a big difference.
- Models are frequently over-sold Fig. 2. It is almost *de rigueur* for modelers to add comforting phrases like "our conclusions are insensitive to the parameters assumed in the model..." and "our model has only one free parameter..." whether or not these statements are true!

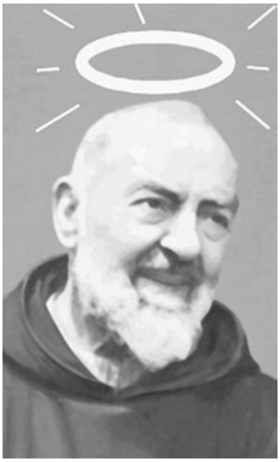

**Fig. 2.** The theorist, spotlessly clean, whose theory explains everything and has no free parameters. The halo and the facial expression signify his wisdom and purity. Courtesy Virginia A. Tikan

Of course, it is the interaction between the observers and the theorists that gives our subject its extraordinary vitality and power. Science without observations would collapse into dull paralysis within months. Science without models would soon degenerate into stamp collecting. But this does not mean that we have to accept either the observations or the models uncritically. In particular, we should not accept models that fail to make observationally testable predictions. They may offer beautiful descriptions of what we observe but, without predictions, we will never know if they have deeper meaning.

The Kuiper belt is still very much in the discovery phase, and we should not expect a scientifically compelling picture of its formation and evolution to emerge overnight. With this warning of a turbulent and uncertain background, we are ready to launch into an overview of the modern Solar system.

# 2 The Modern Solar System

## 2.1 Protoplanetary Disk

### Scale Constraints

The most noticeable feature of the Solar system is that the planets follow nearly circular orbits about the Sun in roughly the same plane. This architecture strongly suggests that the planets formed by accretion in a circum-Solar disk. The properties of this disk, now long-gone, can be inferred only approximately from the modern-day system.

The extent of the solar nebula is not tightly observationally constrained, but again we can set limits. At the inner edge, it is reasonable to suppose that the disk extended practically to the surface of the protosun. Indeed, material flowed through the disk into the Sun as part of its formation. At the outer edge, we surmise that the disk extended to roughly the outer extremity of the well-established part of the Kuiper belt (roughly 50 AU). Observations of disks around other stars show that disks are commonly much larger, extending to hundreds of AU around stars of Solar mass. The timescales for the growth of solid bodies scale with heliocentric distance, $R$, as $R^3$, give or take one power of $R$. One possibility is that the protoplanetary disk may initially have been hundreds of AU in extent but that no large bodies grew in the outer parts. In this case, deeper survey observations should reveal smaller bodies beyond the $\sim$50 AU edge, something that seems not to be true. Another possibility is that the small size of the Kuiper belt (specifically of the classical belt) results from tidal truncation by a passing star, as argued by Ida et al. [66] and others since.

### Structure Constraints

The current mass of the objects in the Solar system (excluding the Sun) is about $10^{-3}$ $M_\odot$, most of which is in Jupiter. Obviously, this sets a strong lower limit to the initial mass of the disk. A more realistic limit is set by careful consideration of the compositions of the planets and the (probably good) assumption that the disk had a basically cosmic composition. For instance, consider the Earth. Its mass consists mostly of heavy elements (called "metals" by terminology-bending astrophysicists), whereas, in a mixture containing a cosmic proportion of H and He, the "metals" carry only $\sim$0.01 of the mass. Therefore, the so-called augmented mass of the Earth (the mass of material of cosmic composition containing an Earth mass of metals) is about 100 $M_\oplus$. This same treatment of the other planets leads to a best estimate of the minimum disk mass of order 0.01 $M_\odot$. Models with this mass are known as MMSN models: Minimum Mass solar nebula models.

The distribution of mass and temperature within the protoplanetary disk are usually approximated by power laws

$$\Sigma(R) = \Sigma(R_0)\left[\frac{R_0}{R}\right]^p \tag{1}$$

$$T(R) = T(R_0)\left[\frac{R_0}{R}\right]^q \tag{2}$$

where $\Sigma(R)$ [kg m$^{-2}$] and $T(R)$ are the column density and temperature of the disk at radius $R$, $R_0$ is a reference radius, often taken as 1 AU (the orbit of Earth) or 10 AU (orbit of Saturn), and the indices $p$ and $q$ describe the radial fall-off of the density and temperature, respectively. Estimates of $\Sigma_0$ and $p$ can be obtained by studying the distribution of mass within the Solar system. If we smear the augmented masses of the planets over annuli extending half way to the nearest planet (e.g., Saturn would be smeared from 7.5 to 15 AU) we obtain $p \sim 3/2$ (with an uncertainty of at least $\pm 1/2$) and $\Sigma(R_0) \sim 50\,\mathrm{kg\,m^{-2}}$ at $R_0 = 10\,\mathrm{AU}$. This is the total (gas plus dust) surface density. The dust surface density is about 100 times smaller. The temperature of a blackbody in radiative equilibrium with sunlight is described by (2) with $T(R_0 = 10) = 88\,\mathrm{K}$ and $q = 0.5$.

The values of disk parameters so derived are not particularly accurate, given the uncertainties in computing augmented masses from current masses and given the likelihood that the orbits of the planets were not always where we now find them. Still, the above give a reasonable starting guess for the structure of the disk. It is natural to think that observations of disks around young stars should provide independent constraints on likely disk parameters. Unfortunately, most existing data generally lack angular resolution high enough for the disk spatial parameters to be directly measured. Instead, the disk parameters are inferred from measurements of the spectral energy distribution using models in which the number of free parameters is larger than the number of observational constraints. Assuming $p = 3/2$, measurements give mean values $q = 0.6 \pm 0.1$ and $T(10\,AU) = 45 \pm 21\,\mathrm{K}$ [4], which fit well with the nominal values. The dust mass inferred from disk observations averages $M_\mathrm{d} = 4 \times 10^{-3}\,\mathrm{M}_\odot$ ( [4]; from 67 classical T-Tauri stars, likely analogs of the young Sun). The dust mass is really a lower limit to the mass in solids: particles much larger than the millimeter wavelengths of observation contribute little to the measured radiation and go undetected. Augmented to cosmic composition, the implied average disk mass is $\sim 0.4\,\mathrm{M}_\odot$. This is substantially larger than MMSN but the scatter in disk masses is large, as are the uncertainties, and there are presumably observational biases against the measurement of lower disk masses.

## Constraints on Disk Timescales and Environment

The most important observational constraints on timescales in the protoplanetary disk are provided by measurements of the products of radioactive decay of short-lived elements in meteorites. The latter are rocks derived by shattering collisions amongst the asteroids and delivered to Earth by gravitational scattering after their orbits become planet-crossing. Minerals in many meteorites incorporate the decay products of short-lived nuclei, showing that the minerals formed on timescales comparable to the half-lives of the decaying elements. The quintessential example is provided by $^{26}$Al, which $\beta$-decays into $^{26}$Mg with a half-life $t_{1/2} = 0.7\,\mathrm{Myr}$ [90]. When $^{26}$Mg is found incorporated within the mineral structure of a meteorite, we may conclude that $^{26}$Al was originally present. To be captured in abundance, $^{26}$Al must have been incorporated into the meteorites within a few half-lives of its formation. Element formation occurs naturally in the explosion of massive stars as supernovae, but the significance of $^{26}$Al has sometimes been questioned because it can be also formed by spallation reactions with particles accelerated to energies >MeV [91]. Such particles might have been emitted by the magnetically super-active young sun. Recent measurements of $^{60}$Ni, which is produced by the decay of $^{60}$Fe with a half-life of 1.5 Myr [116], do not suffer this ambiguity because there is no route to its production through spallation. We conclude with confidence that macroscopic solid bodies formed in the asteroid belt on timescales of a few Myr.

Other timescale constraints come from observations of circumstellar matter in disks around nearby Solar-mass stars. These observations show that circumstellar gas has a lifetime that is less than 10 Myr [10, 161] and potentially just a few Myr. Dust emission from stars also declines rapidly with age (Fig. 3). The initial decline is probably due to growth into particles that are much larger than the wavelength of observation (typically ∼1 mm). There is evidence for thermal excess above the emission from the stellar photospheres in stars as old as ∼0.5 Gyr, and this dust is probably produced in recent times by collisions among unseen bodies in the circumstellar disks, or released by unseen comets. The general decline in the dustiness of nearby stars is occasionally punctuated by objects with surprising dust emission excess. This could be showing that the stars are, for some reason, intrinsically more dusty than others of similar age. An alternative explanation is that the dust has been

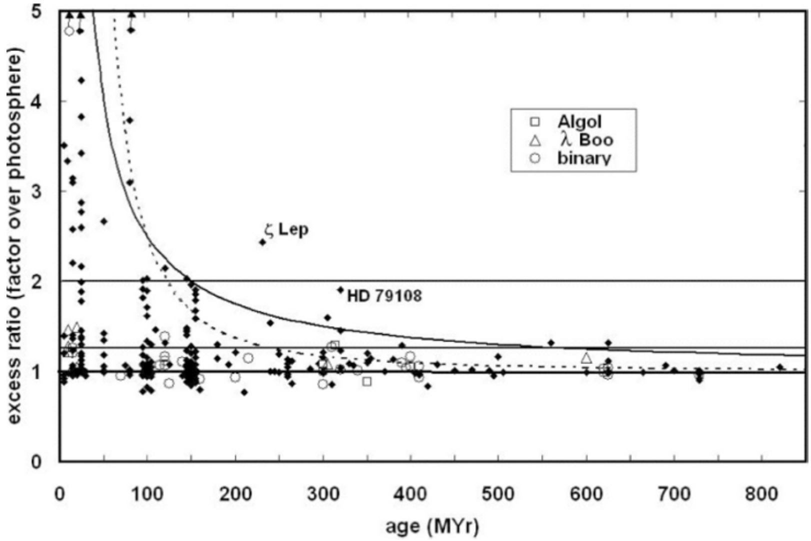


**Fig. 3.** Dust emission from nearby stars at 24 $\mu$m wavelength expressed as a ratio to the flux density expected from the photosphere alone. Values >1 indicate excess emission, most likely from circumstellar dust heated by starlight. The emission generally declines with stellar age, but, at any given age, there is a range of thermal excesses, with occasional dramatic spikes, as at $\zeta$ Lep and HD 79108. The solid curve shows a 1/(time) dependence. Ages of the stars are estimated from cluster membership and from models of their spectra, and are accurate to about a factor of two. One interpretation of the spikes is that dust is impulsively created by collisions between massive bodies. Figure reproduced from [131]

recently created, perhaps by impact and shattering of massive planetesimals in the unseen circumstellar disks [131].

Two pieces of evidence suggest that the Sun formed in a star cluster.

First, some of the short-lived radionuclides (notably $^{60}$Fe) must have been produced, in an exploding star, only shortly before their incorporation into minerals and meteorite parent bodies (asteroids), otherwise, they would have already decayed to insignificance. Supernovae are very rare (the galactic rate is only $\sim$one per 50 years) and typically distant so that the likelihood of having one occur nearly simultaneously with the formation of solid bodies in the disk is small. The simplest interpretation is that the Sun was part of a cluster of stars in which nearby high mass members exploded upon reaching the ends of their stable main-sequence lifetimes. An estimate of the cluster population can be made based on the dual requirements that the cluster must have been populated enough to contain a massive star capable of reaching supernova status but yet not so populated that gravitational perturbations would have noticably disturbed the orbits of the planets. A cluster containing $\sim$2000$\pm$1100 stars seems capable of meeting both conditions [2].

Second, the truncated outer edge of the classical Kuiper belt and the excited dynamical structure of the belt in general suggests to some that the protoplanetary disk might have been tidally truncated by a passing star [66, 114]. Numerical simulations show that to truncate or seriously disturb the disk down to radius $r$ [AU] implies a stellar impact parameter $\sim 3\,r$. The classical belt ends near 50 AU, requiring a Solar mass star to pass $\sim$150 AU from the Sun. In its current environment, the sun and stars are separated by $\sim$1 pc (200,000 AU), and the probability of two stars passing within 150 AU in the 4.6 Gyr age is negligible. Again, a plausible inference is that the mean distance between the Sun and nearby stars was once much smaller: the Sun was in a cluster.

## 2.2 The Three Domains

It is useful to consider the Solar system as divided into three domains, based on the compositions, masses, and radial distances of its constituents. These are as follows:

### The Domain of the Terrestrial Planets

The primary objects are Mercury, Venus, Earth, and Mars, but the asteroids in the main-belt between Mars and Jupiter are also included (the largest asteroid is (1) Ceres; see Table 1). These objects are all distinguished by refractory (non-volatile) compositions dominated by metals [principally iron (Fe) and nickel (Ni)] and compounds of silicon (Si), oxygen (O), magnesium (Mg), and aluminium (Al). The bulk densities are high ($\rho = 3930\,\mathrm{kg\,m^{-3}}$ for Mars up to $5515\,\mathrm{kg\,m^{-3}}$ for Earth, the latter slightly enhanced by self-compression due to gravity), reflecting the lack of volatiles. Densities of many

**Table 1.** Terrestrial Planets

| Object | Mass/$M_\oplus$ | Radius/$R_\oplus$ | $\rho$ [$\mathrm{kg\,m^{-3}}$] | a [AU] | e | i [deg] |
|---|---|---|---|---|---|---|
| Mercury | 0.06 | 0.38 | 5430 | 0.387 | 0.206 | 7.0 |
| Venus | 0.81 | 0.95 | 5424 | 0.723 | 0.007 | 3.4 |
| Earth | 1 | 1 | 5520 | 1.000 | 0.017 | 0.0 |
| Mars | 0.11 | 0.53 | 3930 | 1.523 | 0.093 | 1.8 |
| Ceres | $1.6\times10^{-4}$ | 0.08 | 2080 | 2.766 | 0.078 | 10.6 |

asteroids are smaller, apparently because of porous internal structures created by impact fragmentation and reassembly of these bodies since their formation. The densities of stony meteorites, small fragments from the asteroid belt, are $\rho \sim 3000\,\mathrm{kg\,m^{-3}}$.

All these bodies appear to have formed by "binary accretion," the step-by-step growth occurring when two bodies collide and stick, starting from tiny dust particles in the original nebula about the Sun and reaching up to the sizes of the Earth and Venus. Indeed, the N-body models that are used to study the dynamics and growth of bodies in the outer Solar system have been honed to their highest levels of perfection in the study of terrestrial planet growth. Still, new data continue to surprise and unnerve us. For example, N-body accretion models show that Earth grew to its final mass on a timescale $\sim$100–200 My [18, 129], and this long timescale has remained more or less unchanged for the past several decades, since detailed estimates were first made by G. Wetherill. It stands in contrast to new isotopic data from the Hafnium-Tungsten (Hf-W) decay [67]. Hafnium decays to Tungsten, $^{182}\mathrm{Hf} \rightarrow {}^{182}\mathrm{W}$, with a 9- Myr half life. The quantity of $^{182}\mathrm{W}$ in the Earth's mantle (relative to the core) provides a measure of the amount of the unstable Hf isotope at the epoch of core formation, and so sets the timescale for Earth's differentiation. The W-Hf data show that the Earth accreted the bulk of its mass within 30 Myr, whereas major asteroids such as Vesta formed in an even shorter 3 Myr [67]. This is a half to one order of magnitude discrepancy with the N-body models and remains unexplained.

The relevance to us is that models can give very plausible but wholly incorrect solutions. Without the benefit of independent constraints from the isotopes, we would remain completely unaware that the N-body terrestrial planet growth models are too slow. In the outer Solar system (where independent constraints on the models from isotopes or other sources are unavailable), it is easy to see that we are skating on very, very thin ice.

## The Domain of the Giant Planets

### Gas Giants

Jupiter and Saturn (Figs. 4 and 5), in addition to being two orders of magnitude more massive than the Terrestrial planets (see Table 2), have very differ-

**Fig. 4.** Gas giant Jupiter from the Galileo spacecraft, showing its banded cloud structure and the Great Red Spot. Image from NASA

ent, much more volatile-rich compositions. Jupiter and Saturn are mass-wise dominated by hydrogen ($H_2$) and helium (He) and are known as "gas giants."

The formation of the giant planets is imperfectly understood. Prevailing ideas suggest that, in the Solar system, the gas giant planets formed by a process of nucleated instability, a bit like a rain drop forming by condensation of water molecules on a refractory aerosol. The model was developed by Mizuno and others [111,123]. Briefly, solid bodies collide and grow by binary accretion in the protoplanetary disk, much as they did in the domain of the Terrestrial planets. Upon reaching a critical mass, generally estimated to be $\sim$10 $M_{\oplus}$, the core precipitates the infall of surrounding nebular gas, producing a hydrody-

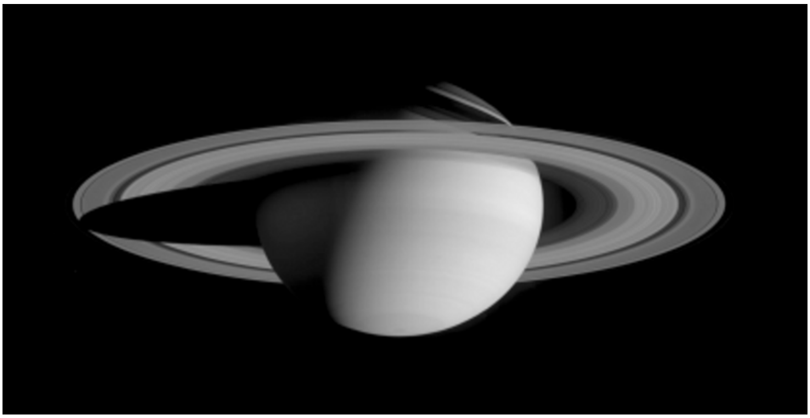

**Fig. 5.** Gas giant Saturn from the Cassini spacecraft. Courtesy NASA

**Table 2.** Giant Planets

| Object | Mass/$M_\oplus$ | Radius/$R_\oplus$ | $\rho$ [kg m$^{-3}$] | a [AU] | e | i [deg] |
|---|---|---|---|---|---|---|
| Jupiter | 316 | 11.21 | 1330 | 5.203 | 0.048 | 1.3 |
| Saturn | 95 | 9.45 | 700 | 9.537 | 0.054 | 2.5 |
| Uranus | 14.5 | 4.01 | 1300 | 19.191 | 0.047 | 0.8 |
| Neptune | 16.6 | 3.88 | 2300 | 30.068 | 0.009 | 1.8 |

namic flow that results in very rapid mass growth of the planet. As the planet mass undergoes a runaway growth, tidal torques exerted by the planet on the protoplanetary disk open a "gap" around the orbit of the planet. Subsequent mass in-flow to the planet continues at a reduced rate.

Growth by nucleated instability clearly involves two distinct timescales. First, the core must grow to critical mass. Second, the nebular gas must be accreted by the core. Core growth, which occurs by binary accretion as for the terrestrial planets, is the slower process. It is the principal cause of concern with the nucleated instability model and so has been the subject of much attention. The key issue is that the core must grow on a timescale that is short compared with the timescale for the dissipation of the gas nebula. Observations of young stars with dusts disks generally fail to reveal attendant gas, leading to the inference that the gas is quickly removed, probably on timescales of a few Myr for sun-like stars and almost certainly on timescales $<$10 Myr [10]. This sets an upper limit to the core growth times and is a primary challenge to the core accretion model. One way in which core growth might have been accelerated is through an increase in the disk column density just beyond the snow-line, owing to the extra mass in solids added by the freeze-out of nebular water vapor [20]. Million year growth times at the orbit of Jupiter are not hard to obtain from current models, but more work is needed to induce Uranus and, especially, Neptune to grow on cosmically reasonable timescales.

A different giant planet growth scenario has been proposed in which the "slow step" of core accretion is side-stepped. In this model, the protoplanetary disk is supposed to have been intrinsically unstable to collapse under its own gravity. Disk instabilities clearly favor higher than MMSN disks (models typically assume disk masses $\sim$10 times the MMSN in order to produce spontaneous collapse), but even MMSN models have been reported to be susceptible to collapse under some circumstances [8]. Formation of giant planets by spontaneous collapse does not suffer the timescale problem of the nucleated instability model (because there is no need to wait for a nucleus to form), but there are other problems related to the long-term stability of the collapsing planet. Investigators differ on this issue. The differences are not fully understood, but might relate to the accuracy with which cooling processes are represented [14].

Neither core accretion nor nebula collapse predicted the over-abundance of heavy elements measured in Jupiter by the Galileo entry probe ( [120], see

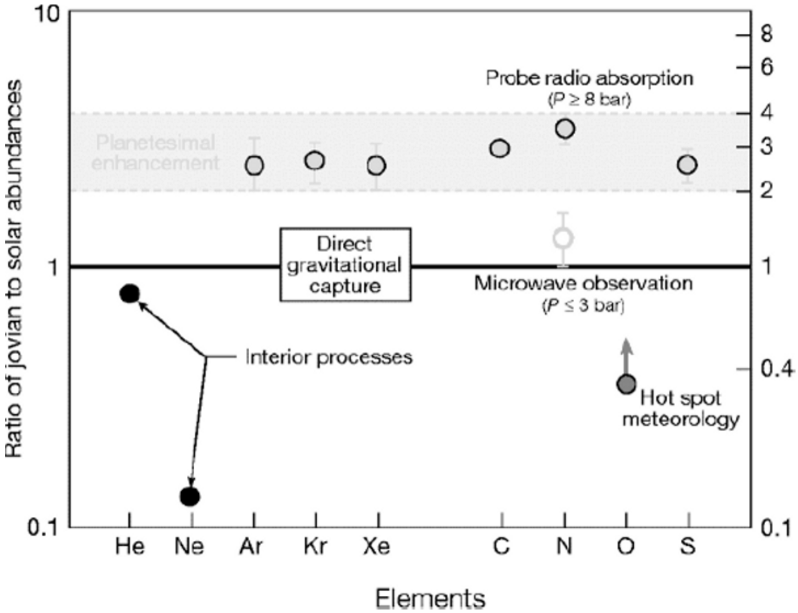


**Fig. 6.** Metal abundances in Jupiter relative to those in the Sun, as measured by the Galileo entry probe. Helium and Neon are low in abundance because they are partly dissolved in the metallic hydrogen core. Oxygen is low, probably because the probe entered Jupiter's atmosphere at an (unrepresentative) hot-spot location, where conditions were atypically dry. The other measured elements are over-abundant relative to their Solar proportions. From [120]

Fig. 6). In fact, pure collapse models implicitly contradict it because gravitational instabilities provide no way to selectively accrete elements according to their molecular weight. Pressure gradient forces might help to concentrate solids near growing planets [56], and one might conjecture that Jupiter's heavy elements were accreted by the capture of ice-rich planetesimals in the extended atmosphere of the newly formed planet. There are problems with providing enough planetesimals to deliver the mass of Jupiter's metal excess above Solar composition. This process further fails to explain N and Ar, which are overabundant in Jupiter by factors of 3 or 4 (Fig. 6) but which are too volatile to be carried by asteroids or the known comets in any appreciable abundance. The suggestion advanced by Owen et al. [120] is that Jupiter's core grew by the accretion of ultra-cold ($\sim$30 K) planetesimals, in which N, Ar, and other volatiles were efficiently trapped (probably by adsorption within amorphous water ice). But 30 K is too cold to fit the protoplanetary disk at 5 AU (c.f. Equation 2, which gives $T = 125$ K at this distance). A convincing resolution of this puzzle has yet to be identified.

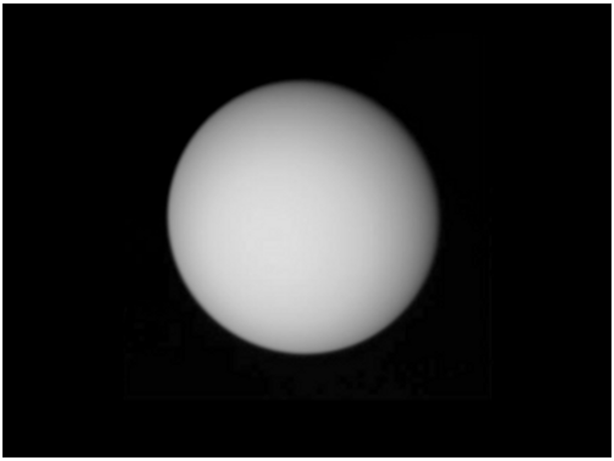

**Fig. 7.** Ice giant Uranus from the Voyager 2 spacecraft. Courtesy NASA

**Ice Giants**

Compared to Jupiter and Saturn, Uranus (Fig. 7) and Neptune (Fig. 8) are an order of magnitude less massive and also compositionally distinct, being depleted in $H_2$ and He. The bulk of their mass is contained in heavier elements that form ices at low temperatures, such as C, N, and O. Uranus and Neptune are known as "ice giants" for this reason. The difficulty in forming Uranus and Neptune on any reasonable timescale has motivated a number of novel, alternative suggestions. For example, in one well-publicized model, Uranus and Neptune are envisioned to have formed between Jupiter and Saturn, were then scattered outwards by mutual perturbations, and, finally, their orbits

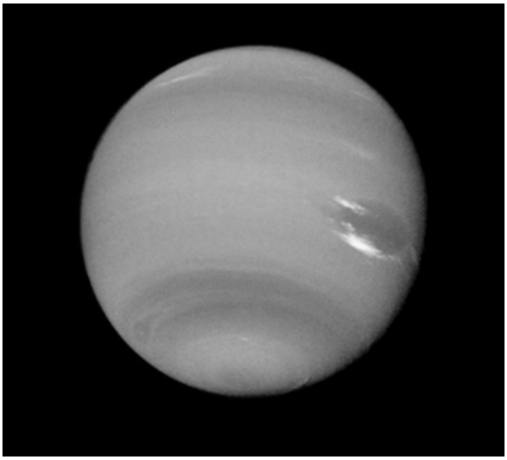

**Fig. 8.** Ice giant Neptune from the Voyager 2 spacecraft. Courtesy NASA

were circularized by friction with an assumed massive disk [149]. To make all this happen, the authors placed the giant planets initially at 6.0, 7.4, 9.0, and 11.1 AU and assumed that they were initially each of $10\,M_\oplus$, with an additional $95\,M_\oplus$ of planetesimals between 12 AU and the assumed edge of the protoplanetary disk at 60 AU. In common with almost all other N-body Solar system simulations, they neglected collective interactions in the $95\,M_\oplus$ disk (these might be expected to generate waves that could be important in the redistribution of angular momentum in the disk [155]). Dynamical effects of the few $\times 10^4\,M_\oplus$ of nebular gas (which must also have been present in order to keep the overall disk composition in approximately cosmic proportions) were also neglected, except that some of this gas was used to feed the runaway growth of the gas giants. The authors assert that their scenario for Uranus and Neptune formation is insensitive to the above assumptions, and, indeed, it is easy to imagine that the first core to experience runaway mass growth should exert a strong gravitational influence on other cores nearby, perhaps scattering them outwards. On the other hand, the initial conditions may have been very different from the ones envisioned in [149]. Worst of all, it is not clear to me what new observations can be taken to test it.

An equally fascinating but rather different scenario for rapid ice giant formation assumes that these planets started out as gas giants and were then eroded down to their observed masses by intense fluxes of ionizing radiation from a nearby, massive star [9]. According to this model, the future ice giants are selectively depleted in mass relative to the surviving gas giants because they are more distant from the sun. Photoionized hydrogen (whose temperature is $\sim 10^4$ K and thermal velocity $\sim 10\,\mathrm{km\,s^{-1}}$) escapes more rapidly from heliocentric orbit at the distances of Uranus and Neptune than at Jupiter and Saturn, leaving the former two planets unprotected from the radiation while the latter two are heavily shielded. Again, the authors do not suggest observational tests of this model, although non-thermal loss of gases from planetary atmospheres often leads to isotopic fractionation effects that might be expected in this extreme case.

## The Domain of the Comets

There are several useful definitions of what it is to be a comet, not all of them mutually consistent. The different definitions are used concurrently, sometimes without a clear understanding of the differences between them. The three different classification schemes are idealized in Fig. 9.

Observationally, a comet is any object showing a gravitationally unbound atmosphere, known as a "coma" (from the Greek for "hair"). The coma is a low-surface brightness region surrounding the central, mass-dominant nucleus. It owes its brightness to a combination of sunlight resonantly scattered from molecules and molecular fragments (radicals) and light scattered from tiny dust particles entrained in the outflowing gas. The visibility of the coma depends on the instrumental sensitivity and angular resolution.

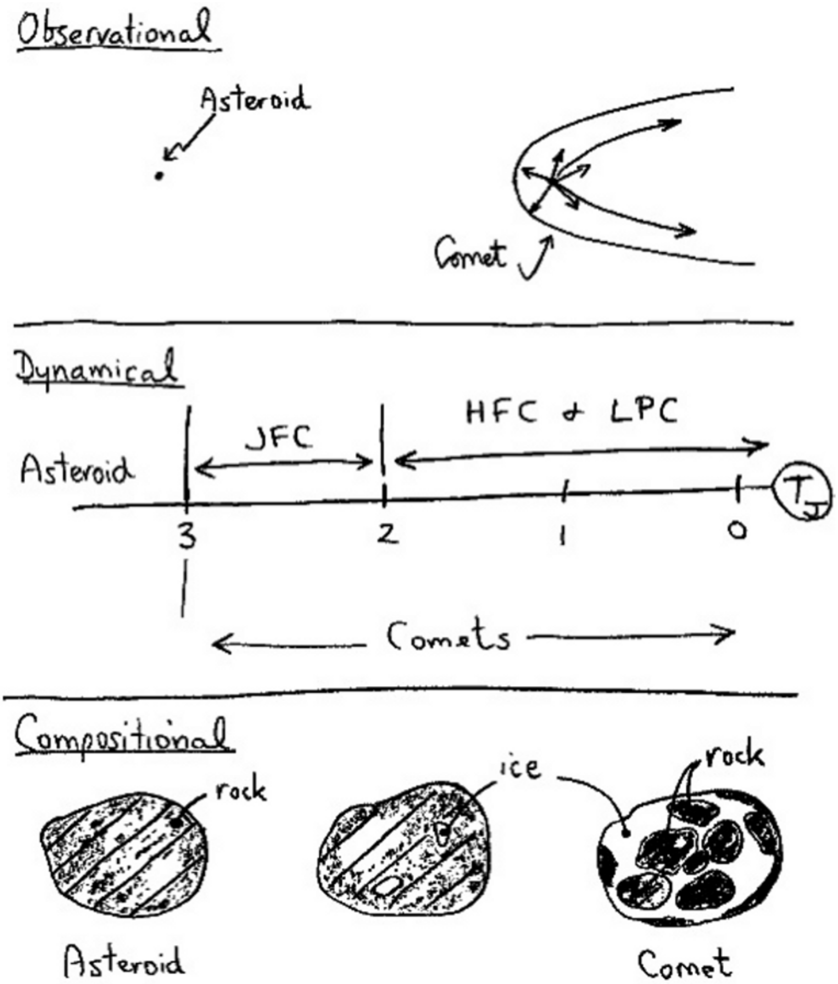


**Fig. 9.** Schematic diagram showing three different criteria for distinguishing comets from asteroids. Observationally, a comet is any body showing a coma (unbound atmosphere) at any point in its orbit. Dynamically, the distinction is made based on some model parameter, typically the Tisserand parameter, $T_J$. JFC, HFC, and LPC denote Jupiter-Family Comets, Halley-Family Comets, and Long-Period Comets. The Main-Belt Comets (MBCs) are located with the asteroids, in the middle panel of the figure. Compositionally, the distinction is based on the presence or absence of bulk ice in the body. The different definitions lead to the same classification in most cases, but there are growing numbers of bodies that are "cometary" by one definition but not the others

For this reason, objects that are discovered by survey telescopes as "asteroids" (i.e., bodies having no atmospheres) are commonly reclassified as comets based on the subsequent detection of comae by observers using more sensitive telescopes. Moreover, the strength of the coma diminishes rapidly with heliocentric distance, falling to invisibility beyond the orbit of Jupiter except in a few unusual cases. On longer timescales, cometary activity can evolve in response to evolutionary process on the surface, in a crust or "mantle" that throttles the release of escaping gas. What appears as a comet now might look completely asteroidal to observers of the twenty second century. Obviously, this observational definition of comet-hood is not at all a perfect one.

Compositionally, a comet may be defined as a small body in which a substantial part of the mass is contained in ice. Practically, we may expect all objects that condensed beyond the "snow-line" to contain bulk water ice. The snow-line is now near the orbit of Jupiter; all small bodies from the Jovian Trojans outward are likely to be compositional comets by this reasoning, whether or not they show comae. In the past, the snow-line may have been closer to the sun, meaning that ice could be present in many of the main-belt asteroids. These bodies are compositionally comets. Unfortunately, we have no meaningful way to estimate the bulk composition of a body without drilling into it, and this definition of comet-hood is consequently hard to apply.

Dynamically, a comet is any body with a Tisserand parameter measured with respect to Jupiter, $T_{\rm J} \leq 3$ (the main-belt asteroids have $T_{\rm J} > 3$). The Tisserand parameter is a constant of the motion in the restricted, circular three-body approximation, defined by

$$T_{\rm J} = \frac{a_{\rm J}}{a} + 2\left[(1-e^2)\frac{a}{a_{\rm J}}\right]^{1/2} cos(i) \tag{3}$$

where $a_{\rm J}$ is the semimajor axis of Jupiter's orbit (assumed circular); $a$, $e$, and $i$ are the semimajor axis, eccentricity and inclination of the small body orbit. Bodies with $T_{\rm J} \leq 3$ strongly interact with the planet, indicating a short dynamical lifetime and a source elsewhere. Those with $T_{\rm J} > 3$ are effectively decoupled from the planet. This definition, although seemingly clean-cut, also suffers from ambiguity. Some main-belt asteroids can be scattered onto orbits with $T_{\rm J} \leq 3$. A few comets (the most famous is 2P/Encke) have $T_{\rm J} > 3$ (although only slightly so), making them dynamically asteroidal.

The timescale for the loss of volatiles from a body is just $\tau_{\rm dv} \sim M/(\mathrm{d}M/\mathrm{d}t))$, where $M$ is the mass and $\mathrm{d}M/\mathrm{d}t$ the rate of loss of mass. Whipple and authors since have assumed that mass loss is predominantly by sublimation [**?**], at a rate that can be calculated from the assumption of radiative equilibrium on the nucleus. There is growing evidence that the mass loss in at least some comets may be dominated by disintegration of the nucleus, in which mass is shed in macroscopic blocks or chunks rather than molecule-by-molecule as in the process of sublimation. Neglecting this possibility for the moment, we write the energy balance equation for a sublimating ice patch as

$$\frac{L_\odot}{4\pi R^2}(1-A)cos(\theta) = \epsilon\sigma T^4 + L(T)\frac{\mathrm{d}m}{\mathrm{d}t} + f_\mathrm{c} + f_\mathrm{g}. \tag{4}$$

Here, $L_\odot$ is the luminosity of the Sun, $R$ is the heliocentric distance, $A$ and $\epsilon$ are the albedo and the emissivity of the surface, $\theta$ is the angle between the direction to the Sun and the surface normal, $L(T)$ is the latent heat of sublimation of the ice at temperature $T$, $dm/dt$ is the mass loss rate per unit area and $f_\mathrm{c}$ represents the conducted energy flux from the surface while $f_\mathrm{g}$ is the flux of energy carried by gas flow into the nucleus. A few things should be noted. The quantity $L_\odot/(4\pi R^2)$ is the flux of sunlight falling on the projected surface. When evaluated at $R = 1\,\mathrm{AU}$, this quantity is called the Solar Constant, $F_\odot$, and has the value $F_\odot = 1360\,\mathrm{W\,m^{-2}}$. The first term on the right-hand side represents the power per unit area lost by radiation into space. The second term is the power per unit area consumed by sublimation. Physically this power is used to break the bonds connecting molecules together in the solid phase. The last term in the equation accounts for thermal conduction and can be either positive or negative, depending on the temperature gradient in the upper layers of the nucleus.

For a non-volatile ($L \rightarrow \infty$) black-body ($A = 0$, $\epsilon = 1$) material oriented perpendicular to the Sun ($\theta = 0$) and neglecting thermal conduction, the temperature is just

$$T = \left[\frac{F_\odot}{\sigma R_\mathrm{AU}^2}\right]^{1/4} \sim \frac{393}{R_\mathrm{AU}^{1/2}}. \tag{5}$$

This corresponds to the temperature at the sub-Solar point on a perfectly absorbing body. The average temperature on a spherical isothermal object will be reduced by a factor $4^{1/4}$, because the average value of $\cos(\theta)$ over the sunlit hemisphere is 1/4, giving $T \sim 278/R_\mathrm{AU}^{1/2}$.

For a sublimating surface, (3) cannot be solved without prior knowledge of the temperature dependence of the latent heat. The Clausius–Clapeyron equation (for the slope of the solid-gas phase boundary in pressure vs. temperature space) can be used or, more directly, measurements of the thermal pressure exerted by sublimating water ice as a function of temperature can be employed. For illustrative purposes, we here consider an extreme approximation.

When close to the Sun (say for $R_\mathrm{AU} < 1\,\mathrm{AU}$) water ice, the dominant cometary volatile, uses so much energy to sublimate that we may write

$$\frac{L_\odot}{4\pi R^2}(1-A)\cos(\theta) \sim L(T)\frac{\mathrm{d}m}{\mathrm{d}t}. \tag{6}$$

as a rough approximation to (4). Then, we see that the characteristic mass loss rate per unit area (again with $\theta = 0$) is just

$$\frac{\mathrm{d}m}{\mathrm{d}t} \sim \frac{F_\odot}{L(T)R_\mathrm{AU}^2} \tag{7}$$

and we have assumed for simplicity that the surface is perfectly absorbing, $A$ =0. Substituting $F_\odot = 1360\,\mathrm{W\,m^{-2}}$ and $L(T) = 2\times10^6\,\mathrm{J\,kg^{-1}}$ (for water ice), we have $\mathrm{d}m/\mathrm{d}t \sim 7\times10^{-4}/R_\mathrm{AU}^2$ [$\mathrm{kg\,s^{-1}\,m^{-2}}$].