PLANETARY SCIENCES

American and Soviet Research

Edited by

Thomas M. Donahue

with

Kathleen Kearney Trivers
David M. Abramson

Proceedings from the US-USSR Workshop
on
Planetary Sciences
January 2-6, 1989

Academy of Sciences of the Union of Soviet Socialist Republics
National Academy of Sciences of the United States of America

Foreword

The Academy of Sciences of the USSR and the National Academy of Sciences of the United States of America sponsored a workshop on Planetary Sciences at the Institute for Space Research in Moscow, January 2-6, 1989. The purpose of the workshop, which was attended by Soviet and American scientists, was to examine the current state of our theoretical understanding of how the planets were formed and how they evolved to their present state. The workshop assessed the type of observations and experiments that are needed to advance understanding of the formation and evolution of the solar system based on the current theoretical framework.

In the past, models of the formation and evolution of the planets have been just that, models. They essentially portrayed possibilities: that events could have transpired in the way depicted without violating any known constraints and that, if they did take place, certain consequences would follow. Now, we may be advancing beyond that stage to the point where it may be possible to settle fairly definitely on certain scenarios and exclude others. Assessment of the present state of theories and the observational base will help determine the extent to which this is the case.

This workshop focused on the present status of observational and theoretical understanding of the clearing of stellar nebulae, planetesimal formation, and planetary accretion; the evolution of atmospheres; the relationship of still existing primitive bodies to these topics; and the relationship of ground-based and *in situ* measurements.

As the papers presented at the workshop and published in this volume show, astronomical observations are now at hand that will reveal the sequence of events occuring in circumstellar disks with sufficient precision to define the models of planetary formation. Moreover, theories for the origin of the solar system have reached a point where it is now possible,

indeed it is essential, to examine them in the broader context of the origins of planetary systems. Until recently, it has not been possible to select among a range of feasible scenarios the one that is most likely to be correct, in the sense that it satisfies observational and theoretical constraints. We are now advancing to that stage.

Similarly, theories for the formation of planetary atmospheres are becoming more sophisticated and respond to an increasingly complex set of observational data on relative abundance of atmospheric species and the state of degassing of the interior of the Earth. But, as the papers presented here will demonstrate, there is still a way to go before all the questions are resolved.

The topics discussed at this workshop were timely, and the debate and discussion were full and informative. The participants learned a great deal, and the scientific basis for cooperation in planetary sciences was strengthened appreciably. The Soviet hosts extended their usual thoughtful and gracious hospitality to the American delegation, and the entire experience was memorable and rewarding.

<div align="right">

T.M. Donahue
Chairman, NAS/NRC Committee on
Cooperation with the USSR on
Planetary Science

</div>

Acknowledgements

Contents

PRECEDING PAGE BLANK NOT FILMED

The Properties and Environment of Primitive Solar Nebulae as Deduced from Observations of Solar-Type Pre-main Sequence Stars[1]

STEPHEN E. STROM
SUZAN EDWARDS[2]
KAREN M. STROM
University of Massachusetts, Amherst

ABSTRACT

This contribution reviews a) current observational evidence for the presence of circumstellar disks associated with solar type pre-main sequence (PMS) stars, b) the properties of such disks, and c) the disk environment. The best evidence suggests that at least 60% of stars with ages $t < 3 \times 10^6$ years are surrounded by disks of sizes ~ 10 to 100 AU and masses ~ 0.01 to $0.1\ M_\odot$. Because there are no known main sequence stars surrounded by this much distributed matter, disks surrounding newborn stars must evolve to a more tenuous state. The time scales for disk survival as massive ($M \sim 0.01$ to $0.1\ M_\odot$), optically thick structures appear to lie in the range $t \ll 3 \times 10^6$ years to $t \sim 10^7$ years. At present, this represents the best astrophysical constraint on the time scale for assembling planetary systems from distributed material in circumstellar disks. The infrared spectra of some solar-type PMS stars seem to provide evidence of *inner disk clearing*, perhaps indicating the onset of planet-building.

The material in disks may be bombarded by energetic (~ 1 kev) particles from stellar winds driven by pre-main sequence stars. However, it is not known at present whether, or for how long such winds leave the stars a) as highly collimated polar streams which do not interact with disk material, or b) as more isotropic outflows.

INTRODUCTION

Recent theoretical and observational work suggests that the process of star formation for single stars of low and intermediate mass ($0.1 < M/M_\odot < 5$) results naturally in the formation of a circumstellar disk, which may then evolve to form a planetary system. This process appears to involve the following steps (Shu et al. 1987):

• The formation of opaque, cold, rotating protostellar "cores" within larger molecular cloud complexes;

• The collapse of a core when self-gravity exceeds internal pressure support;

• The formation within the core of a central star surrounded by a massive (0.01 to $0.1\ M_\odot$) circumstellar disk (*embedded* infrared source stage). At this stage, the star/disk system is surrounded by an optically opaque, infalling protostellar envelope. Gas and dust in the envelope rains in upon both the central stellar core and the surrounding disk, thus increasing both the disk and stellar mass. Infall of low angular momentum material directly onto the central star and accretion of high angular momentum material through the disk provide the dominant contributions to the young stellar object's (YSO) luminosity, far exceeding the luminosity produced by gravitational contraction of the stellar core. Because the infalling envelope is optically opaque, such YSOs can be observed only at wavelengths $\lambda > 1\mu$. They exhibit infrared spectral energy distributions, λF_λ vs. λ, which are a) broad compared to a blackbody distribution and b) flat, or rising toward long wavelengths (see Figure 1a). At some time during the infall phase, the central PMS star begins to drive an *energetic wind* ($L_{wind} \sim 0.1\ L_*$; $v_{wind} \sim 200$ kilometers per second). This is first observable as a highly collimated "molecular outflow" as it transfers momentum to the surrounding protostellar and molecular cloud material and later (in some cases) as a "stellar jet" (Edwards and Strom 1988). The wind momentum is sufficient to reverse infall from the protostellar envelope and eventually dissipates this opaque cocoon, thus revealing the YSO at visible wavelengths (Shu et al. 1987).

• The optical appearance of a young stellar object (YSO) whose luminosity may be dominated in the infrared by accretion through the disk, and in the ultraviolet and optical region, by emission from a hot "boundary layer" (*continuum + emission* TTS *phase*). Accretion of material through the still massive (0.01 to $0.1\ M_\odot$) disk produces infrared continuum radiation with a total luminosity $\gg 0.5$ times the stellar luminosity. At this stage, the infrared spectrum is still much broader than a single, blackbody spectrum (Myers et al. 1987; Adams et al. 1986), and in most cases falls
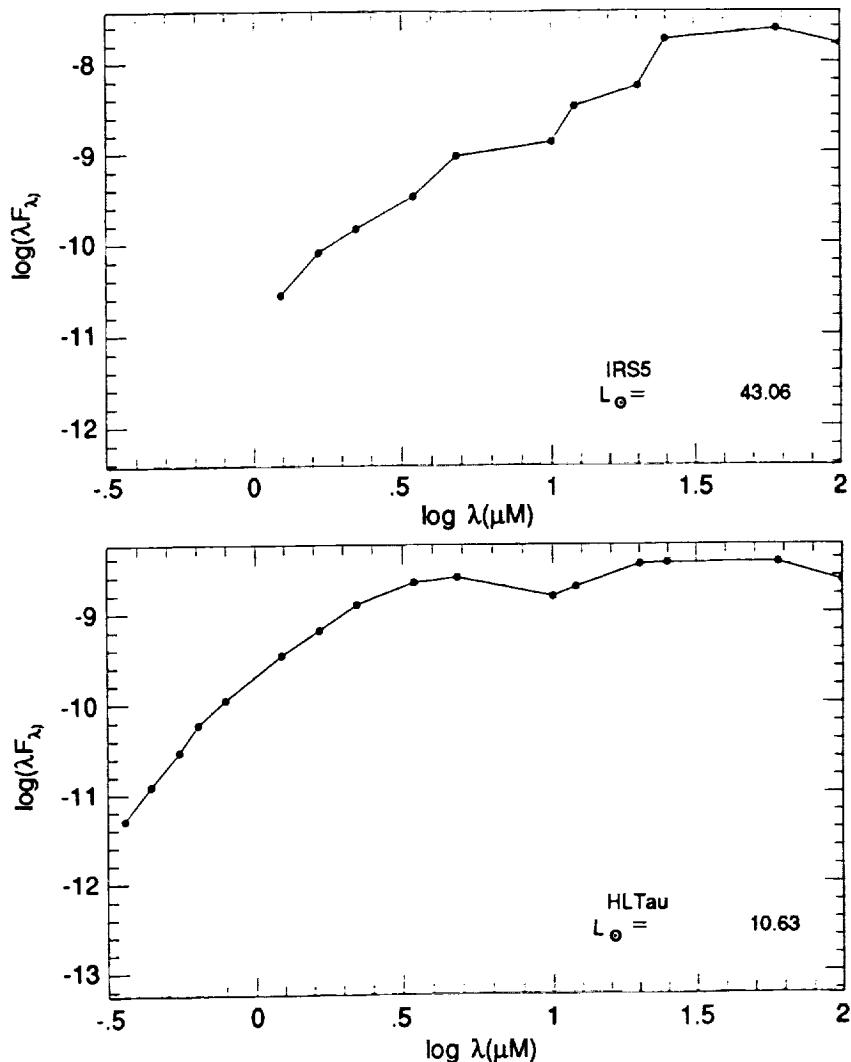
FIGURE 1 (a) A plot of the observed spectral energy distribution for the *embedded infrared source*, L1551/IRS 5. This source drives a well-studied, highly-collimated, bipolar molecular outflow and a stellar jet. Note that the spectrum rises toward longer wavelength, suggesting that IRS 5 is surrounded by a flattened distribution of circumstellar matter, possible remnant material from the protostellar core from which this YSO was assembled. (b) A plot of the observed spectral energy distribution for the *continuum + emmission T Tauri star*, HL Tau. Near-infrared images of HL Tau suggest that this YSO is surrounded by a flattened distribution of circumstellar dust. Its infrared spectrum is nearly flat, which suggests that HL Tau is as well still surrounded by a remnant, partially opaque envelope.

toward longer wavelengths. In a few cases, the spectrum is flat or rises toward long wavelengths, perhaps reflecting the presence of a partially opaque, remnant infalling envelope (Figure 1b).

The accretion of material from the rapidly rotating ($\sim$ 200 kilometers per second), inner regions of the disk to the slowly rotating ($\sim$ 20 kilometers per second) stellar photosphere also produces a narrow, hot (T $\geq$ 8000 K) emission region: the "boundary layer" (Kenyon and Hartmann 1987; Bertout *et al.* 1988). Radiation from the boundary layer overwhelms that from the cool PMS star photosphere. Consequently, the photospheric absorption spectrum cannot be seen against the boundary layer emission at $\lambda < 7000$ Å. Strong permitted and forbidden emission lines (perhaps tracing emission associated with energetic stellar winds and the boundary layer) are also present during this phase; hence the classification "continuum + emission" objects. All such objects drive energetic winds, some of which are manifest as highly collimated stellar jets (Strom *et al.* 1988; Cabrit *et al.* 1989).

• The first appearance of the stellar photosphere, as the contribution of the disk to the YSO luminosity decreases (*T Tauri or TTS (0.2 < M/M$_\odot$ Tauri < 1/5) and Herbig emission star or HES (1.5 < M/M$_\odot$ < 5) phase*). Relatively massive disks are still present, but the accretion rate and mass outflow rate both diminish. The infrared luminosity from accretion and the boundary layer emission decrease as a consequence of the reduced accretion rate through the disk. "Reprocessing" of stellar radiation absorbed by dust in the accretion disk and re-radiated in the infrared contributes up to 0.5 times the stellar luminosity (depending on the inclination of the star/disk system with respect to the line of sight). The observed infrared spectra show the combined contribution from a Rayleigh-Jeans ($\lambda F_\lambda \sim \lambda^{-3}$) component from the stellar photosphere and a broader, less rapidly falling ($\lambda^{-2/3}$ to $\lambda^{-4/3}$) component arising from both disk accretion and passive reprocessing by circumstellar dust (Figure 2). Photospheric absorption spectra are visible, though sometimes partially "veiled" by the hot boundary layer emission. H emission is strong (typical equivalent widths ranging from $10 < W_\lambda < 100$ Å). In TTS, Ca II and selected forbidden and pace permitted metallic lines are often prominent in emission. Energetic winds persist, although their morphology and interaction with the circumstellar environment is unknown at this stage. Highly collimated optical jets are not seen, but spatially unresolved [O I] line profile structures require that the winds be at least moderately collimated (on scales of $\sim$ 100 AU).

• The settling of dust into the midplane of the disk followed by the clearing of *distributed material* in the disk, as dust agglomerates into planetesimals, first in the inner regions of the disk where terrestrial planets form, and later in the outer disk where giant planets and comets are
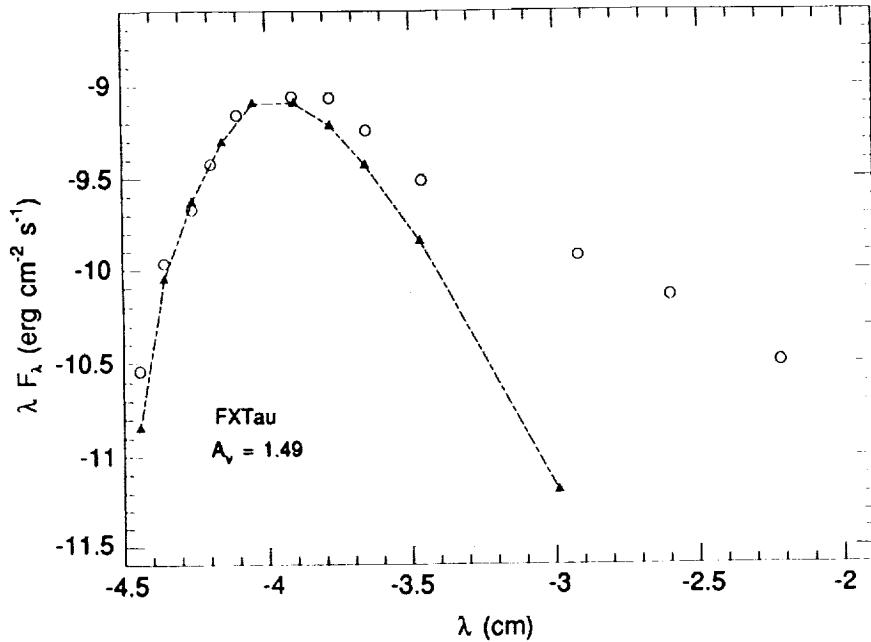
FIGURE 2   A plot of the reddening-corrected spectral energy distribution ($0.35\mu < \lambda \leq$ 100 μ) for the *T Tauri star*, FX Tau (open circles). Also plotted is the spectral energy distribution of a dwarf standard star (filled triangles) of spectral type M1 V corresponding to that of FX Tau. The flux of the TTS and the standard star have been forced to agree at R (0.65μ).

assembled. The disappearance of accretion signatures and energetic winds (*"naked" T Tauri or NTTS phase?*).

• The appearance of the star on the hydrogen-burning main sequence, accompanied by its planetary system and possibly by a tenuous remnant or secondary dust disk (analogous to the edge-on disk surrounding B. Pictoris imaged recently by Smith and Terrile 1984).

Is this picture correct even in broad outline for all single stars? Are disks formed around members of binary and multiple star systems (which constitute at least 50% of the stellar population in the solar neighborhood)? For those stars that form disks, what is the range of disk sizes and masses? What is the range in time scales for disk evolution in the inner and outer disks, and how do these time scales compare with those inferred for our solar system from meteoritic and primitive body studies, and theoretical modeling of the early solar nebula? In what fraction of star/disk systems can the gas in the outer disk regions survive removal by energetic winds long enough to be assembled into analogs of the giant planets?

RECENT OBSERVATIONAL RESULTS

Current Observational Evidence for Disks
Associated with Pre-Main Sequence Stars

Observations carried out over the past five years provide strong evidence for circumstellar disks associated with many young stars (ages $< 3 \times 10^6$ years) of a wide range of masses. These disks appear to be massive (M ~ 0.01 to 0.1 $M_\odot$) precursor structures to the highly evolved, low-mass (M $\sim 10^{-7}$ $M_\odot$)) disks discovered recently around B-Pictoris and its analogs (Smith and Terrile 1984; Backman and Gillett 1988):

• The direct and speckle imaging at near-infrared wavelengths (Grasdalen *et al.* 1984; Beckwith *et al.* 1984; Strom *et al.* 1985) reveal flattened, disk-like structures associated with three YSOs: HL Tau (a low-mass continuum + emission star), R Mon (an intermediate-mass continuum + emission star), and L1551/IRS 5 (a low-mass embedded infrared source). These structures are seen via light scattered in our direction by sub-micron and micron-size dust grains. Associated structures are also seen in mm-continuum and CO line images obtained with the Owens Valley interferometer (Sargent and Beckwith 1987), although the relationship between the optical and near-infrared and mm-region structures is not clear at this point. Shu (1987, private communication) has argued that the flattened, scattered light structures detected to date trace not disks, but rather remnants of infalling, protostellar cores (see also Grasdalen *et al.* 1989).

• The high-spectral resolution observations of [O I] and [S II] lines in continuum + emission, T Tauri, and Herbig emission stars provide indirect but compelling evidence of such disks. The forbidden line emission is associated with the outer ($r \sim 10$ to 100 AU) regions of winds driven by PMS stars. However, in nearly all cases studied to date, only blue-shifted emission is observed (see Figure 3) thus requiring the presence of structures whose opacity and dimension is sufficient to obscure the receding part of the outflowing gas diagnosed by the forbidden lines (Edwards *et al.* 1987; Appenzeller *et al.* 1984).

• The broad, far-infrared ($\lambda > 10\mu$) spectra characteristic of *all* continuum + emission, T Tauri, and Herbig emission stars arise from heated dust located over a wide range of distances (~ 0.1 to > 100 AU) from a central pre-main sequence (PMS) star (Rucinski 1985; Rydgren and Zak 1986). In order to account for the fact that these IR-luminous YSOs are visible at optical wavelengths, it is necessary to assume that the observed far-IR radiation arises in an optically thick but *physically thin* circumstellar envelope: a disk. If the heated circumstellar dust responsible for the observed far-infrared radiation intercepted a large solid angle, the
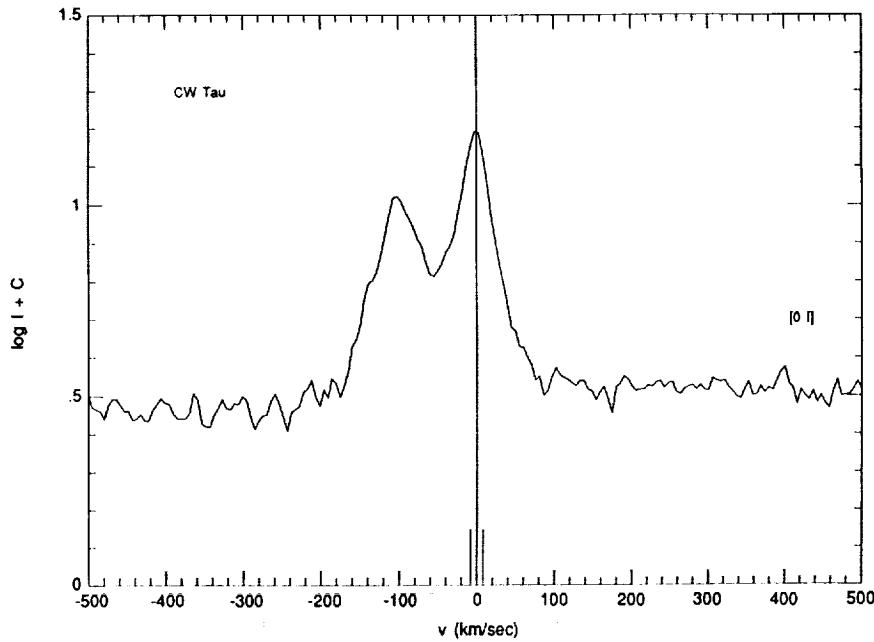
FIGURE 3 A plot of the [0 I] λ 6300 Å profile observed for the T Tauri star, CW Tau. Note the broad, double-peaked profile extending to blue-shifted (negative) velocities; there is no corresponding red-shifted emission. The [0 I] emission is believed to trace low density, outflowing gas located at distances r > 10 AU from the surface of CW Tau. The absence of red-shifted emission is attributed to the presence of an opaque circumstellar disk whose size is comparable to or larger than the region in which [0 I] emission is produced.

associated PMS stars could not be seen optically (Myers *et al.* 1987; Adams *et al.* 1987). The observed far-IR fluxes require a mass of emitting dust 10^{-3} to 10^{-4} M$_\odot$ or a total disk mass of 0.1 to 0.01 M$_\odot$ (assuming a gas/dust ratio appropriate to the interstellar medium).

• The optical and infrared spectra of a class of photometrically eruptive YSOs known as FU Ori objects appear to arise in self-luminous, viscous accretion disks characterized by a temperature-radius relation of the form T ~ $r^{-3/4}$ (Hartmann and Kenyon 1987a,b; Lynden-Bell and Pringle 1974). Because material in this disk must be in Keplerian motion about a central PMS star, absorption lines formed in the disk reflect the local rotational velocity. High spectral resolution observations show that lines formed in the outer, cooler regions of the disk are narrower than absorption features formed in the inner, hotter, more rapidly rotating disk regions (Hartmann and Kenyon 1987a,b; Welty *et al.* 1989), thus providing

important *kinematic* confirmation of disk structures associated with PMS stars.

- The mm-line and continuum observations of HL Tau and L1551/IRS 5 made with the OVRO mm interferometer suggest that circumstellar gas and dust is bound to the central PMS star and, in the case of HL Tau, in Keplerian motion about the central object (Beckwith and Sargent 1987).

Frequency of Disk Occurrence

What fraction of stars are surrounded by disks of distributed gas and dust at birth? If excess infrared and mm-wave continuum emission is produced by heated dust in disks, then all continuum + emission, T Tauri, and Herbig emission stars must be surrounded by disks. The inferred disk masses ($0.01 < M_{disk}/M_\odot < 0.1$, comparable to the expected mass of the primitive solar nebula) and optical depths ($\tau_V \sim 1000$) for TTS and HES are relatively large (Beckwith *et al.* 1989 and Edwards *et al.* 1987 for estimates based on mm-continuum and far-IR measurements respectively).

However, the HR diagram presented by Walter *et al.* (1988) suggests that $\sim 50\%$ of low mass pre-main sequence stars with ages comparable to those characterizing TTS ($t < 3 \times 10^6$ years) are "naked" T Tauri stars (NTTS) which lack measureable infrared emission, and therefore *massive*, optically thick disks. The observations of Warner *et al.* (1977) suggest that a comparable percentage (50% to 70%) of young ($t < 3 \times 10^6$ years) intermediate mass stars ($M \sim 1.5–2.0\ M_\odot$) also lack infrared excesses.

Recently, Strom *et al.* (1989) examined the frequency distribution of near-IR (2.2μ) excesses, $\Delta K \equiv \log \{F_{2.2\mu}$ (PMS star)$/ F_{2.2\mu}$ (standard star)$\}$, associated with 47 NTTS and 36 TTS in Taurus-Auriga (see Figure 4). They find that 84% of the TTS and 36% of the NTTS have significant excesses, $\Delta K \geq 0.10$ dex. Thus, nearly 60% of solar-type PMS stars with ages $t < 3 \times 10^6$ years located in this nearby star-forming complex have *measurable*[3] infrared excesses; these excesses most plausibly arise in disks. However, the sample includes only *known* PMS stars for which adequate photometry is available; more NTTS may yet be discovered when more complete x-ray and proper motion searches of the Taurus-Auriga clouds become available. It is also important to note that these disk frequency statistics *exclude* several PMS stars which show small or undetectable near IR excesses, but relatively strong mid- to far-IR excesses (see Figure 5):

---

[3] Objects that lack measurable infrared excesses *could be* surrounded by low mass, tenuous disks ($M \ll 10^{-5}\ M_\odot$) disks. For example, a disk of mass comparable to that surrounding β Pic ($M \sim 10^{-7}\ M_\odot$) could not be detected around a PMS star in the Taurus-Auriga clouds given the current sensitivity of IR measurements.
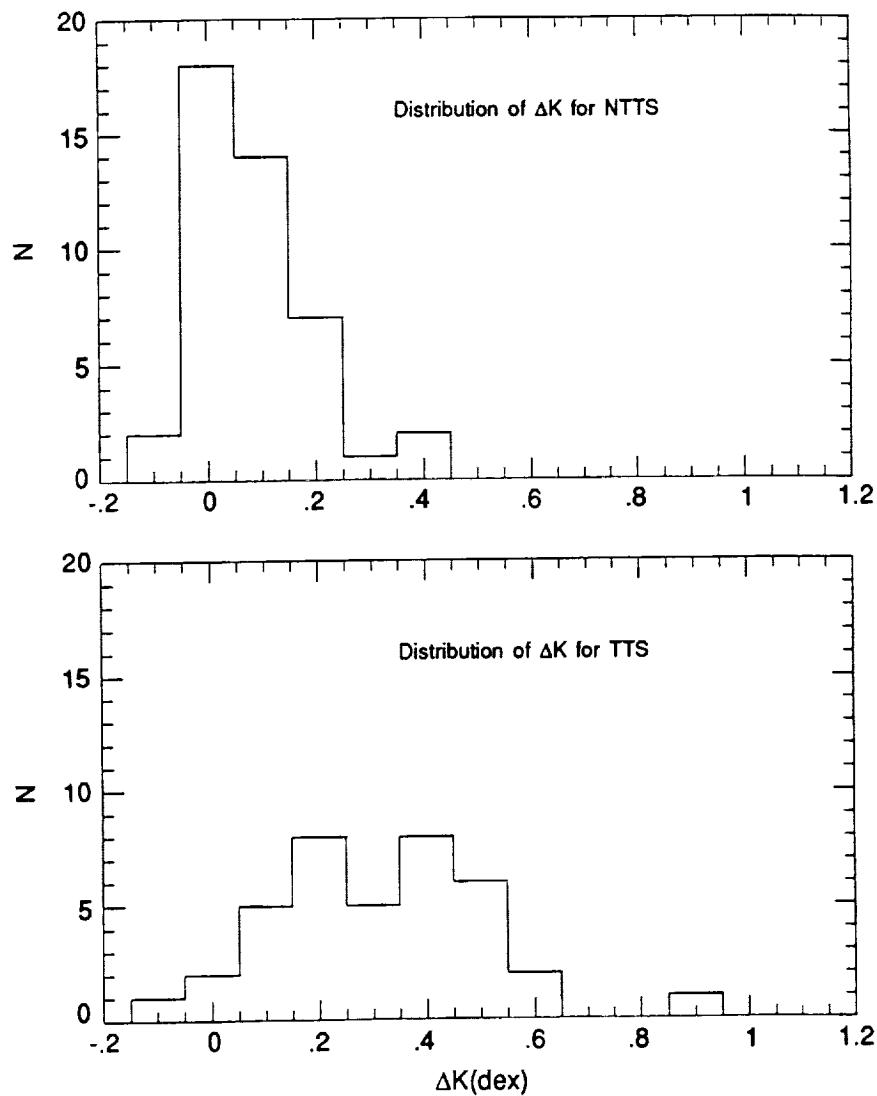
FIGURE 4   The frequency distribution of the 2.2μ excess, $\Delta K \equiv \log \left\{ F_{2.2\mu} \text{ (PMS)}/F_{2.2\mu} \right.$ (standard)$\left. \right\}$, for NTTS (top) and TTS (bottom). Note that a) 36% of the NTTS show excesses, $\Delta K \geq 0.10$ dex, and b) that while the distribution for the NTTS peaks toward smaller values of ΔK, there is significant overlap in the two distribution. It does not appear as if NTTS as a class lack infrared excesses.

these objects may represent PMS stars surrounded by circumstellar disks which are optically thin near the star (and therefore produce too little near-IR radiation to be detected), but optically thick at distances r > 1 AU (see below).

Do pre-main sequence stars with ages t ≤ 3 × 10^6 years that lack infrared excesses (40% of all solar-type PMS stars) represent a population of stars born without disks? Or have their disks been destroyed by tides raised by a companion star? Or have some fraction of these young PMS stars already built planetary systems?

The Effect of Stellar Companions on Disk Survival

Do disks form around members of binary and multiple star systems? If so, are these disks perturbed by tidal forces and perhaps disrupted when the disk size is comparable to the separation between stellar components? To date, the overwhelming majority of binaries discovered among low- and intermediate-mass PMS stars in nearby star-forming complexes have been wide ($\Delta\theta$ > 1″; r > 150 AU) doubles with separations well in excess of the radius of our solar system (and thus possibly of lesser interest to addressing the above questions). In the last few years, ground-based observations have uncovered a few examples of a) spectroscopic binaries with velocity amplitudes, Δv > 10 km/s (separations < 3 AU; Hartmann *et al.* 1986); b) binaries with separations in the range 0.1 to 50 AU detected from lunar occultation observations of YSOs in Taurus-Auriga and Ophiuchus (Simon, private communication); and c) binaries discovered in the course of optical and infrared speckle interferometric observations (Chelli *et al.* 1988). Of the known binaries in Taurus-Auriga with observed or inferred separations $\Delta\theta$ < 0.5″ (r < 70 AU; DF Tau, FF Tau, HV Tau, HQ Tau, T Tau), all but FF Tau appear to have IR excesses similar to those of TTS thought to be surrounded by disks. This somewhat surprising result implies that circumstellar envelopes of mass > 10^{-4} $M_\odot$ and of dimension 10 to 100 AU are present *even in close binary systems.*

Disk Evolutionary Time Scales

On what time scales do disks evolve from massive "primitive" to low mass, perhaps "post planet-building" disks? Current observations suggest that more than 50% of low- to intermediate-mass PMS stars with ages t < 3 × 10^6 years are surrounded by disks with masses in excess of 0.01 $M_\odot$ of distributed material (see above). There are *no* known main sequence stars surrounded by this much distributed matter, although a few stars such as Vega and β Pictoris (Smith and Terrile 1984; Backman and Gillett 1988)
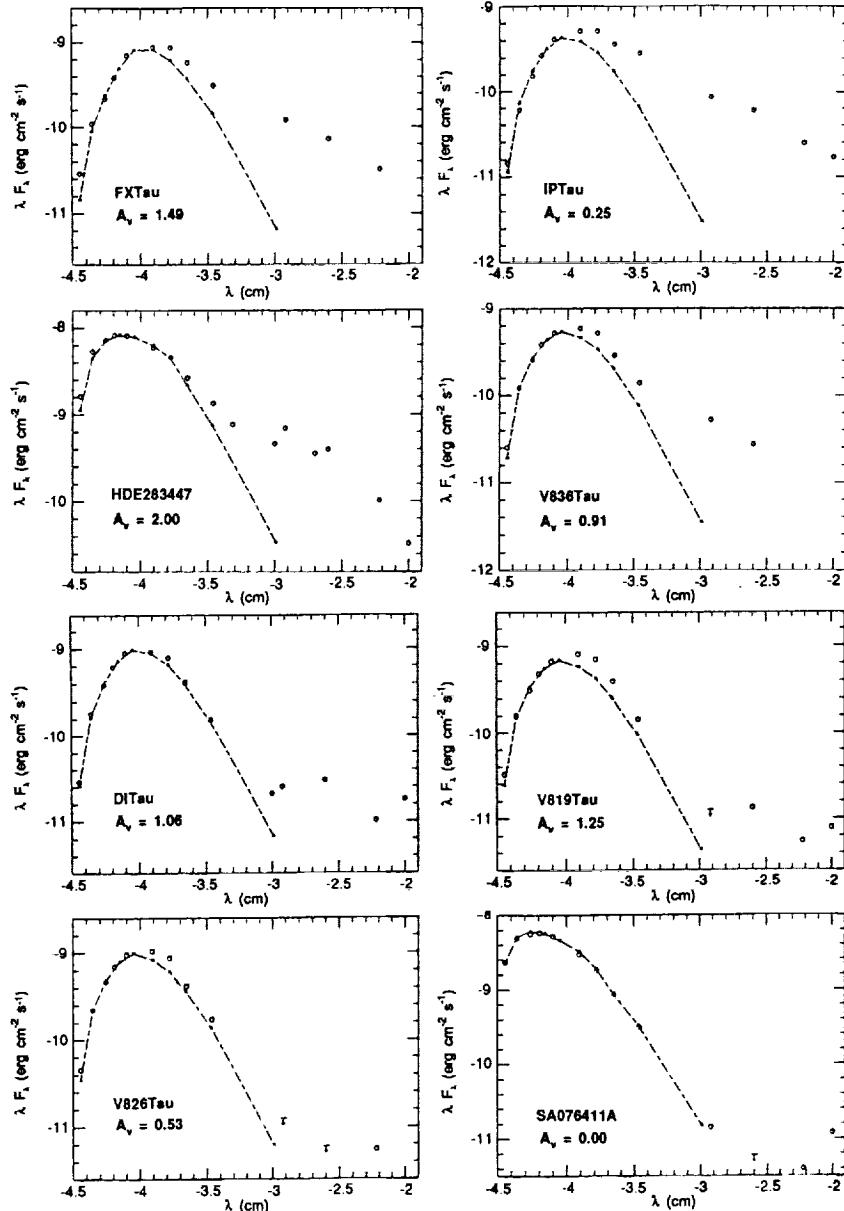
FIGURE 5   A plot of the reddening-corrected spectral energy distributions ($0.35\mu < \lambda \leq 100\mu$) for 7 NTTS and the T Tauri star, FX Tau (see Figure 2); the observed points for the NTTS and TTS are plotted as open circles. Also plotted are the spectral energy distributions of dwarf standard stars (filled triangles) of spectral types corresponding to those of the NTTS and TTS. The fluxes of the NTTS and those of the standard stars have been forced to agree at R (0.65μ). Note the small near-infrared excess and relatively large mid- to far-infrared excesses for several of the NTTS. A spectral energy distribution of this character might be produced by a circumstellar disk in which the optical depth of emitting material in the inner disk is small, while that in the outer disk is large.

are surrounded by disks with masses $\sim 10^{-7}$ $M_\odot$ ($\sim$ 0.1 Earth masses). Thus, *disks surrounding newborn stars must evolve to a more tenuous state.*

Recent work by Strom *et al.* (1989) suggests that nearly 60% of PMS stars in Taurus-Auriga with ages younger than 3×10^6 years, and only 40% of older stars, show evidence of significant near-infrared excesses $\Delta K \geq$ 0.1 dex (see above). These results are illustrated in Figure 6. *Fewer than 10% of PMS stars older than 10^7 years show $\Delta K \geq 0.1$.* If excess near-IR emission arises in the warm, inner regions of circumstellar disks, then we can use these statistics to discuss the range of time scales for disks to evolve from massive, optically thick structures (with large K values) to low-mass, tenuous entities (with small ΔK values).

If *all* solar-type stars form massive (0.01 to 0.1 $M_\odot$ disks, then by t = 3×10^6 years, 40% of PMS stars (the fraction of young PMS stars with $\Delta K < 0.10$) are surrounded by remnant disks too tenuous to detect. The evolutionary time scale for such rapidly evolving disks must be t $\ll 3 \times 10^6$ years. Because fewer than 10% of PMS stars *older* than 10^7 years show $\Delta K \geq 0.1$ (Strom *et al.* 1989), the disks surrounding all but 10% of PMS stars must have completed their evolution by this time. The majority of disks must have evolutionary time scales in the range 3×10^6 to 10^7 years. This range represents the best astrophysical constraint on the likely time scale for planet building available at present.

As noted earlier, the above statistics obtain for all *known* TTS and NTTS in Taurus-Auriga for which adequate photometry is available. Although the exact fraction of PMS stars surrounded by disks may change somewhat as more complete surveys for PMS stars become available, our qualitative conclusions regarding the *approximate* time scale range for disk evolution are unlikely to be vitiated.

Disk Sizes and Morphologies

In our solar system, all known planets lie within 40 AU of the Sun. Yet the circumstellar disk imaged around β Pic extends to a distance approximately several thousand AU from the central star. Do primitive solar nebulae typically extend to radii considerably larger than our own planetary system? If so, how far, and how much material do they contain? How do properties such as size and surface mass distribution change with time? Do such changes reflect the consequences of angular momentum transport within the disk? Of planet building episodes?

Current estimates of disk sizes are indirect, and are based on: (1) the size of the YSO wind region required to account for the observed blue-shifted [O I] and [S II] forbidden line emission fluxes; disks must be large enough to occult the receding portion of the stellar wind (Appenzeller *et al.* 1984; Edwards *et al.* 1987) and (2) the projected radiating area required to
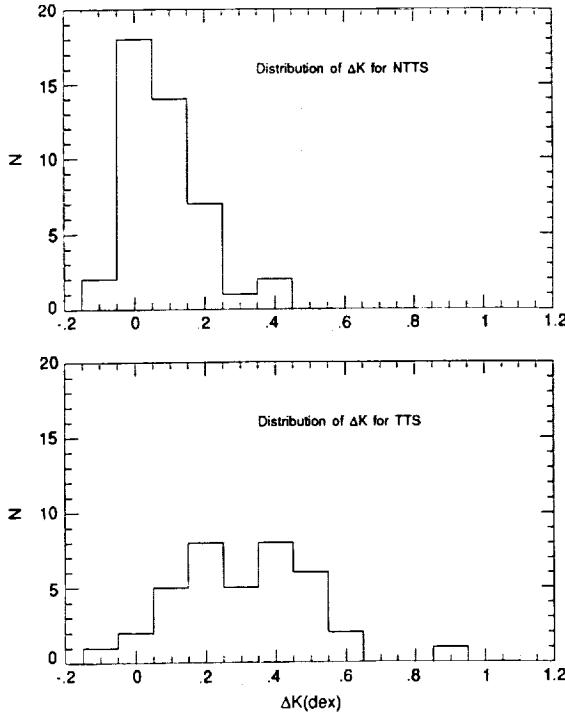
FIGURE 6 The frequency distribution of the near-infrared excess ΔK (see text) for stars with ages $t > 3 \times 10^6$ years and $t \leq 3 \times 10^6$ years. Note that a) nearly 60% of young PMS stars have measurable ($\Delta K > 0.10$) near-infrared excesses and b) that the fraction of PMS stars with such excesses decreases for ages $t \gg 3 \times 10^6$ years. If IR excesses derive from emitting dust embedded in massive, optically thick circumstellar disks surrounding PMS stars, then the fraction of stars surrounded by such disks must decrease with time. Our data suggests that the time scales for evolving from a massive, optically thick disk to a low mass, tenuous disk must range from $\sim 3 \times 10^6$ to 10^7 years. This range represents the best available astrophysical constraint on the time scale for planet building.

explain the observed far-infrared radiation emanating from optically thick, cool dust in the outer disk regions (Myers *et al.* 1987; Adams *et al.* 1987). In both cases, these estimates provide lower limits to the true extent of the disk. Edwards *et al.* (1987) have shown that these independent methods predict comparable lower limit disk size estimates: $r_{disk} > 10$ to 100 AU, for a sample of continuum + emission, T Tauri, and Herbig emission stars. At the distance of the nearest star-forming regions, such disks intercept an angular radius, $r > 0.07$ to 0.7 arc seconds. Thus far, it has proven difficult to image disks of this size from the ground. Decisive information regarding disk isophotal sizes and surface brightnesses must await sensitive

imaging with HST, whose stable point spread function and high angular resolution will permit imaging of low-surface brightness circumstellar disks around bright PMS stars. When available, HST measurements of disk sizes will prove an invaluable complement to ground-based sub-mm and mm-continuum measurements which provide strong constraints on the disk *mass*, but which lack the spatial resolution to determine *size*. Together, these data will yield average surface mass densities and estimates of midplane optical depths—critical parameters for modeling the evolution of primitive solar nebulae.

Prior to HST, ground-based observations of the ratio of near- to far-IR excess radiation may provide a qualitative indication of the distribution of material in circumstellar disks. For example, Strom *et al.* (1989) discuss several solar-type PMS stars which show small near-IR excesses compared to far-IR excesses (Figure 5). They suggest that the disks surrounding these stars have developed *central holes* as a first step in their evolution from massive, optically thick structures (such as those surrounding TTS) to tenuous structures (such as those surrounding β Pic and Vega). Such central holes may provide the first observational evidence of planet-building around young stars.

The Disk Environment

High resolution ground-based spectra suggest that energetic winds ($L_{wind} > 0.01$ L_*; v $\sim$ 200 kilometers per second) characterize all TTS and HES (Edwards and Strom 1988). However, mass-loss rates are not well determined: estimates range from 10^{-6} to 10^{-9} $M_\odot$ per year and are greatly hampered by uncertainties in our knowledge of the wind geometry. The broad, blueshifted, often double-peaked forbidden line profiles of [O I] and [S II] (see Figure 3), suggest that typical TTS and HES winds are not spherically symmetric and may be at least moderately collimated. The models proposed to account for the forbidden line profiles include a) latitude-dependent winds characterized by higher velocities and lower densities in the polar regions, b) sub-arc second, highly collimated polar jets; and c) largely equatorial mass outflows obliquely shocking gas located at the raised surface of a slightly flaring disk (Hartmann and Raymond 1988).

Knowledge of the wind geometry is necessary if we are to derive more accurate estimates of PMS star mass loss rates from [O I] profiles. Depending on their mass loss rate and geometry, PMS star winds may have a profound effect on the survival of gas in circumstellar disks and on the physical and chemical characteristics of the grains:

• Energetic winds can remove gas from the disk during the epoch of planet building, thereby eliminating an essential "raw material" for assembling massive giant planets in the outer disk.

• Exposure of interstellar grains to ~ 1 kev wind particles carried by a wind with $M \sim 10^{-9}$ $M_{\odot}$ per year for times of 10^6 to 10^7 years, can alter the chemical composition of grain mantles. For a grain with a water-ammonia-methane-ice mantle, energetic particle irradiation can in principle: (1) create large quantities of complex organic compounds on grain surface and (2) drastically reduce the grain albedo (Greenberg 1982; Lanzerotti *et al.* 1985; Strazzula 1985). Recent observations of the dust released from the nucleus of comet Halley show these grains to be "black" (albedos of ~ 0.02 to 0.05) and probably rich in organic material (Chyba and Sagan 1987). Does this cometary dust owe its origin to irradiation of grains by the Sun's T Tauri wind during the early lifetime of the primitive solar nebula?

HST will allow us to image low-density, outflowing ionized gas in the light of [O I] $\lambda6300$ Å and thus allow us to trace YSO wind morphologies directly. HST observations will determine whether energetic outflows interact with the disk (as opposed to leaving the system in highly collimated polar jets). In combination with improved estimates of mass outflow rates, we can then determine the integrated flux of energetic particles *through the disk* for a sample of PMS objects of differing age and thus assess the role of winds in the evolution of disks.

ACKNOWLEDGMENTS

REFERENCES

Adams, F.C., C. Lada, and F.H. Shu. 1987. Ap. J. 312:788.
Appenzeller, I., I. Jankovics, and R. Ostreicher. 1984. Astr. Ap. 141:108.
Backman, D. and F.C. Gillett. 1988. In: Linsky, J.L. and R. Stencel (eds.). Proceedings of the Fifth Cambridge Cool Star Workshop. Springer, Verlag.
Beckwith, S., B. Zuckerman, M.F. Skrutskie, and H.M. Dyck. 1984. Ap. J. 287:793.
Beckwith, S., and A. Sargent. 1987. Ap. J. 323:294.
Beckwith, S. *et al.* 1989. Ap. J., in press.

Bertout, C., G. Basri, and J. Bouvier. 1988. Ap. J., in press.

Cabrit, S., S. Edwards, S.E. Strom, and K.M. Strom. 1989. A. J., in preparation.

Chelli, A., L. Carrasco, H. Zinnecker, I. Cruz-Gonzalez, and C. Perrier. 1988. Astr. Ap., in press.

Chyba, C., and C. Sagan. 1987. Infrared-emission by organic grains in the coma of comet Halley. Nature 330 (6146):350-353.

Edwards, S., S. Cabrit, S.E. Strom, I. Heyer, K.M. Strom, and E. Anderson. 1987. Ap. J. 321:473.

Edwards, S., and S.E. Strom. 1988. In: Linsky, J.L., and R. Stencel (eds.). Proceedings of the Fifth Cambridge Cool Star Workshop. Springer, Verlag.

Grasdalen, G.L., S.E. Strom, K.M. Strom, R.W. Capps, D. Thompson, and M. Castelaz. 1984. Ap. J. 283:L57.

Grasdalen, G.L., G. Sloan, N. Stoudt, S. E. Strom, and A.D. Welty. 1989. Ap. J. (Letters), in press.

Greenberg, J.M. 1982. In: Wilkening, L., with M.S. Matthews (eds.). Comets. University of Arizona Press, Tucson.

Hartmann, L., R. Hewett, S. Stahler, and R.D. Mathieu. 1986. Ap. J. 309:275.

Hartmann, L., and S.J. Kenyon. 1987a. Ap. J. 312:243.

Hartmann, L., and S.J. Kenyon. 1987b. Ap. J. 322:393.

Hartmann, L. and W. Raymond. 1988. Ap. J., submitted.

Jones, B., and G.H. Herbig. 1979. A.J. 84:1872.

Kenyon, S., and L.W. Hartmann. 1987. Ap. J. 323:714.

Lanzerotti, L.J., W.L. Brown, and R.E. Johnson. 1986. Nuclear Instruments and Methods in Physics Research B14: 373.

Lynden-Bell, D., and J.E. Pringle. 1974. MNRAS 168:603.

Myers, P.C., Fuller, G.A., R.D. Mathieu, C.A. Beichman, P.J. Benson, R.E. Schild, and J.P. Emerson. 1987. Ap. J. 319: 340.

Rucinski, S. 1985. A.J. 90:2321.

Rydgren, A.E., and D.S. Zak. 1986. Pub. A.S.P. 99:141.

Shu, F., F. Adams, and S. Lizano. 1987. Ann. Rev. Astr. Ap. 25: 23.

Smith, B., and R. Terrile. 1984. Science 226:1421.

Strazzulla, G. 1985. Icarus 61:48.

Strom, S.E., K.M. Strom, G.L. Grasdalen, R.W. Capps, and D. Thompson. 1985. A.J. 290:587.

Strom, K.M., S.E. Strom, S. Kenyon, and L. Hartmann. 1988. A. J. 95:534.

Strom, S.E., K. M. Strom, and S. Edwards. 1988. In: Pudritz, R., and M. Fich (eds.). NATO Advanced Study Institute: Galactic and Extragalactic Star Formationed. Reidel, Dordrecht.

Strom, K., S. Cabrit, S. Edwards, M. Skrutskie, and S.E. Strom. 1989. A. J., submitted.

Walter, F. 1987. PASP 99:31.

Walter, F., A. Brown, R.D. Mathieu, P.C. Myers, and F.J. Vrba. 1988. A.J. 96:297.

Warner, J.W., S.E. Strom, and K.M. Strom. 1977. Ap. J. 213: 427.

Welty, A., S.E. Strom, L.W. Hartmann, and S.J. Kenyon. 1989. Ap. J., in preparation.

Numerical Two-Dimensional Calculations
of the Formation of the Solar Nebula

PETER H. BODENHEIMER
Lick Observatory

ABSTRACT

The protostellar phase of stellar evolution is of considerable impor-
tance with regard to the formation of planetary systems. The initial mass
distribution and angular momentum distribution in the core of a molecular
cloud determine whether a binary system or a single star is formed. A rel-
atively slowly rotating and centrally condensed cloud is likely to collapse to
a disk-like structure out of which planets can form. The above parameters
then determine the temperature and density structure of the disk and the
characteristics of the resulting planetary system.

There has been considerable recent interest in two-dimensional nu-
merical hydrodynamical calculations with radiative transfer, applied to the
inner regions of collapsing, rotating protostellar clouds of about 1 $M_\odot$.
The calculations start at a density that is high enough so that the gas
is decoupled from the magnetic field. During the collapse, mechanisms
for angular momentum transport are too slow to be effective, so that an
axisymmetric approximation is sufficiently accurate to give useful results.
Until the disk has formed, the calculations can be performed under the
assumption of conservation of angular momentum of each mass element.
In a numerical calculation, a detailed study of the region of disk formation
can be performed only if the central protostar is left unresolved.

With a suitable choice of initial angular momentum, the size of the
disk is similar to that of our planetary system. The disk forms as a relatively
thick, warm equilibrium structure, with a shock wave separating it from the
surrounding infalling gas. The calculations give temperature and density

17

distributions throughout the infalling cloud as a function of time. From these, frequency-dependent radiative transfer calculations produce infrared spectra and isophote maps at selected viewing angles. The theoretical spectra may be compared with observations of suspected protostellar sources. These disks correspond to the initial conditions for the solar nebula, whose evolution is then driven by processes that transport angular momentum.

OBSERVATIONAL CONSTRAINTS ON THE PROPERTIES OF THE INITIAL SOLAR NEBULA

From observations of physical and cosmochemical properties of the solar system and from astronomical observations of star-forming regions and young stars, certain constraints can be placed on the processes of formation and evolution of the solar nebula.

a) Low-mass stars form by the collapse of initially cold (10 K), dense (10^5 particles cm^{-3}) cores of molecular clouds. The close physical proximity of such cores with T Tauri stars, with imbedded infrared sources which presumably are protostars, and with sources with bipolar outflows, presumably coming from stars in a very early stage of their evolution, lends support to this hypothesis (Myers 1987).

b) The specific angular momenta (j) of the cores, where observed, fall in the range $10^{20} - 10^{21}$ cm^2 s^{-1} (Goldsmith and Arquilla 1985; Heyer 1988). In the lower end of this range, the angular momenta are consistent with the properties of our solar system: for example, Jupiter's orbital motion has j $\approx 10^{20}$ cm^2 s^{-1}. In the upper end of the range, collapse with conservation of angular momentum would lead to a halt of the collapse as a consequence of rotational effects at a characteristic size of $\sim$ 2000 AU, far too large to account for the planetary orbits. In fact, hydrodynamical calculations suggest that collapse in this case would in fact lead to fragmentation into a binary or multiple system.

c) The infrared radiation detected in young stars indicates the presence of disks around these objects (Hartmann and Kenyon 1988). A particularly good example, where orbital motions have been observed, is HL Tau (Sargent and Beckwith 1987). The deduced masses of the disk and star are 0.1 $M_\odot$ and 1.0 $M_\odot$, respectively. The radius of the disk is about 2000 AU. Roughly half of all young pre-main-sequence stars are deduced to have disks, mostly unresolved, with masses in the range 0.01-0.1 $M_\odot$ and sizes from 10 to 100 AU (Strom et al. 1989).

d) The rotational velocities of T Tauri stars are small, typically 20 km s^{-1} or less (Hartmann et al. 1986). The distribution of angular momentum in the system consisting of such a star and disk is quite different from that in the core of a molecular cloud, which is generally assumed to be

uniformly rotating with a power-law density distribution. Substantial angular momentum transport, from the central regions to the outer regions, must take place early in the evolution. The required transport is unlikely to occur during collapse; therefore it must occur during the disk evolution phase before the star emerges as a visible object.

e) The lifetime of the pre-main-sequence disks is difficult to determine from observation, but it probably does not exceed 10^7 years (Strom *et al.* 1989). The mechanisms for angular momentum transport, which deplete disk mass by allowing it to fall into the star, must have time scales consistent with these observations, as well as with time scales necessary to form gaseous giant planets.

f) The temperature conditions in the early solar nebula can be roughly estimated from the distribution of the planets' and satellites' mass and chemical composition (Lewis 1974). The general requirements are that the temperature be high enough in the inner regions to vaporize most solid material, and that it be low enough at the orbit of Jupiter and beyond to allow the condensation of ices. Theoretical models of viscous disks produce the correct temperature range, as do collapse models of disk formation with shock heating.

g) The evidence from meteorites is difficult to interpret in terms of standard nebular models. First, there is evidence for the presence of magnetic fields, and second, the condensates indicate the occurrence of rapid and substantial thermal fluctuations. Suggestions for explaining this latter effect include turbulent transport of material and non-axisymmetric structure (density waves) in the disk.

h) The classical argument, of course, is that the coplanarity and circularity of the planets' orbits imply that they were produced in a disk.

i) A large fraction of stars are observed to be in binary and multiple systems (Abt 1983); the orbital values of j in the closer systems are comparable to those in our planetary system. It has been suggested (Boss 1987; Safronov and Ruzmaikina 1985) that if the initial cloud is slowly rotating and centrally condensed, it is likely to form a single star rather than a binary. Pringle (1989) has pointed out that if a star begins collapse after having undergone slow diffusion across the magnetic field, it will be centrally condensed and will therefore form a single star. If, however, the collapse is induced by external pressure disturbances, the outcome is likely to be a binary. On the other hand, Miyama (1989) suggests that single star formation occurs in initial clouds with $j \approx 10^{21}$ cm^2 s^{-1}. After reaching a rotationally supported equilibrium that is stable to fragmentation, the cloud becomes unstable to nonaxisymmetric perturbations, resulting in angular momentum transport and collapse of the central regions.

THE PHYSICAL PROBLEM

The above considerations illustrate several of the important questions relating to the formation of the solar nebula: What are the initial conditions for collapse of a protostar? At what density does the magnetic field decouple from the gas? What conditions lead to the formation of a single star with a disk rather than a double star? Can the embedded IRAS sources be identified with the stage of evolution just after disk formation? What is the dominant mechanism for angular momentum transport that produces the present distribution of angular momentum in the solar system? The goal of numerical calculations is to investigate these questions by tracing the evolution of a protostar from its initial state as an ammonia core in a molecular cloud to the final quasi-equilibrium state of a central star, which is supported against gravity by the pressure gradient, and a circumstellar disk, which is supported in the radial direction primarily by centrifugal effects. A further goal is to predict the observational properties of the system at various times during the collapse. A full treatment would include a large number of physical effects: the hydrodynamics, in three space dimensions, of a collapsing rotating cloud with a magnetic field; the equation of state of a dissociating and ionizing gas of solar composition, cooling from molecules and grains in optically thin regions; frequency-dependent radiative transfer in optically thick regions; molecular chemistry; the generation of turbulent motions as the disk and star approach hydrostatic equilibrium; and the properties of the radiating accretion shock which forms at the edge of the central star and on the surfaces of the disk (Shu *et al.* 1987).

The complexity of this problem makes a general solution intractable even on the fastest available computing machinery. For example, the length scales range from 10^{17} centimeters, the typical dimension of the core of a molecular cloud, to 10^{11} centimeters, the size of the central star. The density of the material that reaches the star undergoes an increase of about 15 orders of magnitude from its original value of $\sim 10^{-19}$ g cm^{-3}. Also, the numerical treatment of the shock front must be done very carefully. The number of grid points required to resolve the entire structure is very large in two space dimensions; in three dimensions it is prohibitively large. Even if the detailed structure of the central object is neglected and the system is resolved down to a scale of 0.1 AU, the Courant-Friedrichs-Lewy condition in an explicit calculation requires that the time step be less than one-millionth of the collapse time of the cloud. Therefore, a number of physical approximations and restrictions have been made in all recent numerical calculations of nebular formation. For example, magnetic fields have not been included, on the grounds that the collapse starts only when the gas has become almost completely decoupled from the field because of the negligible degree of ionization at the relatively

high densities and cold temperatures involved. Also, in most calculations, turbulence has been neglected during the collapse. Even if it is present, the time scale for transport of angular momentum by this process is expected to be much longer than the dynamical time. It turns out that angular momentum transport can be neglected during the collapse, and therefore an axisymmetric (two-dimensional) approximation is adequate during this phase. Three-dimensional effects, such as angular momentum transport by gravitational torques become important later, during the phase of nebular evolution. A further approximation involves isolating and resolving only specific regions of the protostar. In one-dimensional calculations (Stahler *et al.* 1980), it has been possible to resolve the high-density core as well as the low-density envelope of the protostar. However, in two space dimensions, proper resolution of the region where the nebular disk forms cannot be accomplished simultanously with the resolution of the central star. In several calculations the outer regions of the protostar are also not included, so the best possible resolution can be obtained on the length scale 1-50 AU. Thus the goal outlined above, the calculation of the evolution of a rotating protostar all the way to its final stellar state, has not yet been fully realized.

The stages of evolution of a slowly rotating protostar of about 1 $M_\odot$ can be outlined as follows:

a)   The frozen-in magnetic field transfers much of the angular momentum out of the core of the molecular cloud, on a time scale of 10^7 years.

b)   The gradual decoupling between the magnetic field and the matter allows the gas to begin to collapse, with conservation of angular momentum.

c)   The initial configuration is centrally condensed. During collapse, the outer regions, with densities less than about 10^{-13} g cm^{-3} remain optically thin and collapse isothermally at 10 K. The gas that reaches higher densities becomes optically thick, most of the released energy is trapped, and heating occurs.

d)   The dust grains, which provide most of the opacity in the protostellar envelope, evaporate when the temperature exceeds 1500 K. An optically thin region is created interior to about 1 AU. Further, at temperatures above 2000 K, the molecular hydrogen dissociates, causing renewed instability to collapse.

e)   The stellar core and disk form from the inner part of the cloud. The remaining infalling material passes though accretion shocks at the boundaries of the core and disk; most of the infall kinetic energy is converted into radiation behind the shock. The surrounding infalling material is optically thick, and the object radiates in the infrared, with a peak at around 60-100 μm.

f)   A stellar wind is generated in the stellar core, by a process that is not well understood. The wind breaks through the infalling gas at the rotational poles, where the density gradient is steepest and where most of the material has already fallen onto the core. This bipolar outflow phase lasts about 10^5 years.

g)   Infall stops because of the effects of the wind, or simply because the material is exhausted. The stellar core emerges onto the Hertzsprung-Russell diagram as a T Tauri star, still with considerable infrared radiation coming from the disk.

h)   The disk evolves, driven by processes that transfer angular momentum, on a timescale of 10^6 to 10^7 years. Angular momentum is transferred outwards through the disk while mass is transferred inwards. The rotation of the central object slows, possibly through magnetic braking in the stellar wind. Possible transport processes in the disk include turbulent (convectively driven) viscosity, magnetic fields, and gravitational torques driven by gravitational instability in the disk or by non-axisymmetric instabilities in the initially rapidly rotating central star.

The following sections describe numerical calculations of phases b through e, from the time when magnetic effects become unimportant to the time when at least part of the infalling material is approaching equilibrium in a disk.

REVIEW OF TWO-DIMENSIONAL CALCULATIONS OF THE FORMATION PHASE

Modern theoretical work on this problem goes back to the work of Cameron (1962, 1963), who discussed in an approximate way the collapse of a protostar to form a disk. In a later work, Cameron (1978) solved numerically the one-dimensional (radial) equations for the growth of a viscous accretion disk, taking into account the accretion of mass from an infalling protostellar cloud. The initial cloud was assumed to be uniformly rotating with uniform density. The hydrodynamics of the inflow was not calculated in detail; rather, infalling matter was assumed to join the disk at the location where its angular momentum matched that of the disk. A similar approach was taken by Cassen and Summers (1983) and Ruzmaikina and Maeva (1986), who, however, took into account the drag caused by the infalling material, which has angular momentum different from that of the disk at the arrival point. The latter authors discuss the turbulence that develops for the same reason (see also Safronov and Ruzmaikina 1985). This section concentrates on full two-dimensional calculations of the collapsing cloud during the initial formation of the disk.

One approach to this problem (Tscharnuter 1981; Regev and Shaviv 1981; Morfill *et al.* 1985; Tscharnuter 1987) is based on the assumption that

turbulent viscosity operates during the collapse. The resulting transport of angular momentum out of the inner parts of the cloud might be expected to suppress the fragmentation into a binary system and to reduce the angular momentum of the central star to the point where it is consistent with observations of T Tauri stars. The procedure is to assume a simple kinematic viscosity $\nu = 0.33 \, \alpha c_s L$, where c_s is the sound speed, L is the length scale of the largest turbulent eddies, and α is a free parameter. Subsonic turbulence is generally assumed, so that α is less than unity. Once the collapse is well underway, the sound crossing time is much longer than the free fall time, so that angular momentum transport is actually relatively ineffective. Binary formation is probably suppressed, but the material that falls into the central object still has high angular momentum compared with that of a T Tauri star.

Nevertheless, the model presented by Morfill *et al.* (1985) provides interesting information regarding the initial solar nebula. In contrast to the earlier calculation of Regev and Shaviv (1981), which used the isothermal approximation, this two-dimensional numerical calculation included the full hydrodynamical equations, applicable in both the optically thin and optically thick regions. Radiation transport was included in the Eddington approximation. The calculations started with 3 $M_\odot$ at a uniform density of 10^{-20} g cm^{-3} and with uniform angular velocity. Two different values for j were tested, 10^{21} and 10^{20} cm^2 s^{-1}, with qualitatively similar results. The case with lower angular momentum is the one of most interest. The collapse proceeds and results in the formation of a central condensation surrounded by a disk. However, the core does not reach hydrostatic equilibrium but exhibits a series of dynamical oscillations, driven, according to the authors, by heat generated through viscous dissipation in the region near the edge of the core. The dynamical expansion is accompanied by an outgoing thermal pulse.

So that the development of the disk could be studied, the computational procedure was modified to treat the region interior to 2×10^{12} cm as an unresolved core, and thereby to suppress the oscillations. Matter and angular momentum were allowed to flow into this central region but not out of it. The kinetic energy of infall was assumed to be converted into radiation at the same boundary. The calculation was continued until about 0.5 $M_\odot$ had accumulated in the core, and about 0.1 $M_\odot$ had collapsed into a nearly Keplerian disk, with radial extent of about 20 AU. The calculation was stopped at that point because of insufficient spatial resolution in the disk region, and because the ratio (β) of rotational kinetic energy to gravitational potential energy of the core exceeded 0.27, so dynamical instability to non-axisymmetric perturbations would be likely (Durisen and Tohline 1985). The development of a triaxial central object is likely to result in the transport of angular momentum by gravitational torques (Yuan and Cassen

1985; Durisen *et al.* 1986). Angular momentum would be transported from the central object to the disk, and the value of β for the core would be reduced below the critical value. However, its remaining total angular momentum would be still too large to allow it to become a normal star.

A further important feature of the calculation was its prediction of the temperatures that would be generated in the planet-forming region. Over a time scale of 3×10^4 years the temperature of material with the same specific angular momentum as that of the orbit of Mercury ranged from 400-600 K. The predicted temperature for Jupiter remained fairly constant at 100 K, while that for Pluto approximated 15 K. In the inner region of the nebula these temperatures are slightly cooler than those generally thought to exist during planetary formation or those calculated in evolving models of a viscous solar nebula (Ruden and Lin 1986).

A further calculation was made by Tscharnuter (1987) with similar physics but a different initial condition. A somewhat centrally condensed and non-spherical cloud of 1.2 $M_\odot$ starts collapse from a radius of 4×10^{15} cm, a mean density of 8×10^{-15} g cm^{-3}, and j $\approx 10^{20}$ cm^2 s^{-1}. A major improvement was a refined equation of state. The use of this equation of state to calculate the collapse of a spherically symmetric protostar starting from a density of 10^{-19} g cm^{-3} produces violent oscillations in central density and temperature after the stellar core has formed. The instability is triggered when the adiabatic exponent $\Gamma_1 = (\partial \ln P / \partial \ln \rho)_s$ falls below 4/3, and the source of the energy for the reexpansion is association of hydrogen atoms into molecules.

In the two-dimensional case, the much higher starting density and the correspondingly higher mass inflow rate onto the core, as well as the effects of rotation, are sufficient to suppress the instability. A few relatively minor oscillations, primarily in the direction of the rotational pole, dampen quickly. The numerical procedure uses a grid that moves in the (spherical) radial direction and thus is able to resolve the central regions well, down to a scale of 10^{10} cm. This calculation is carried to the point where a fairly well-defined core of 0.07 $M_\odot$ has formed, which is still stable to non-axisymmetric perturbations ($\beta = 0.08$). A surrounding disk structure is beginning to form, out to a radius of about 1 AU. The density and temperature in the equatorial plane at that distance are about 3×10^{-9} g cm^{-3} and 2500 K, respectively. Further accretion of material into the core region is likely to increase the value of β. The calculation was not continued because of the large amount of computer time required.

A different approach to the problem of the two-dimensional collapse of the protostar has been considered by Adams and Shu (1986) and Adams, *et al.* (1987). The aim is to obtain emergent spectra through frequency-dependent radiative transfer calculations. In order to bypass the difficulties

of a full two-dimensional numerical calculation, they made several approximations. The initial condition is a "singular" isothermal sphere, in unstable equilibrium, with sound speed c_s, uniformly rotating with angular velocity Ω. In the initial state the density distribution is given by $\rho \propto R^{-2}$, where R is the distance to the origin, and the free-fall accretion rate onto a central object of mass M is given by $\dot{M} = 0.975 \, c_s^3/G$. The hydrodynamical solution for the infalling envelope is taken to be that given by Terebey et al. (1984), a semianalytic solution under the approximation of slow rotation. The thermal structure and radiation transport through the envelope can be decoupled from the hydrodynamics (Stahler et al. 1980). The model at a given time consists of an (unresolved) core, a circumstellar disk, and a surrounding infalling, dusty, and optically thick envelope. The radiation produced at the accretion shocks at the core and disk is reprocessed in the envelope, and emerges at the dust photosphere, primarily in the mid-infrared. The thermal emission of the dust in the envelope is obtained by approximating the rotating structure as an equivalent spherical structure; however, the absorption in the equation of transfer is calculated taking the full two-dimensional structure into account. The model is used to fit the observed infrared radiation from a number of suspected protostars, by variation of the parameters M, c_s, Ω, η_D, and η_*, where the last two quantities are the efficiencies with which the disk transfers matter onto the central star and with which it converts rotational energy into heat and radiation, respectively. These models provide good fits to the spectra of the observed sources for typical parameters $M = 0.2 - 1.0 \, M_\odot$, $c_s = 0.2 - 0.35 \, km \, s^{-1}$, $\Omega = 2 \times 10^{-14} - 5 \times 10^{-13} \, rad \, s^{-1}$, $\eta_D = 1$ and $\eta_* = 0.5$. Of particular interest is the fact that in many cases the deduced values of Ω fall in the range $j \approx 10^{20} \, cm^2 \, s^{-1}$, which is appropriate for "solar nebula" disks. The contribution from the disk broadens the spectral energy distribution and brings it into better agreement with the observations than does the non-rotating model. More recent observational studies of protostellar sources (Myers et al. 1987; Cohen et al. 1989) also are consistent with the hypothesis that disks have formed within them.

RECENT MODELS WITH HYDRODYNAMICS AND RADIATIVE TRANSPORT

Full hydrodynamic calculations of the collapse, including frequency-dependent radiative transport, have recently been reported by Bodenheimer et al. (1988). The purpose of the calculations was to obtain the detailed structure of the solar nebula at a time just after its formation and to obtain spectra and isophotal contours of the system as a function of viewing angle and time. Because of the numerical difficulties discussed above, the protostar was resolved only on scales of $10^{13} - 10^{15}$ cm. These calculations

have now been redone with the extension of the outer boundary of the grid to 5×10^{15} cm, with improvements in the radiative transport, and with a somewhat better spatial resolution, about 1 AU in the disk region (Bodenheimer *et al.* 1990).

The initial state, a cloud of 1 $M_\odot$ with a mean density of 4×10^{-15} g cm^{-3}, can be justified on the grounds that only above this value does the magnetic field decouple from the gas and allow a free-fall collapse, with conservation of angular momentum of each mass element, to start (Nakano 1984; Tscharnuter 1987). The initial density distribution is assumed to be a power law, the temperature is assumed to be isothermal at 20 K, and the angular velocity is taken to be uniform with a total angular momentum of 10^{53} g cm^2 s^{-1}. Because the cloud is already optically thick at the initial state, the temperature increases rapidly once the collapse starts. The inner region with $R \leq 1$ AU is unresolved; the mass and angular momentum that flow into this core are calculated. At any given time, a crude model of this material is constructed under the assumption that it forms a Maclaurin spheroid. From a calculation of its equatorial radius R_e, the accretion luminosity $L = GMM/R_e$ is obtained. For each timestep Δt the accretion energy $L\Delta t$ is deposited in the inner zone as internal energy and is used as an inner boundary condition for the radiative transfer. Most of the energy radiated by the protostar is provided by this central source. During the hydrodynamic calculations, radiative transfer is calculated according to the diffusion approximation, which is a satisfactory approximation for an optically thick system. Rosseland mean opacities were taken from the work of Pollack *et al.* (1985). After the hydrodynamic calculations were completed, frequency-dependent radiative transfer was calculated for particular models according to the approach of Bertout and Yorke (1978), with their grain opacities which include graphite, ice, and silicates.

The results of the calculations show the formation of a rather thick disk, with increasing thickness as a function of distance from the central object. As a function of time the outer edge of the disk spreads from 1 AU to 60 AU, because of the accretion of material of higher angular momentum. The shock wave on the surface is evident, and the internal motions in the disk are relatively small compared with the collapse velocities. At the end of the calculation the mass of the disk is comparable to that of the central object, and it is not gravitationally unstable according to the axisymmetric local criterion of Toomre (1964). The central core of the protostar, inside 10^{13} cm, contains about 0.6 $M_\odot$ and sufficient angular momentum so that $\beta \approx 0.4$. This region is almost certainly unstable to bar-like perturbations. Theoretical spectra show a peak in the infrared at about 40μ; when viewed from the equator the wavelength of peak intensity shifts redward from that at the pole. A notable difference between equator and pole is evident in the isophotal contours. At 40 μm, for example, the peak intensity

shifts spatially to points above and below the equatorial plane because of heavy obscuration there. This effect becomes more pronounced at shorter wavelengths. Maximum temperatures in the midplane of the disk reached 1500 K in the distance range 1-10 AU. At the end of the calculation, after an elapsed time of 2500 years, these temperatures ranged from 700 K at 2 AU to 500 K at 10 AU and were decreasing with time.

FURTHER EVOLUTION OF THE SYSTEM

In the preceding example most of the infalling material joined the disk or central object on a short time scale, because of the high initial density. For a lower initial density, processes of angular momentum transport in the disk would begin before accretion was completed. The problem of the rapidly spinning central regions is apparently not solved by including angular momentum transport by turbulent viscosity during the collapse phase. Furthermore, no plausible physical mechanism for generating turbulence on the appropriate scale has been demonstrated. The angular momentum transport resulting from gravitational torques arising from the non-axisymmetric structure of the central regions is likely to leave them with values of β near 0.2 (Durisen *et al.* 1986). Therefore, even further transport is required. A related mechanism has been explored by Boss (1985, 1989). He has made calculations of protostar collapse, starting from uniform density and uniform angular velocity, with a three-dimensional hydrodynamic code, including radiation transport in optically thick regions. Small, initial non-axisymmetric perturbations grow during the collapse, so that the central regions, on a scale of 10 AU, become significantly non-axisymmetric even before a quasi-equilibrium configuration is reached. The deduced time scales for angular momentum transport depend on the initial conditions but range from 10^3 to 10^6 years for systems with a total mass of 1 $M_\odot$. However, since the evolution has not actually been calculated over this time scale, it is not clear how long the non-axisymmetry will last or how it will affect the angular momentum of the central core.

It is likely that some additional process is required to reduce j of the central object down to the value of 10^{17} characteristic of T Tauri stars. The approach of Safronov and Ruzmaikina (1985) is to assume that the initial cloud had an even smaller angular momentum (j $\approx 10^{19}$ cm^2 s^{-1}) than that assumed in most other calculations discussed here. The cloud would then collapse and form a disk with an equilibrium radius much less than that of Jupiter's orbit. Outward transport of angular momentum into a relatively small amount of mass is then required to produce the solar nebula. Magnetic transport could be important in the inner regions, which are warm and at least partially ionized (Ruzmaikina 1981). However, outside about 1 AU (Hayashi 1981) the magnetic field decays faster than it

amplifies, and magnetic transport is ineffective. A supplementary mechanism must be available to continue the process. One possibility is the turbulence generated in the surface layers of the disk, caused by the shear between disk matter and infalling matter. Another possibility is that the initial cloud had higher j, and the rapidly spinning central object is braked through a centrifugally driven magnetic wind which can remove the angular momentum relatively quickly (Shu *et al.* 1988).

As far as the evolution of the disk itself is concerned, other important mechanisms that have been suggested include (a) gravitational instability; (b) turbulent viscosity induced by convection, and (c) sound waves and shock dissipation. The former can occur if the disk is relatively massive compared with the central star or if the disk is relatively cold. Although it is still an open question whether this instability can result in the formation of a binary or preplanetary condensations, the most likely outcome is the spreading out of such condensations, because of the shear, into spiral density waves (Larson 1983). Lin and Pringle (1987) have estimated the transport time to be about 10 times the dynamical time. Processes (b) and (c) have time scales more in line with the probable lifetimes of nebular disks. Convective instability in the vertical direction (Lin and Papaloizou 1980), induced by the temperature dependence of the grain opacities, gives disk evolutionary times of about 10^6 years (Ruden and Lin 1986). An alternate treatment of the convection (Cabot *et al.* 1987a,b) gives a time scale longer roughly by a factor of 10. Sound waves induced by various external perturbations give transport times in the range 10^6 to 10^7 years (Larson 1989). A complete theory of how the disk evolves after the immediate formation phase may involve several of the mechanisms just mentioned, and its development will require a considerable investment of thought and numerical calculation.

ACKNOWLEDGEMENTS

REFERENCES

Abt, H.A. 1983. Ann. Rev. Astron. Astrophys. 21:343.
Adams, F.C., C.J. Lada, and F.H. Shu. 1987. Astrophys. J. 312:788.
Adams, F.C., and F.H. Shu. 1986. Astrophys. J. 308:836.
Bertout, C., and H.W. Yorke. 1978. Page 648. In: Gehrels, T. (ed.). Protostars and Planets. University of Arizona Press, Tucson.

Bodenheimer, P., M. Rozyczka, H.W. Yorke, and J.E. Tohline. 1988. Page 139. In: Dupree, A.K., and M.T.V.T. Lago (eds.). Formation and Evolution of Low-Mass Stars. Kluwer, Dordrecht.

Bodenheimer, P., H.W. Yorke, M. Rozyczka, and J.E. Tohline. 1990. J. Astrophys. J. 355:651.

Boss, A.P. 1985. Icarus 61:3.

Boss, A.P. 1987. Astrophys. J. 319:149.

Boss, A.P. 1989. Astrophys. J. 345:554.

Cabot, W., V.M. Canuto, O. Hubickyj, and J.B. Pollack. 1987a. Icarus 69:387.

Cabot, W., V.M. Canuto, O. Hubickyj, and J.B. Pollack. 1987b. Icarus 69:423.

Cameron, A.G.W. 1962. Icarus 1:13.

Cameron, A.G.W. 1963. Icarus 1:339.

Cameron, A.G.W. 1978. Moon and Planets 18:5.

Cassen, P., and A. Summers. 1983. Icarus 53:26.

Cohen, M., J.P. Emerson, and C.A. Beichman. 1989. Astrophys. J. 339:455.

Durisen, R.H., R.A. Gingold, J.E. Tohline, and A.P. Boss. 1986. Astrophys. J. 305:281.

Durisen, R.H., and J.E. Tohline. 1985. Page 534. In: Black, D.C., and M.S. Matthews (eds.). Protostars and Planets II. University of Arizona Press, Tucson.

Goldsmith, P.F., and R. Arquilla. 1985. Page 137. In: Black, D.C. and M.S. Matthews (eds.). Protostars and Planets II. University of Arizona Press, Tucson.

Hartmann, L., R. Hewitt, S. Stahler, and R.D. Mathieu. 1986. Astrophys. J. 309:275.

Hartmann, L., and S. Kenyon. 1988. Page 163. In: Dupree, A.K. and M.T.V.T. Lago (eds.). Formation and Evolution of Low-Mass Stars. Kluwer, Dordrecht.

Hayashi, C. 1981. Prog. Theor. Phys. Suppl. 70:35.

Heyer, M.H. 1988. Astrophys. J. 324:311.

Larson, R.B. 1983. Rev. Mexicana Astron. Astrof. 7:219.

Larson, R.B. 1989. Page 31. In: Weaver, H.A., and L. Danly (eds.). The Formation and Evolution of Planetary Systems. Cambridge University Press, Cambridge.

Lewis, J.S. 1974. Science 186:440.

Lin, D.N.C., and J. Papaloizou. 1980. Mon. Not. R. astr. Soc. 191:37.

Lin, D.N.C., and J.E. Pringle. 1987. Mon. Not. R. astr. Soc. 225:607.

Miyama, S. 1989. Page 284. In: Weaver, H.A., and L. Danly (eds.). The Formation and Evolution of Planetary Systems. Cambridge University Press, Cambridge.

Morfill, G.E., W. Tscharnuter, and H.J. Volk. 1985. Page 493. In: Black, D.C. and M.S. Matthew (eds.). Protostars and Planets II. University of Arizona Press, Tucson.

Myers, P.C. 1987. Page 33. In: Peimbert, M. and J. Jagaku (eds.). Star Forming Regions (IAU Symposium 115). Reidel, Dordrecht.

Myers, P.C., G.A. Fuller, R.D. Mathieu, C.A. Beichman, P.J. Benson, R.E. Schild, and J.P. Emerson. 1987. Astrophys. J. 319:340.

Nakano, T. 1984. Fund. Cosmic Phys. 9:139.

Pollack, J.B., C. McKay, and B. Christofferson. 1985. Icarus 64:471.

Pringle, J.E. 1989. Mon. Not. R. astr. Soc., 239:361.

Regev, O., and G. Shaviv. 1981. Astrophys. J. 245:934.

Ruden, S.P., and D.N.C. Lin. 1986. Astrophys. J. 308:883.

Ruzmaikina, T.V. 1981. Adv. Space Res. 1:49.

Ruzmaikina, T.V., and S.V. Maeva. 1986. Astron. Vestn. 20(3): 212.

Safronov, V.S., and T.V. Ruzmaikina. 1985. Page 959. In: Black, D.C., and M.S. Matthews (eds.). Protostars and Planets II. University of Arizona Press, Tucson.

Sargent, A.I., and S. Beckwith. 1987. Astrophys. J. 323:294.

Shu, F.H., F.C. Adams, and S. Lizano. 1987. Ann. Rev. Astron. Astrophys. 25:23.

Shu, F.H., S. Lizano, F.C. Adams, and S.P. Ruden. 1988. Page 123. In: Dupree, A.K. and M.T.V.T. Lago (eds.). Formation and Evolution of Low-Mass Stars. Kluwer, Dordrecht.

Stahler, S.W., F.H. Shu, and R.E. Taam. 1980. Astrophys. J. 241:637.

Strom, S.E., S. Edwards, and K.M. Strom. 1989. Page 91. In: Weaver, H.A., and L. Danly (eds.). The Formation and Evolution of Planetary Systems. Cambridge University Press, Cambridge.

Terebey, S., F.H. Shu, and P. Cassen. 1984. Astrophys. J. 286:529.
Toomre, A. 1964. Astrophys. J. 139:1217.
Tscharnuter, W. 1981. Page 105. In: Sugimoto, D., D.Q. Lamb, and D.N. Schramm (eds.). Fundamental Problems in the Theory of Stellar Evolution (IAU Symposium 93). Reidel, Dordrecht.
Tscharnuter, W. 1987. Astron. Astrophys. 188:55.
Yuan, C., and P. Cassen. 1985. Icarus 64:41.

Three-Dimensional Evolution of
Early Solar Nebula

ALAN P. BOSS
Carnegie Institution of Washington

INITIAL CONDITIONS FOR PROTOSTELLAR COLLAPSE

Mathematically speaking, solar nebula formation is an initial value problem. That is, it is believed that given the proper initial conditions and knowledge of the dominant physical processes at each phase, it should be possible to calculate the evolution of a dense molecular cloud core as it collapses to form a protosun and solar nebula. Also, through calculating the evolution of the dust grains in the nebula, it should be possible to learn how the initial phases of planetary accumulation occurred. While this extraordinarily ambitious goal has not yet been achieved, considerable progress has been made in the last two decades, with the assistance of current computational resources. This paper reviews the progress toward the goal of a complete theory of solar nebula formation, with an emphasis on three spatial dimension (3D) models of solar nebula formation and evolution.

In principle, astronomical observations should provide the initial conditions required for theoretical calculations of solar nebula formation. This assumes that physical conditions in contemporary regions of solar-type star formation in our galaxy are similar to the conditions $\approx 4.56 \times 10^9$ years ago. Whether or not this is a reasonable assumption, considering that the age of the solar system is a fair fraction of the Hubble time, there really is no other means of constraining the initial conditions for solar system formation.

Millimeter wave and infrared telescopes have revealed that low-mass (solar-type) stars are presently forming preferentially in groups distributed throughout large molecular cloud complexes in the disk of our galaxy. These

31

molecular clouds have complex, often filamentary structures on the largest scales ($\sim$100pc). On the smallest scales ($\sim$0.1pc), where finite telescope resolution begins to limit the observations, molecular clouds are composed of centrally condensed cloud cores surrounded by cloud envelopes. The cloud cores are gravitationally bound and, should they begin to collapse upon themselves because of self-gravity, are quite likely to form stars (Myers and Benson 1983). Indeed, when infrared observations of embedded protostellar objects (presumed to be embedded pre-main-sequence T Tauri stars or very young protostars) are combined with millimeter wave maps of molecular cloud cores, it appears that roughly one half of the cloud cores contain embedded protostars (Beichman *et al.* 1986). The physical conditions in molecular cloud cores, or their predecessors, should then provide the best indication of the initial conditions that are appropriate for solar nebula formation.

Practically speaking, there are several reasons why astronomical observations can only provide us with a range of possible initial conditions for solar nebula formation. First, since even the short time scales associated with low-mass star formation ($\sim$10^5 – 10^6 years) are quite long compared to human lifetimes, one can never be sure what a particular collapsing cloud core will produce. Second, although interferometric arrays have the potential to greatly increase our understanding of cloud core properties, the length scales appropriate for the initial phases of collapse are only marginally resolved by current millimeter wave telescopes. Third, because many (if not most) cloud cores have already collapsed to form protostars, their properties may not be appropriate for constraining the earliest phases of collapse. If the immediate predecessors of cloud cores could be identified, then the constraints on the initial conditions for protostellar collapse would be improved considerably.

Given these limitations, observations of cloud cores suggest the following initial conditions for solar nebula formation: 1) Cloud cores have masses in the range of 0.1 to 10$M_\odot$, implying that evolution within the molecular cloud complex has already reduced the mass of sub-structures by several orders of magnitude, from masses characteristic of giant molecular clouds ($\sim$10^5 – 10$^6 M_\odot$), to masses in the stellar range; 2) Dense cloud cores have maximum densities on the order of 10^{-19} – 10^{-18} g cm^{-3}, sizes less than 0.1 pc, and temperatures close to 10K. These are basically the same initial conditions that have long been used to model protosolar collapse (Larson 1969).

One of the remaining great uncertainties is the amount of rotation present in cloud cores, because even the most rapidly rotating clouds have Doppler shifts comparable to other sources of line broadening, such as thermal broadening, translational cloud motions, and turbulence. While there is strong evidence that at least some dense clouds rotate close to

centrifugal equilibrium (specific angular momentum J/M $\sim 10^{21}$ cm^2s^{-1} for solar mass clouds), it is unknown how common such rapid rotation is, or how slowly clouds can rotate. Three-dimensional calculations (described in the next sub-section) indicate that the initial amount of rotation is critical for determining whether clouds collapse to form single or multiple protostars.

Finally, most three-dimensional calculations of protostellar collapse have ignored the possible importance of magnetic fields, in no small part because of the computational difficulties associated with their inclusion in an already formidable problem. OH Zeeman measurements of magnetic field strengths in molecular clouds yield values ($\sim 30\mu G$) in cool clouds implying that magnetic fields dominate the dynamics on the largest scales ($\sim 10 - 100pc$). However, there is evidence from the lack of correlations between magnetic field directions and dense cloud minor and rotational axes that, on the smaller scales (and higher densities) of dense cloud cores, magnetic fields no longer dominate the dynamics (Heyer 1988). Loss of magnetic field support is probably caused by ambipolar diffusion of the ions and magnetic field lines during contraction of the neutral bulk of the cloud. The evidence for decreased importance of magnetic fields at densities greater than $\sim 10^{-20}$ g cm^{-3} suggests that nonmagnetic models may be adequate for representing the gross dynamics of solar nebula formation.

SINGLE VERSUS BINARY STAR FORMATION

The first numerical models of the collapse of interstellar clouds to form solar-type stars disregarded the effects of rotation, thereby reducing the problem to spherical symmetry and the mathematics to one-dimensional (1D) equations (Bodenheimer 1968). The assumption of spherical symmetry ensures that a single star will result, but unfortunately such calculations can say nothing about binary star or planet formation. The first major dynamical problem in solar nebula formation is avoiding fragmentation into a binary protostar.

Larson (1969) found that 1D clouds collapse non-homologously, forming a protostellar core onto which the remainder of the cloud envelope accretes. Once the envelope is accreted, the protostar becomes visible as a low-mass, pre-main-sequence star (T Tauri star). Models based on these assumptions have done remarkably well at predicting the luminosities of T Tauri stars (Stahler 1983). However, there is a recent suggestion (Tscharnuter 1987a) that the protostellar core may not become well established until much later in the overall collapse than was previously thought. The 1D models also showed that the first six or so orders of magnitude increase

in density during protostellar collapse occur isothermally, a thermodynamical simplification employed in many of the two-dimensional (2D) and three-dimensional (3D) calculations that followed.

The first models to include the fact that interstellar clouds must have finite rotation were restricted to two dimensions, with an assumed symmetry about the rotation axis (axisymmetry) being assumed (Larson 1972; Black and Bodenheimer 1976). These models showed that a very rapidly rotating, isothermal cloud collapses and undergoes a centrifugal rebound in its central regions, leading to the formation of a self-gravitating ring. While a similar calculation yielded a runaway disk that initially cast doubt on the physical reality of ring formation (Norman *et al.* 1980), it now appears that most of the ring versus disk controversy can be attributed to differences in the initial conditions studied and to the possibly singular nature of the standard test problem. With a single numerical code, Boss and Haber (1982) found three possible outcomes for the collapse of rotating, isothermal, axisymmetric clouds: quasi-equilibrium Bonnor-Ebert spheroids, rings, and runaway disks, with the outcome being a simple function of the initial conditions. Two-dimensional clouds with initially high thermal and rotational energies do not undergo significant collapse, but relax into rotationally flattened, isothermal equilibrium configurations that are the 2D analogues of the 1D Bonnor-Ebert sphere. High specific angular momentum 2D clouds that undergo significant collapse form rings, while slowly rotating 2D clouds ($J/M < 10^{20}$ cm^2 s^{-1}) can collapse to form isothermal disks (Terebey *et al.* 1984).

Whereas the Sun is a single star, a cloud that forms a ring is likely to fragment into a multiple protostellar system (Larson 1972). Because of this, subsequent axisymmetric presolar nebula models have concentrated on slowly rotating clouds (Tscharnuter 1978, 1987b; Boss 1984a). These models, which included the effects of radiative transfer and detailed equations of state, showed that even for a very slowly rotating cloud, formation of a central protosun is impossible without some means of transporting angular momentum outward and mass inward. Consequently, Tscharnuter (1978, 1987b) has relied on turbulent viscosity to produce the needed angular momentum transport, even during the collapse phase, when there are reasons to doubt the efficacy of turbulence (Safronov 1969).

Because isothermal protostellar clouds, and even slowly rotating non-isothermal clouds (Boss 1986), may fragment prior to ring formation (Bodenheimer *et al.* 1980), a 3D (i.e., fully asymmetric) calculation is necessary in order to ensure that a given collapsing cloud produces a presolar nebula rather than a binary system. Forming the presolar nebula through formation of a triple system followed by orbital decay of that triple system to yield a runaway single nebula and a close binary does not appear to be feasible (Boss 1983), because the runaway single nebula is likely to undergo binary

fragmentation during its own collapse. Hence 3D models of solar nebula formation have also concentrated on low J/M clouds, in a search for clouds that do not undergo rotational fragmentation during their isothermal or nonisothermal collapse phases. Three-dimensional calculations, including radiative transfer in the Eddington approximation, have shown that solar mass clouds with J/M $< 10^{20}$ cm^2 s^{-1} are indeed required in order to avoid binary formation (Boss 1985; Boss 1986).

There are two other ways of suppressing binary fragmentation other than starting with insufficient J/M to form and maintain a binary protostellar system. First, when the initial mass of a collapsing cloud is lowered sufficiently, fragmentation is halted, yielding a lower limit on the mass of protostars formed by the fragmentation of molecular clouds of around 0.01 $M_\odot$ (Boss 1986). The minimum mass arises from the increased importance of thermal pressure as the cloud mass is decreased; thermal pressure resists fragmentation. This limit implies that there may be a gap between the smallest mass protostars ("brown dwarfs") and the most massive planets (the mass of Jupiter is ≈ 0.001 $M_\odot$). Second, clouds that are initially strongly centrally condensed can resist binary fragmentation simply because of their initial geometric prejudice toward forming a single object (Boss 1987). While initially uniform density and initially moderately condensed clouds readily fragment, given large J/M and/or low thermal pressure, it appears to be impossible to fragment a cloud starting from an initial power law density profile (see Figure 1). Considering that the majority of stars are found in binary or multiple systems, it does not appear likely that power law initial density profiles are widespread in regions of star formation, but such a profile could have led to solar nebula formation. Thus, the 3D calculations have shown that formation of the Sun requires the collapse of either a very slowly rotating, high-thermal energy cloud, or else the collapse of a cloud starting from a power law initial density profile. In contrast to 1D calculations, however, no 3D calculation has been able to collapse a cloud core all the way to the pre-main-sequence. In part, this is because of the greatly increased computational effort necessary to evolve a multidimensional cloud through the intermediate phases. Equally important though, is the problem that when rotation is included, accumulation of the central protosun requires an efficient mechanism for outward angular momentum transport. Identifying this mechanism and its effects is one of the major remaining uncertainties in solar nebula models.

ANGULAR MOMENTUM TRANSPORT MECHANISMS

Given the formation of a rotationally flattened, presolar nebula through the collapse of a cloud core that has avoided binary fragmentation, the next major dynamical problem is accumulating the protosun out of the disk
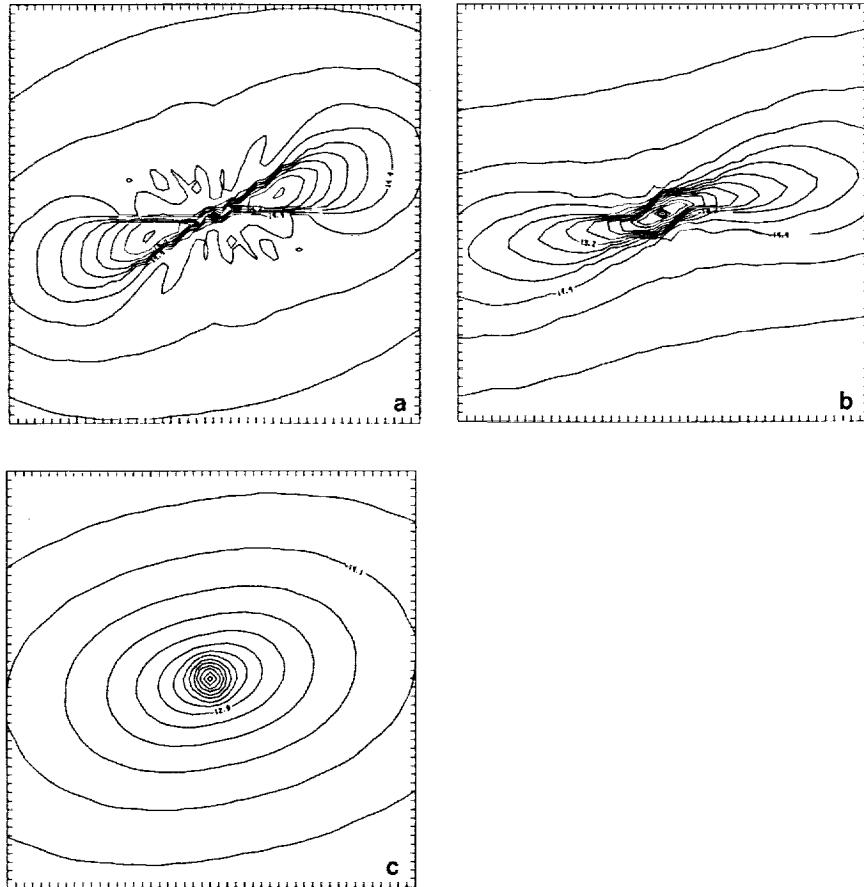
FIGURE 1   Density contours in the midplane of three models of protostellar collapse with varied initial density profiles (Boss 1987). The rotation axis falls in the center of each plot; counterclockwise rotation is assumed. Each contour represents a factor of two change in density; contours are labelled with densities in g cm^{-3}. (a) Initially uniform density profile, (b) initially Gaussian density profile, and (c) initially power law profile (r^{-1}). As the initial degree of central concentration increases, the amount of nonaxisymmetry produced during collapse decreases. Qualititatively similar results hold when the initial cloud mass or initial angular velocity is decreased; binary formation is stifled. Diameter of region shown: (a) 580 AU, (b) 300 AU, (c) 110 AU.

matter. Angular momentum must flow outward, if mass is to accrete onto the protosun, and also if the angular momentum structure of the solar system is to result from a cloud with more or less uniform J/M. The physical process responsible for this dynamical differentiation is thought to have operated within the solar nebula itself, rather than in the material collapsing to form the nebula.

Three different processes have been proposed for transporting mass and angular momentum in the solar nebula: viscous shear, magnetic stresses, and gravitational torques (see also Bodenheimer, this volume). Molecular viscosity is far too small to be important, so turbulent viscosity must be invoked if viscous stresses are to dominate. The most promising means for driving turbulence in the solar nebula appears to be through convective instability in the vertical direction, perpendicular to the nebula midplane (Lin and Papaloizou 1980). The main uncertainty associated with convectively driven viscous evolution, aside from the effective strength of the turbulent stresses, is the possibility that such a nebula is unstable to a diffusive instability that would break up the nebula into a series of concentric rings (Cabot et al. 1987).

As previously mentioned, magnetic fields need not be dominant during the early phases of presolar collapse, and frozen-in magnetic fluxes scale in such a way that they never become important, if they are not important initially. While some meteorites show evidence for remanent magnetic fields requiring solar nebula field strengths on the order of 30μT (Sugiura and Strangway 1988), the magnetic pressure $(B^2/8\pi)$ corresponding to such field strengths is still considerably less than even thermal pressures in hot solar nebula models (Boss 1988), implying the negligibility of magnetic fields for the gross dynamics of the solar nebula.

The remaining candidate for angular momentum transport is gravitational torques between nonaxisymmetric structures in the solar nebula. Possible sources of nonaxisymmetry include intrinsic spiral density waves (Larson 1984), large-scale bars (Boss 1985), and triaxial central protostars (Yuan and Cassen 1985). Early estimates of the efficiency of gravitational torques (Boss 1984b) implied that a moderately nonaxisymmetric nebula can have a time scale for angular momentum transport just as short as a strongly turbulent accretion disk. Three-dimensional calculations of the aborted fission instability in rapidly rotating polytropes (Durisen et al. 1986) were perhaps the first to demonstrate the remarkable ability of gravitational torques to remove orbital angular momentum from quasi-equilibrium, nonaxisymmetric structures similar to the solar nebula.

At later phases of nebula evolution, nonaxisymmetry and spiral density waves can also be driven by massive protoplanets. The possible effects range from gap clearing about the protoplanet, in which case the protoplanet must evolve along with the nebula (Lin and Papaloizou 1986), to rapid orbital

decay of the protoplanet onto the protosun (Ward 1986). While the effects of viscous or magnetic stresses can be studied with 2D (axisymmetric) solar nebula models, in order to model the effects of gravitational torques, a nonaxisymmetric (generally 3D) solar nebula model must be constructed.

THREE-DIMENSIONAL SOLAR NEBULA MODELS

Only a few attempts have been made at studying the nonaxisymmetric structure of the early solar nebula. Cassen *et al.* (1981) used a type of N-body code to study the growth of nonaxisymmetry in an infinitely thin, isothermal model of the solar nebula. Cassen *et al.* (1981) found that when the nebula is relatively cool (~100K) and more massive than the central protosun, nonaxisymmetry grows within a few rotational periods, resulting either in spiral arm formation, or even fragmentation into giant gaseous protoplanets in the particularly extreme case of a nebula 10 times more massive than the protosun. Cassen and Tomley (1988) are presently engaged in using this code to study the onset of gravitational instability in nebula models with simulated thermal gradients.

Boss (1985) used a 3D hydrodynamics code to model the early phases of solar nebula formation through collapse of a dense cloud core, and found that formation of a strong bar-like structure resulted. However, because the explicit nature of the code limited Boss (1985) from evolving the model very far in time, these results are only suggestive of the amount of nonaxisymmetry that could arise in the solar nebula. Recently, Boss (1989) has tried to circumvent this computational problem by calculating a suite of 3D models starting from densities high enough to bypass the intermediate, quasi-equilibrium evolution phases that obstruct explicit codes. While these initial densities for collapse (~10^{-13} – 10^{-12} g cm^{-3}) are clearly not realistic given the present understanding of star formation, it can be argued (Boss 1989) that starting from these high densities should not greatly distort the results.

The 3D models of Boss (1989) show that gravitational torques can be quite efficient at transporting angular momentum in the early solar nebula. The models show that collapsing presolar clouds become appreciably nonaxisymmetric (as a result of a combination of nonlinear coupling with the infall motions, rotational instability, and/or self-gravitation), and that trailing spiral arm patterns often form spontaneously; trailing spiral arms lead to the desired outward transport of angular momentum. The most nonaxisymmetric models tend to be massive nebulae surrounding low mass protosuns in agreement with the results of Cassen *et al.* (1981). Extrapolated time scales for angular momentum transport, and hence nebula evolution, can be as short as ~10^3 years for strongly nonaxisymmetric models, or about ~10^6 – 10^7 years for less nonaxisymmetric models. Because

these time scales are comparable to or less than model ages for naked T Tauri stars (Walter 1988), solar-type, pre-main-sequence stars that show no evidence for circumstellar matter, it appears that gravitational torques can indeed be strong enough to account for the transport of the bulk of nebula gas onto the protosun on the desired time scales. While these initial estimates are encouraging, it remains to be learned exactly how a solar nebula evolves due to gravitational torques.

IMPLICATIONS FOR PLANETARY FORMATION

The 3D solar nebula models of Boss (1989) show little tendency for breaking up directly into small numbers of giant gaseous protoplanets, contrary to one of the models of Cassen *et al.* (1981). This difference is probably a result of several features of the Boss (1989) models. The inclusion of 3D radiative transfer means that the compressional heating accompanying nebula formation can be included, leading to considerably higher temperatures than assumed in Cassen *et al.* (1981), and hence greater stability against break-up. Also, the gradual buildup of the nebula through collapse in the Boss (1989) models means that incipient regions of gravitational instability can be sheared away into trailing spiral arms by the differential rotation of the nebula before the regions become well-defined. These models thus suggest that planet formation must occur through the accumulation of dust grains (Safronov 1969; Wetherill 1980).

Considering that dust grain evolution is not yet included in 3D codes, detailed remarks about the earliest phases of dust grain accumulation are not possible. However, the models of Boss (1989) can be used to predict surface densities of dust grains in the solar nebula, and these surface densities are quite important for theories of planetary accumulation. For example, Goldreich and Ward (1973) suggested that a dust surface density at 1 AU of $\sigma_d \sim 7.5$ g cm^{-2} would be sufficient to result in a gravitational instability of a dust sub-disk (Safronov 1969) that could speed up the intermediate stages of planetary accumulation. More recently, Lissauer (1987) has proposed the rapid formation of Jupiter through runaway accretion of icy-rock planetesimals in a nebula with $\sigma_d > 15$ g cm^{-2} at 5 AU. Rapid formation is required in order to complete giant planet formation prior to dispersal of the solar nebula. Using a gas to dust ratio of 200:1 at 1 AU and 50:1 at 5 AU, these critical surface densities correspond to gas surface densities of 1500 g cm^{-2} at 1 AU and 750 g cm^{-2} at 5 AU. Similar minimum densities are inferred from reconstituting the planets to solar composition (Weidenschilling 1977). The models of Boss (1989) have surface densities in the inner solar nebula that are nearly always sufficient to account for terrestrial planet formation. However, surface densities in the outer solar nebula are less than the critical amount, unless the nebula is quite massive
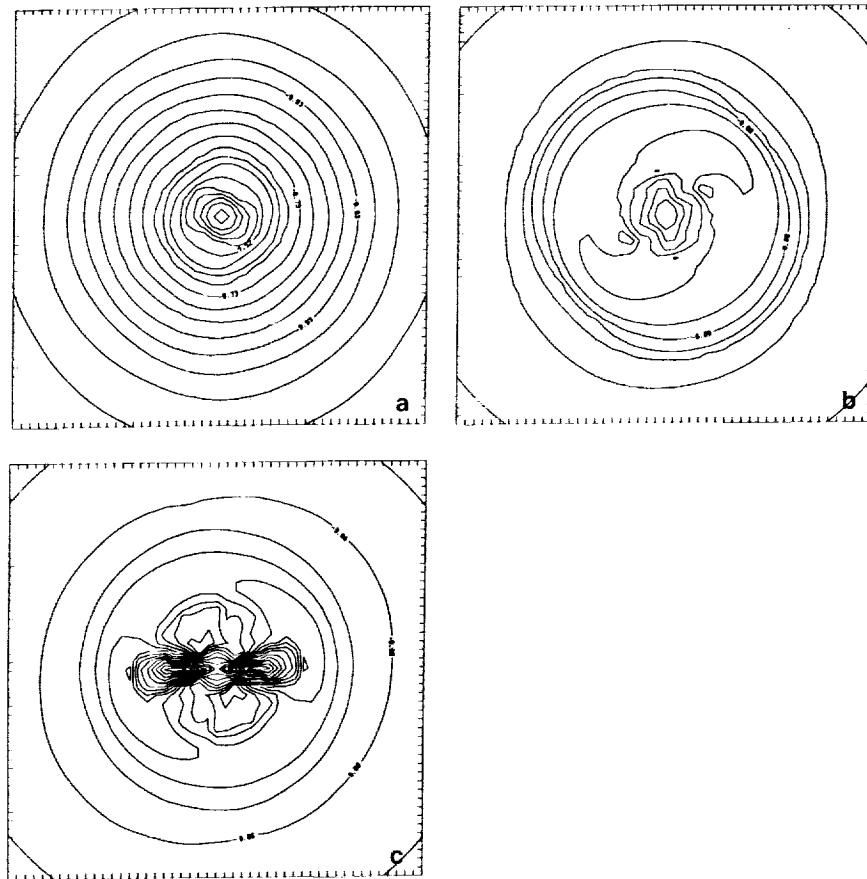
FIGURE 2 Density contours in the midplane of three solar nebula models formed by collapse onto protosuns with varied initial masses M_s (Boss 1989), plotted as in Figure 1. (a) $M_s = 1\ M_\odot$, (b) $M_s = 0.01\ M_\odot$, (c) $M_s = 0M_\odot$. The initial nebula mass was $1M_\odot$ for each model, and the initial specific angular momentum was $J/M = 6.2 \times 10^{19}$ cm^2s^{-1}. The resulting nebulae become increasingly nonaxisymmetric as the initial protosun mass is decreased; (b) forms trailing spiral arms that result in efficient transport of angular momentum, while (c) actually fragments into a transient binary system. These models also illustrate the ability of a massive central object to stabilize a protostellar disk. Region shown is 20 AU across for each model.

($\sim 1 M_\odot$). Any protoJupiter that is formed rapidly in a massive nebula is likely to be lost during subsequent evolution, either because gap clearing will force the protoplanet to be transported onto the protosun with the rest of the gas (Lin and Papaloizou 1986), or because motion of the protoplanet relative to the gas will result in orbital decay onto the protosun (Ward 1986). The planetary system is the debris leftover from formation of the voracious Sun, and so prematurely formed protoplanets are at peril.

Variations in the initial density and angular velocity profiles do not appear to be able to produce sufficiently high surface densities at 5 AU in low-mass solar nebula (Boss 1989), so it does not appear that the surface densities required for planet formation can be accounted for simply by collapse onto a nebula. Diffusive redistribution of water vapor could preferentially accumulate ices wherever temperatures drop to 160K (Stevenson and Lunine 1988), but this mechanism can only be invoked to explain the formation of one of the outer planets. The most promising means for enhancing surface densities of the outer solar nebula appears to be through nebula evolution subsequent to formation. Viscous accretion disks can increase the surface density in the outer regions where the angular momentum is being deposited (Lin and Bodenheimer 1982; Lissauer 1987). While the long-term evolution of a 3D nebula subject to gravitational torques is as yet unknown, gravitational torques should produce a similar result (Lin and Pringle 1987). Determining the evolution of a nonaxisymmetric solar nebula thus appears to be a central issue in finding a solution to the problem of rapid Jupiter formation.

Finally, the 3D models of Boss (1988, 1989) have important implications for the thermal structure of the solar nebula. Provided the artifice of starting from high initial densities does not severely overestimate nebula temperatures, it appears that the compressional energy released by infall into the gravitational well of a solar-mass object can heat the midplane of the inner solar nebula to temperatures on the order of 1500 K for times on the order of 10^5 years. Such temperatures are high enough to vaporize all but the most refractory components of dust grains. In particular, because the vaporization of iron grains around 1420 K removes the dominant source of opacity, temperatures may be regulated to values close to $\sim$1500 K by the thermostatic effect of the opacity. A hot inner solar nebula can account for the gross depletion of volatiles on the terrestrial planets (relative to solar) by allowing the volatiles to be removed along with the H and He of the nebula. Hot solar nebula models were introduced by Cameron (1962) in one of the first solar nebula investigations, but have since fallen into disfavor (Wood 1988), so it will be interesting to see whether high temperatures in the inner nebula can be successfully resurrected, and whether they will prove to be useful in explaining planetary and asteroidal formation.

ACKNOWLEDGMENTS

REFERENCES

Beichman, C.A., P.C. Myers, J.P. Emerson, S. Harris, R. Mathieu, P.J. Benson, and R.E. Jennings. 1986. Candidate solar-type protostars in nearby molecular cloud cores. Astrophys. J. 307:337-349.

Black, D.C., and P. Bodenheimer. 1976. Evolution of rotating interstellar clouds. II. The collapse of protostars of 1, 2, and 5 $M_\odot$. Astrophys. J. 206:138-149.

Bodenheimer, P. 1968. The evolution of protostars of 1 and 12 solar masses. Astrophys. J. 153:483-494.

Bodenheimer, P., J.E. Tohline, and D.C. Black. 1980. Criteria for fragmentation in a collapsing rotating cloud. Astrophys. J. 242:209-218.

Boss, A.P., and J.G. Haber. 1982. Axisymmetric collapse of rotating, interstellar clouds. Astrophys. J. 255:240-244.

Boss, A.P. 1983. Fragmentation of a nonisothermal protostellar cloud. Icarus 55:181-184.

Boss, A.P. 1984a. Protostellar formation in rotating interstellar clouds. IV. Nonisothermal collapse. Astrophys. J. 277:768-782.

Boss, A.P. 1984b. Angular momentum transfer by gravitational torques and the evolution of binary protostars. Mon. Not. R. astr. Soc. 209:543-567.

Boss, A.P. 1985. Three-dimensional calculations of the formation of the presolar nebula from a slowly rotating cloud. Icarus 61:3-9.

Boss, A.P. 1986. Protostellar formation in rotating interstellar clouds. V. Nonisothermal collapse and fragmentation. Astrophys. J. Suppl. 62:519-552.

Boss, A.P. 1987. Protostellar formation in rotating interstellar clouds. VI. Nonuniform initial conditions. Astrophys. J. 319: 149-161.

Boss, A.P. 1988. High temperatures in the early solar nebula. Science 241:505-628.

Boss, A.P. 1989. Evolution of the solar nebula. I. Nonaxisymmetric structure during nebula formation. Astrophys. J., 345:554-571.

Cabot, W., V.M. Canuto, O. Hubickyj, and J.B. Pollack. 1987. The role of turbulent convection in the primitive solar nebula. II. Results. Icarus 69:423-457.

Cameron, A.G.W. 1962. The formation of the Sun and planets. Icarus 1:13-69.

Cassen, P. M., B.F. Smith, R.H. Miller, and R.T. Reynolds. 1981. Numerical experiments on the stability of preplanetary disks. Icarus 48:377-392.

Cassen, P., and L. Tomley. 1988. The dynamical behavior of gravitationally unstable solar nebula models. Bull. Amer. Astron. Soc. 20:815.

Durisen, R.H., R.A. Gingold, J.E. Tohline, and A.P. Boss. 1986. Dynamic fission instabilities in rapidly rotating n = 3/2 polytropes: a comparison of results from finite-difference and smoothed particle hydrodynamics codes. Astrophys. J. 305:281-308.

Goldreich, P., and W.R. Ward. 1973. The formation of planetesimals. Astrophys. J. 183:1051-1061.

Heyer, M.H. 1988. The magnetic evolution of the Taurus molecular clouds. II. A reduced role of the magnetic field in dense core regions. Astrophys. J. 324:311-320.

Larson, R.B. 1969. Numerical calculations of the dynamics of a collapsing protostar. Mon. Not. R. astr. Soc. 145:271-295.

Larson, R.B. 1972. The collapse of a rotating cloud. Mon. Not. R. astr. Soc. 156:437-458.

Larson, R.B. 1984. Gravitational torques and star formation. Mon. Not. R. Astr. Soc. 206:197-207.

Lin, D.N.C., and J. Papaloizou. 1980. On the structure and evolution of the primordial solar nebula. Mon. Not. R. astr. Soc. 191:37-48.

Lin, D.N.C., and P. Bodenheimer. 1982. On the evolution of convective accretion disk models of the primordial solar nebula. Astrophys. J. 262:768-779.

Lin, D.N.C., and J. Papaloizou. 1986. On the tidal interaction between protoplanets and the protoplanetary disk. III. Orbital migration of protoplanets. Astrophys. J. 309: 846-857.

Lin, D.N.C., and J.E. Pringle. 1987. A viscosity prescription for a self-gravitating accretion disk. Mon. Not. R. astr. Soc. 225:607-613.

Lissauer, J.J. 1987. Time scales for planetary accretion and the structure of the protoplanetary disk. Icarus 69:249-265.

Myers, P.C., and P.J. Benson. 1983. Dense cores in dark clouds. II. NH_3 observations and star formation. Astrophys. J. 266: 309-320.

Norman, M.L., J.R. Wilson, and R.T. Barton. 1980. A new calculation on rotating protostar collapse. Astrophys. J. 239:968-981.

Safronov, V.S. 1969. Evolution of the protoplanetary cloud and formation of the Earth and the Planets. Nauka, Moscow.

Stahler, S.W. 1983. The birthline for low-mass stars. Astrophys. J. 274:822-829.

Stevenson, D.J., and J.I. Lunine. 1988. Rapid formation of Jupiter by diffusive redistribution of water vapor in the solar nebula. Icarus 75:146-155.

Sugiura, N., and D.W. Strangway. 1988. Magnetic studies of meteorites. Pages 595-615. In: Kerridge, J.F., and M.S. Matthews (eds.). Meteorites and the Early Solar System. University of Arizona Press, Tucson.

Terebey, S., F.H. Shu, and P. Cassen. 1984. The collapse of the cores of slowly rotating isothermal clouds. Astrophys. J. 286:529-551.

Tscharnuter, W.M. 1978. Collapse of the presolar nebula. Moon Planets 19:229-236.

Tscharnuter, W.M. 1987a. Models of star formation. Page 96. In: Meyer-Hofmeister, E., H.C. Thomas, and W. Hillebrandt (eds.). Physical Processes in Comets, Stars, and Active Galaxies. Springer-Verlag, Berlin.

Tscharnuter, W.M. 1987b. A collapse model of the turbulent presolar nebula. Astron. Astrophys. 188:55-73.

Walter, F.M. 1988. Implications for planetary formation timescales from the nakedness of the low-mass PMS stars. Pages 71-78. In: Weaver, H.A., F. Paresce, and L. Danly (eds.). Formation and Evolution of Planetary Systems. Space Telescope Science Institute, Baltimore.

Ward, W.R. 1986. Density waves in the Solar Nebula: Differential Lindblad Torque. Icarus 67:164-180.

Weidenschilling, S.J. 1977. The distribution of mass in the planetary system and solar nebula. Astrophys. Space Sci. 51:153-158.

Wetherill, G.W. 1980. Formation of the terrestrial planets. Ann. Rev. Astron. Astrophys. 18:77-113.

Wood, J.A. 1988. Chondritic meteorites and the solar nebula. Ann. Rev. Earth Planet. Sci. 16:53-72.

Yuan, C., and P. Cassen. 1985. Protostellar angular momentum transport by spiral density waves. Icarus 64:435-447.

Formation and Evolution of the Protoplanetary Disk

Tamara V. Ruzmaikina and A.B. Makalkin
Schmidt Institute of the Physics of the Earth

ABSTRACT

The structure of the solar system, the similarity of isotope compositions of its bodies (in terms of nonvolatile elements), and observational data on the presence of disks (and possibly of planets) around young solar-type stars are evidence of the joint formation of the Sun and the protoplanetary disk. The removal of angular momentum to the periphery (necessary for the formation of the Sun and protoplanetary disk) is possible at the formation stage in the center of the contracting cloud (protosolar nebula) of the stellar-like core (Sun's embryo). The possibility that the core forms before fragmentation begins to impose a constraint on the value of the angular cloud momentum. This value is highly dependent on the distribution of angular momentum within the cloud.

This paper discusses a disk formation model during collapse of the protosolar nebula with $J \sim 10^{52}$g cm^{2}s^{-1}, yielding a low-mass protoplanetary disk. The disk begins to form at the growth stage of the stellar-like core and expands during accretion to the present dimensions of the solar system. Accretion at the edge of the disk significantly affects the nature of matter fluxes in the disk and its thermal evolution.

In addition to the internal heat source (viscous dissipation), there is an external one which affects the temperature distribution in the disk: radiation (diffused in the accretion envelope) of the shock wave front at the core and in the nearest portion of the disk. Absorbed in the disk's surface layers, this radiation heats these layers, and reduces the vertical temperature gradient in the disk to a subadiabatic point. It also renders convection impossible. Convection becomes possible after accretion ceases.

INTRODUCTION

The proximity of planes of planetary orbit in the solar system indicates that planets were formed in the thin, elongated protoplanetary disk surrounding the young Sun.

The coincidence of isotope compositions of the Sun, the Earth, and meteorites for basic nonvolatile elements, and the similarity of the chemical compositions of the Sun and Jupiter are evidence that the Sun and the protoplanetary disk originated from the same concentration of interstellar medium. In view of observed data on the rapid ($\sim 10^6 \div 10^7$ years) removal of gas from the proximity of solar-type stars which had formed (T Tauri stars), it is natural to infer that the convergence of the isotope composition of the Sun and the planets also means that they formed contemporaneously.

According to the theory developed by Jeans early in this century, stars are formed as a result of collapse under the impact of the gravitational effects of compacted regions of the interstellar medium. The collapse occurs when the forces of self-gravity exceed the sum of forces restricting compression. The latter include the thermal pressure gradient, magnetic pull, and centrifugal forces. Data from infrared and radioastronomy tell us that the stars are formed in molecular clouds: low temperature regions ($\simeq 10$ K) with relatively high density $10^{-22} \div 10^{-20}$ g $\cdot$ cm^{-3}, in which hydrogen and other gases (besides the noble gases) are in a molecular state, and condensing matter is included in grains.

Study of the radio lines of molecules has shown that the clouds are highly inhomogeneous. They contain compact areas, cores with densities $\rho \sim 10^{-20} \div 10^{-18}$ g $\cdot$ cm^{-3} and masses M $\sim 0.1 \div 10$ M$_\odot$, infrared sources and compact zones of ionized hydrogen with ages $10^4 \div 10^5$ years. Relatively weak, variable T Tauri stars are also seen in certain molecular clouds. They approximate the Sun in terms of mass ($0.5 \div 2$M$_\odot$), but are much younger, with ages $10^5 \div 10^6$ years (Adams *et al.* 1983). Their formation is related to the compression of cores of molecular clouds under the impact of the forces of self-gravity.

Contemporary theory of evolution holds that the fate of a star is determined by its mass and chemical composition. In view of the similarity of these parameters, we can identify T Tauri stars with the young Sun and use observed data on these stars to construct a theory of the formation of the Sun and the protoplanetary disk.

CIRCUMSTELLAR DISKS

Elongated, disk-shaped, gas-dust envelopes with characteristic masses 0.1M$_\odot$ (Sargent and Beckwith 1987; Smith and Terrile 1984) have been discovered around several young stars (HL Tau, DG Tau, β Pic, and a

source of infrared radiation L 1551 IRS5). The radii of the gas-dust disks, which were determined by eclipses of the stars, are estimated at ~ 10 ÷ 10^2 AU (see Strom's article in this volume). The radii of the disk-shaped envelopes in CO molecule lines are estimated on the order of 10^3 AU. Data from observations of the IRAS infrared astronomical satellite have shown that 18% of 150 near stars which have been studied exhibit infrared excesses. That is, they emit more in the infrared range than matches their temperatures (Backman 1987). It was proven for a number of closer stars (Aumann *et al.* 1984) that excess infrared radiation is not created by the star itself, but by the envelope of dust grains. The grains absorb the light of the star and reemit it in the infrared range. It is possible that disks exist for all stars which exhibit infrared excesses (Bertrout *et al.* 1988). Thus, there is a greater probability that the formation of a star is accompanied by the formation of a gas-dust disk around it.

IRAS discovered in the envelopes of stars α Lyr, α PsA, and β Pic that the central region with a radius of 30 AU is dust-free (Backman 1987). This empty region could not have been retained after the star's formation stage, since the dust grains from the surrounding envelope shift inside under the Poynting-Robertson effect and fill up the empty space over a time scale of < 10^5 years. This is too brief a time scale in comparison with the age of the star. A possible cause of the existence of an empty region around stars is that the large planetesimals and planets may remove those dust grains shifting towards the center.

The first finding of the search for planets around solar-type stars was obtained using an indirect search method: determining with high accuracy (up to 10 to 13 meters per second) the ray velocities of the nearby stars. Seven of 16 stars examined were found to have long-period Doppler Shifts of velocities with amplitudes of 25-45 meters per second. It is probable that these variations in ray velocities are produced by invisible components (planets with masses from one to nine Jovian masses) moving in orbit around the stars (Campbell *et al.* 1988).

These findings, together with observational data on the disks around young stars, are evidence that the formation of the planetary system is a natural process which is related to star formation.

CONDITIONS FOR THE FORMATION OF STARS WITH DISKS

The most likely cause of the formation of protoplanetary disks and planetary systems, including the solar system, is the rotation of molecular clouds. The rotation of clouds and separate dense regions in them is a function of the differential rotation of the galaxy and turbulence in the interstellar medium. The rotation velocities of molecular clouds and their nuclei are determined by the value of the gradient of the spectral line's ray

velocity along the cloud profile. Angular velocities of rotation Ω, measured in this manner, are included in the interval $10^{-15}-10^{-13}\text{s}^{-1}$ (Myers and Benson 1983). For small cores of $< 0.1\text{pc}$, rotation is only measurable for $\Omega > 2.10^{-14}\text{s}^{-1}$. Rotation was not discovered with this kind of accuracy for approximately 30% of the cores.

Angular momentum value and internal distribution $J(r)$ are important cloud/core characteristics for star formation. These characteristics depend both on the value Ω on the outer edges of the cloud and its internal distribution. Data on the dependence $\Omega(r)$ have been measured for individual, sufficiently elongated molecular clouds. Thus, for cloud B 361, $\Omega \simeq$ const in the inner region and falls on the periphery in the outer region (Arquilla 1984). A concentration of matter towards the center is also observed within the compact cores. The distribution Ω is not known. The maximum laws of rotation which appear to be reasonable are being explored in theoretical studies:

$$\Omega = \text{const with } \rho = \text{const}, \tag{1a}$$

$$\Omega = \text{const with } \rho \propto r^{-2}. \tag{1b}$$

The first describes the angular momentum distribution in a homogeneous and solid-state rotating cloud. It holds for a core which has separated from the homogeneous rotating medium where the specific angular momentum of each cloud element is conserved. The second corresponds to a solid-state rotating, singular isothermal sphere. This is an isothermal cloud with $\rho = C_s^2/(2\pi G r^2)$, in which the forces of self-gravity are balanced out by internal pressure (C_s is the speed of sound). In order for this kind of distribution to be established, the core must exist long enough during the stage preceding collapse for angular momentum redistribution to occur and for solid-state rotation to be established. The characteristic time scale for the existence of cores prior to the onset of contraction is $\tau_{sf} \sim \sim 10^7$ years. This is significantly more than the contraction of an individual core of $\tau_{ff} \sim \sim 10^5$ years (Adams *et al.* 1983).

Random (turbulent) motion with near-sound speeds is present in the cores (Myers 1983). The viscosity created by this motion may be represented as (Schakura and Sunyaev 1973)

$$\nu_T = 1/3 v_T l_T \simeq \propto R c_s, \tag{2}$$

where R denotes the core radius, sc is the speed of sound in it, $\propto$ is the nondimensional value, and v_T and l_T are the characteristic turbulent motion velocity and scale. The time scale for angular momentum redistribution in the core under the effect of viscosity $\tau_\nu \sim R^2/\nu_T$. The condition $\tau_\nu \sim \tau_{sf}$ is fulfilled for

$$\propto \, \sim R/c_s \tau_{sf} \, . \tag{3}$$

As a numerical example, let us consider the core TMC-2 . It has a mass $M \simeq M_\odot$, $c_s = 3 \cdot 10^4$cm/s and $v_T \simeq 0.5 \, c_s$, $R \simeq 0.1$ pc (Myers 1983). Substitution in (2) and (3) of these numerical values and $\tau_{sf} = 3 \cdot 10^{14}$s give us $\propto \, \sim 0.03$. This fits with $l_T/R \sim 0.1$. This estimate demonstrates that the efficiency of angular momentum redistribution in various cores may differ depending on the scale of turbulent motion. Therefore, cores with both rotation laws, (see 1a and 1b) and the intermediate ones between them, may exist.

The angular momentum of a solid-state rotating spherical core of mass M is equal to

$$J \simeq 10^{52} K \frac{\Omega}{10^{-15} s^{-1}} \left(\frac{M}{M_\odot} \right)^{5/3} \left(\frac{\bar{p}}{10^{-19} g cm^{-3}} \right)^{-2/3} g cm^2 s^{-1}, \tag{4}$$

where $k = 2/5$ and $2/9$ for the rotation laws (1a) and (1b), and $\bar{p}$ denotes the mean core density.

The rotation of clouds plays an important role in star formation. Stellar statistics demonstrate that more than one half $\gtrsim 60\%$) of solar-type stars enter into binary or multiple systems which, as a rule, exhibit angular momentums $> 10^{52}$ g·cm^2·s^{-1} (Kraycheva et al. 1978). This means that when a binary (multiple) system is formed, the bulk of a cloud's angular momentum is concentrated in the orbital movement of stars relative to each other. The formation of a single star with a disk is an alternative and additional route by which a forming star expels excess angular momentum.

In a circumsolar Kepler disk, the angular momentum per unit mass is, actually, $j = (GM_\odot R)^{1/2}$. At a Jovian distance (5 AU) $j \sim 10^{20}$cm^2 s^{-1}. This is 100 times more than the maximum possible and 10^5 times more than the present angular momentum related to the Sun's rotation. Therefore, even a low-mass but elongated disk can accumulate a large portion of cloud angular momentum, thereby allowing a single star to form.

Estimates of mass M_D and angular momentum J_D of a circumstellar protoplanetary disk, performed by adding presolar composition dissipated hydrogen and helium to the planet matter, yielded (Weidenschilling 1977),

$$10^{-2} M_\odot \lesssim M_D \lesssim 10^{-1} M_\odot, \text{and } 3 \cdot 10^{51} \lesssim J_D \lesssim 2 \cdot 10^{52} g \cdot cm^2 \cdot s^{-1},$$

that is,

$$J_D/M_D \gg J_\odot/M_\odot.$$

Using a system of two bodies with constant aggregate mass and full angular momentum as an example, Lynden-Bell and Pringle (1974) demonstrated that the system's total energy decreases as mass is transferred from the smaller to the larger body, and the concentrations of angular momentum in orbital motion are less than the massive body. Consequently, dissipation of rotational energy, accompanied by removal of the angular momentum to the periphery and its concentration in a low amount of mass, is necessary to form a single star with a protoplanetary disk as a cloud contracts.

Effectiveness of angular momentum redistribution is the key issue of protoplanetary disk formation. The high abundance of binary or multiple stars of comparable masses and the analysis of the contraction dynamics of rotating clouds indicate that angular momentum redistribution is not always effective enough for a single star with a disk to be formed. The entire process, from the beginning contraction all the way to the formation of a star, has thus far only been examined for a nonrotating cloud (Larson 1969; Stahler et al. 1980). Calculations have shown that the initial stage of contraction occurs at free-fall velocities and is accompanied by an increase in the concentration of matter towards the center. Pressure increase, coupled with a rise in temperature, triggers a temporary deceleration in the contraction of a cloud's central region within a density range of $\rho_c \sim 10^{-13} \div 10^{-8}$ g·cm^{-3}. This is followed by one more stage of dynamic contraction that is initiated by molecular hydrogen dissociation. Dissociation terminates at 10^{-2}g cm^{-3}; the contraction process again comes to a halt; and a quasihydrostatic stellar-like core is formed with an initial mass of $M_c \sim 10^{-2}M_\odot$ and central density $\rho_c \sim 10^{-2}$g·cm^{-3}. This core is surrounded by an envelope which initially contains 99% mass and falls onto the core over a time scale of $10^5 \div 10^6$ years. Naturally, a single nonrotating diskless star is formed from the contraction of this kind of protostellar cloud.

It is clear from these general ideas that contraction of a rotating protostellar cloud occurs in a similar manner, when the centrifugal force in a cloud is low throughout in comparison to the gravitational force and internal pressure gradient.

The role of rotation is enhanced with increased density in the contraction process where angular momentum is conserved. (For example, the ratio of rotational energy to gravitational energy is $\beta \propto \rho^{1/3}$ for a spherically symmetrical collapse). Two-dimensional and three-dimensional calculations for the contraction of rapidly rotating protostellar clouds have shown that as a certain β_{cr} is reached in the cloud central region, a ring (two-dimensional) or nonaxisymmetrical (three-dimensional) instability develops. According to Bodenhemier (1981) and Boss (1987) this instability triggers fragmentation: $\beta_{cr} \simeq 0.08$ for ring instability at the hydrodynamic

contraction stage (Boss 1984). Naturally, the smaller the angular momentum of the cloud's central region, the later the instability arises. It has been suggested that a cloud's fate depends significantly on the stage of contraction at which instability arises: the cloud turns into a binary or multiple star system when β_{cr} is attained at the initial hydrodynamic stage of contraction (with $\rho_c \lesssim 10^{-12}$ g·cm^{-3}). It becomes a single star with a disk when nontransparency increases, causing the contraction of the portion to decelerate before instability develops. The condition that a nontransparent core be formed before fragmentation occurs, imposes a constraint on the cloud angular momentum: $J \lesssim 10^{53}$ g cm^2·s^{-1} for rotation law (1a) (Safronov and Ruzmaikina 1978; Boss 1985).

Fragmentation may be halted by a sufficiently efficient removal of angular momentum from the center. Turbulence or the magnetic field (Safronov and Ruzmaikina 1978) have been proposed as removal mechanisms. Another is gravitational friction (Boss 1984) generated from excitation by a central nonaxisymmetrical condensation of the density wave in the envelope surrounding the core. Yet to be determined is whether angular momentum removal at this stage can actually prevent fragmentation. The difficulty is that contraction deceleration due to enhanced nontransparency is temporary. It is followed by the stage of hydrodynamic contraction triggered by molecular hydrogen dissociation. During this stage, density increases by several orders and β may attain β_{cr}. The formation of a single star with a disk (or without it) appears to be highly probable when the cloud (or its central portion, to be more exact) exhibit such a slight angular momentum that instability does not develop until the formation of a low-mass, stellar-like core with $M_c < 10^{-2} M_\odot$ (Ruzmaikina 1980, 1981). When it is born, the core must be magnetized as a result of enhancement of the interstellar magnetic field during contraction (Ruzmaikina 1980, 1985). The poloidal magnetic field strength in the stellar-like core is estimated at 10 $\div$ 10^3 Gauss. This field ensures angular momentum redistribution in the core over a time scale less than its evolution time scale ($\gtrsim 10^2$ years) and initiates an outflow of the core's matter, forming an embryonic disk instead of fragmentation.

With disruption of the core's axial symmetry, angular momentum removal to the periphery may also be carried out by the spiral density wave generated in the envelope (Yuan and Cassen 1985). Consequently, the formation of a single, stellar-like core appears to be sufficient for the formation of a single star.

The possibility of the formation of a stellar-like core with $M_c \sim 10^{-2} M_\odot$ imposes constraints on the maximum value of the protostellar cloud's angular momentum. With an angular momentum distribution within the cloud as described in ratio (1a) (solid-state rotation with homogeneous

density), the maximum angular momentum value of the cloud at which a single, stellar-like core can form is within the range (Ruzmaikina 1981)

$$0.3 \ 10^{52}(M/M_\odot)^{5/3} \lesssim J^a_{max} < 2 \ 10^{52}(M/M_\odot)^{5/3}. \tag{5}$$

With the law of rotation contained in (1b) (solid-state rotation with $\rho \propto r^{-2}$), J_{max} is approximately $(M_\odot/M_c)^{4/3} \simeq 500$ times greater than with (1a), that is, $J^b_{max} \gtrsim \gtrsim 10^{54}g \ cm^2 \cdot s^{-1}$ with $M = 1M_\odot$. This is easy to determine by equating the angular momentums of the central sphere with mass $M_c \sim 10^{-2}M_\odot$, for distributions (1a) and (1b), respectively.

It follows from the above estimates that: (1) single stellar-like cores can be formed during the contraction of clouds whose angular momenta are included in a broad range. This range overlaps to a large degree the angular momentums of cores in molecular clouds. Therefore, there may be a considerably high probability that a single, stellar-like core can be formed in the contracting dense region (core) of a molecular cloud. This depends on the angular momentum distribution established inside the core of molecular clouds at the pre-contraction stage; (2) J^a_{max} approximates or exceeds by several times the angular momentum of a "minimal mass" solar nebula (Weidenschilling 1977). J^b_{max} is greater than or on the order of the angular momentum of a massive solar nebula (Cameron 1962). Let us note that the idea of a massive solar nebula was recently revived by Marochnik and Mukhin (1988), who reviewed the estimate of the Oort cloud's mass on the hypothesis that the typical mass of cometary bodies in the cloud is equal to the mass of Halley's comet. (Data gathered by the Vega missions have put estimates of its mass at two orders greater than was previously believed.)

For a broad range of J values, scenarios appear possible whereby a single Sun embryo is first formed in the contracting protosolar nebula. A disk then forms around it. Disk parameters and the nature of its evolution are dependent on angular momentum value. However, the presence of a single, stellar-like core in the center has a stabilizing effect on the disk's central portion and may prevent its fragmentation. We will later discuss in more detail a protoplanetary disk formation model with cloud contraction of $J \sim 10^{52}g \ cm^2 \ s^{-1}$ which appears preferable for a solar nebula (Ruzmaikina 1980, 1982; Ruzmaikina and Maeva 1986; see also review papers Safronov and Ruzmaikina 1985; Ruzmaikina *et al.* 1989).

EARLY EVOLUTION OF THE PROTOPLANETARY DISK

Let us consider the stage of protosolar nebula contraction when a single, stellar-like core and a compact embryonic disk are formed in the center, both surrounded by an accretive shell. The embryonic disk could

have been formed from external equatorial core layers under the impact of magnetic pull (Ruzmaikina 1980, 1985) or by direct accretion of the rotating envelope at a distance from the axis greater than the equatorial radius of the core (Tereby et al. 1984). The directions of flow of the accretive material intersect the equatorial plane inside the so-called centrifugal radius R_k which exceeds (at least at the final stage of accretion) the radius of the protosun $R_c = 3 \div 5\ R_\odot$

$$R_K = \frac{J^2}{K^2 GM^3}.\qquad\qquad(6)$$

$R_K = 0.15 \div 0.5$ AU where $M = 1\ M_\odot$, $J = 10^{52} \mathrm{g \cdot cm^2 \cdot s^{-1}}$, and $K = 2/5 \div 2/9$, respectively.

A mechanism for the coformation of the Sun and the protoplanetary disk during protosolar nebula contraction with $J \sim 10^{52}\mathrm{g\ cm^2 \cdot s^{-1}}$ has been proposed in studies by Ruzmaikina (1980) and Cassen and Moosman (1981) and investigated by Ruzmaikina (1982), Cassen and Summers (1983), and Ruzmaikina and Maeva (1986). The disk proposed in these models is a turbulent one. The following points have been offered to justify this: large Reynolds number ($\gtrsim 10^{10}$) for currents between the disk and the accretive envelope and currents generated by differential disk rotation; and the possible development of vertical (Z) direction convection. Investigations have shown that a sufficiently weak turbulence with $\propto \sim 10^{-2}$ can trigger an increase in the disk radius to the current size of the solar system within the time scale of the Sun's formation (10^5 years). Approximately 1% of the kinetic energy of the accretive material is needed to support this kind of turbulence. Near-sound turbulence produces disk growth to 10^3 AU.

It is noteworthy that the process of disk growth occurs inside the protosolar nebula as it continues to contract. Nebula matter (gas and dust) accrete on the forming Sun and the disk (Figure 1). Matter situated in the envelope and close to the equatorial plane encounters the face of the disk. Turbulence causes accreting matter flowing about the disk to mix with disk matter. Addition of the new matter is especially effective on the face, where this matter falls on the disk almost perpendicular to the surface. It loses its radial velocity in the shock wave and is retarded long enough for effective mixing. Complete mixing on the remaining surface of the disk only occurs in layer Δh. This layer is small at subsonic turbulence in comparison with disk thickness $\Delta h/h \sim \propto^{1/2}$ (Ruzmaikina and Safronov 1985). Ruzmaikina and Maeva (1986) looked at the process of protoplanetary disk formation for a model with $J = 2 \cdot 10^{52}\mathrm{g\ cm^2 \cdot s^{-1}}$ and $M = 1.1 M_\odot$, taking into account accretion of material both to the Sun and to the disk. Turbulence viscosity in the disk was alleged to equal $v_T = (1 \div 6) \cdot 10^{15}$ (M/M$_\odot \cdot$ R AU)$^{1/2}$ cm^{2}s^{-1}. This fits with $\propto \approx 3 \cdot 10^{-3} \div 4 \cdot 10^{-2}$. As a result, by the completion of the accretion stage (which lasts 10^5 years), the disk radius
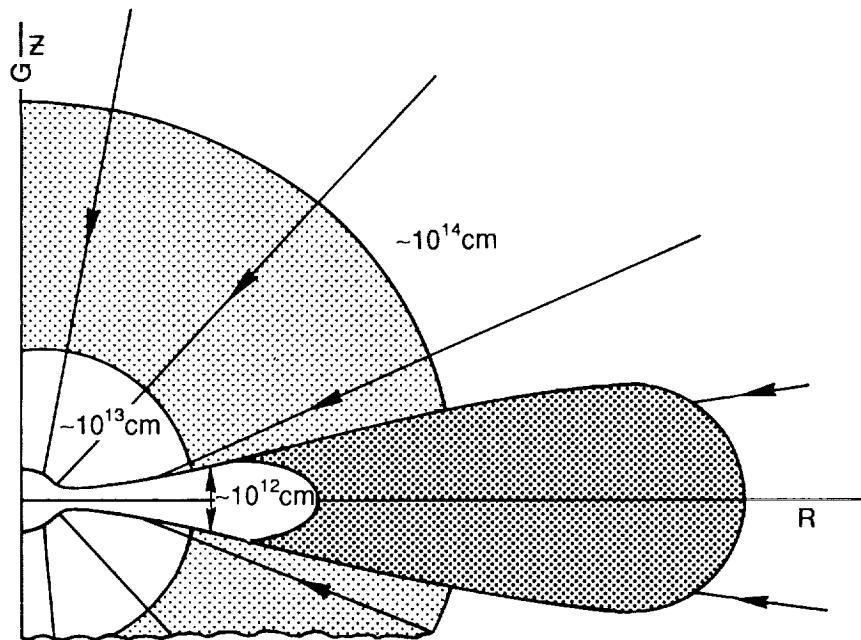
FIGURE 1 Schematic profile of a forming protoplanetary disk immersed in an accretion envelope.

was equal to 24 - 70 AU, its mass is about 0.1 $M_\odot$, and the remaining mass is concentrated in the Sun. The distribution of radial U_R velocity of matter flow in the disk undergoes a complex evolution: at the initial stage, matter in the larger portion of the disk flows towards the center. The velocity is only positive near the disk edge. However, there gradually emerges one more area with a positive radial velocity which is broadening over time. Two regions with $U_R < 0$; $R < R_i \simeq 0.3 \div 0.5$ AU and $0.6\ R_D < R < 0.95\ R_D$ two regions with $U_R > 0$; $R_i < R < 0.6\ R_D$ and $R > 0.95\ R_D$ exist in the disk by the time accretion is completed.

TEMPERATURE CONDITIONS AND CONVECTION IN THE PROTOPLANETARY DISK

The question of temperature distribution and fluctuation in the disk is important for an understanding of the physical and chemical evolution of preplanetary matter. Temperature greatly affects the kinetics of chemical reactions, matter condensation, and vaporization, the efficiency rate at which dust grains combine during collision, and the conditions within planetesimals.

The protoplanetary disk is thin, that is, at any distance R from the center, the inequality $h/R \ll 1$ is true for the thickness of a disk's homogeneous atmosphere of $h = c_s \Omega^{-1}$. Therefore, heat transfer occurs primarily transversely to the disk in the l-c-z direction, between the central plane and the surface. At the same time, a disk of mass $10^{-2}- 10^{-1} M_\odot$ and radius 10 — 100 AU is optically thick (Lin and Papaloizou 1980). Therefore, if there is an internal source heating the disk, the temperature in the central plane is higher than on the surface. Such a source is the internal friction in conditions of differential rotation. Mechanical energy dissipation in the disk is proportional to $\nu_T R^2 (d\Omega/dR)^2$. If we propose turbulence as a viscosity mechanism, the viscosity value averaged for disk thickness could be written using Schakura and Sunyaev's α-parameter as $\nu_T = \alpha\, c_s h$ where the speed of sound is taken from the central plane. As we noted above, the value α must equal 10^{-2} for a disk to form over 10^5 years. A number of models were constructed using Lynden-Bell and Pringle's viscous disk evolution theory (1974). These models considered the further evolution of the protoplanetary disk after protosolar nebula matter has stopped precipitating on it. A large portion of disk mass is transported inside and accretes to the Sun at this stage. At the same time the disk radius increases owing to conservation of the angular momentum. According to the estimates, surface density decreases by one order over 10^6 years with $\alpha = 10^{-2}$. (Ruden and Lin 1986; Makalkin and Dorofeeva 1989).

Internal disk structure and, in particular, the vertical temperature profile were also considered for this stage (viscous disk diffusion). Lin and Papaloizou's model (1980) proposes Z-direction convection as a turbulence mechanism in the disk. Correspondingly, the temperature gradient in this direction is slightly higher than adiabatic. P-T conditions and matter condensation in the protoplanetary disk's internal portion (Cameron and Fegley 1982) were calculated on the basis of this model. Cabot et al. (1987) yielded a more accurate vertical structure which accounts for the dependence of opacity on temperature: beginning with the central plane, the convective layer with a superadiabatic temperature gradient is superseded at a higher elevation by a layer in which the gradient is below adiabatic. One more layer alternation may occur above this if the photosphere temperature is below that of ice condensation. However, on the average, the temperature profile is quite close to adiabatic for the entire thickness of the disk. Temperature distributions in these models are in fairly close agreement with the temperature estimate generated by Lewis (1974) where he used cosmochemical data. His estimate is indicated by the crosses in Figure 2.

Convective models only take into account the internal source of protoplanetary disk heating: turbulence dissipation. However, external sources may provide an appreciable input to disk heating. They are particularly
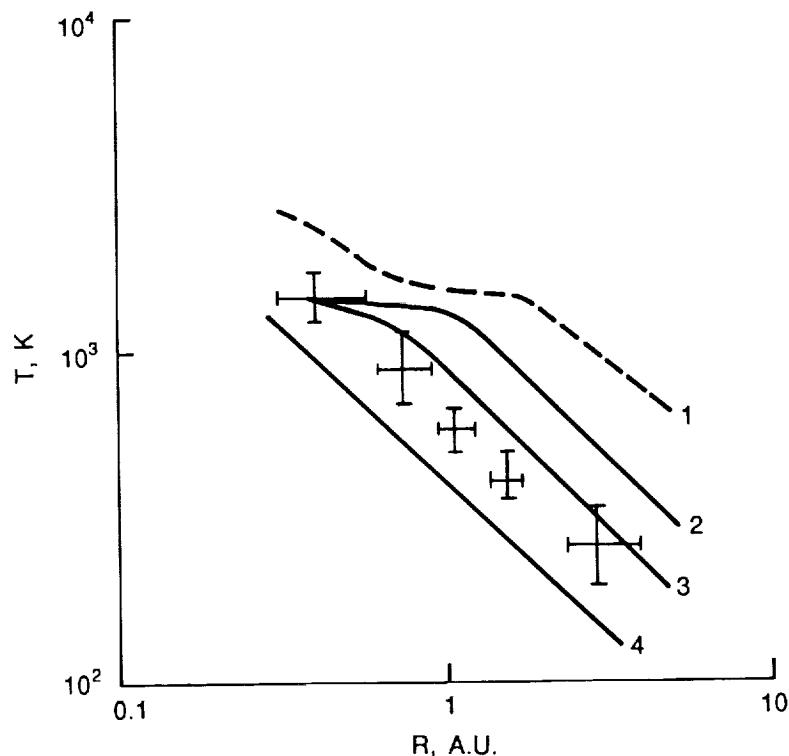
FIGURE 2   Temperature distribution in the central plane of the protoplanetary disk: 1 is the maximum temperatures at the disk formation stage. The figures 2, 3, and 4 are at the subsequent stage of viscous disk evolution (from the study by Makalkina and Dorofeeva 1989): 2 is $1 \cdot 10^5$ years after the stage begins; 3 is $2 \cdot 10^5$ years later; and 4 is after 10^6 years. Ranges of temperature ambiguities as per Lewis' cosmochemical model (1974) are indicated by the overlapping areas.

significant during disk formation. At the accretion stage of contraction of a protosolar nebula with $J \sim 10^{52} g \cdot cm^2 \cdot s^{-1}$, energy is emitted at the shock front of the protosolar core's surface and the portion of the disk with $R_k \sim 10^{12}$ centimeters nearest to it (which is only several times greater than the core radius and two to three orders less than the radius of disk R_D by the end of the accretion stage). This radiation is absorbed and reemitted in the infrared range in the accretion envelope around the core. A significant portion of the disk is immersed in the envelope's optically thick portion and is appreciably heated by its radiation (Makalkin 1987). With a characteristic accretion time scale $\tau_a \sim 10^5$ years, the radius of the optically thick

portion of the accretion envelope (the radius of the dust photosphere) is approximately 10^{14} centimeters.

This estimate was made for a spherically symmetric collapse model (Stahler *et al.* 1981). However, it is just in terms of the order of value and for a model with a moderate angular momentum. The estimate of the internal radius of the optically thick portion of the envelope (the radius of the dust vaporization front), which is approximately 10^{13} centimeters (Figure 1), is also true. Where there is a disk around the core, a very rough estimate of luminosity $L = GM_c\dot{M}/R_k$, and a more stringent estimate based on the theory of Adams and Shu (1986), demonstrate that energy ensuring a luminosity value $L \simeq 20 \div 25L_\odot$ is emitted at the shock front near the core ($R \lesssim R_k$) with $\tau_a \approx 10^5$ years and $J \approx 10^{52} g \cdot cm^2 \cdot s^{-1}$. That is, this value is several times less than in Stahler *et al.*'s spherically symmetrical model (1980) where $L = 66L_\odot$. An estimate of the temperature in the envelope using Adams and Shu's method (1985) produces for $L \simeq 20 \div 25L_\odot$ a temperature in the envelope of $T_{env} \simeq 1100$ K at $R = 1$ AU. We note for comparison that at the same R $T_{env} \simeq 1600$ K with $L = 70$ $L_\odot$ and $T_{env} = 800$ K at $L = 15L_\odot$ (for the spherical model of Adams and Shu (1985) with $\tau_a \simeq 10^6$ years). The exact form of the shock front surface and the dependence of the radiation flow on angular coordinates are still unknown. It is considered in a zero approximation that at $R \gg R_k$, isotherms have a near spherical form.

The specific dissipation energy of turbulent motion in the protoplanetary disk can be expressed as $D = 9/4\ \rho\nu_T\Omega^2$. It follows from this that the flow of radiant energy from the disk (per unit of area of each of the two surfaces in its quasistationary mode) is equal to $D_1 = 9/8\Sigma\ \nu_T\ \Omega^2 = \sigma T^4_{eff}$ where Σ denotes the disk's surface density, and σ is the Stefan-Boltsman constant. It is easy to see that at $\propto \sim 10^{-2}$ the effective disk temperature T_{eff} is significantly lower than the temperature in the accretion envelope T_{env} at the same R. Therefore, the effect of outflow from the disk to heat the envelope at $R \gg R_k$ can be disregarded. Even where a developed convection is present, the flow of radiant energy inside the protoplanetary disk F is approximately three times greater than the convective value (Lin and Papaloizou 1985). This is similar to what occurs in hot accretion disks. The solution to the formula for radiation transport in the disk $dF/dz = D$, where the mean opacity is dependent on temperature according to the power law, $\overline{Ke} \propto T^\xi$, is expressed as (Makalkin 1987)

$$(T_o/T_s)^{4-\xi} = 1 + 3/64\ (4 - \xi)(D_1/\sigma T^4_s)\overline{Ke}_s\Sigma, \tag{7}$$

where T_o and T_s denote temperatures in the disk's central plane and on its surface, $\overline{Ke}_s = \overline{Ke}(T_s)$. In the absence of external heating sources, $T_s = T_{eff}$ is fulfilled and, correspondingly, $D_1 = \sigma T^4_s$. In the case of

a disk immersed in the accretion envelope, T_s in formula (7) is equal to the temperature in the envelope T_{env} at the same R. T_{env} is several times greater than T_{eff}, and hence, $D_1 \ll \sigma T^4_s$. The mean opacity $\overline{Ke}$, examined by Pollack et al. (1985) for a protoplanetary disk, taking into account the chemical composition and dimensions of dust grains, can be approximated as a function T^ξ with various values of ξ in different temperature intervals. Below the temperature of dust condensation, the values of ξ range from 0.6 – 1.5. Above this temperature Lin and Papaloizou's approximation (1985) can be used for $\overline{Ke}(T)$. The ratio (7) can be applied not only to the entire thickness of the disk, but also to its different layers in order to account for variations in the function $\overline{Ke}(T)$. Then Σ and D_1 correspond to the appropriate layer, and T_o and T_s to its lower and upper boundaries. We note that formula (7) differs significantly both in its appearance and its result from the widely accepted, simplified formula $(T_o/T_s)^4 = 3/8 \ \overline{Ke}_o\Sigma$, where $\overline{Ke}_o = \overline{Ke}(T_o)$.

Curve 1 in Figure 2 illustrates the temperature in the disk's central plane at the end of its formation stage. We calculate this temperature using formula (7) for the disk model, (Ruzmaikina and Maeva 1986) with values $= 110^4 \text{gcm}^{-2}$ and $\nu_T = 1.210^{15}\text{cm}^2\text{s}^{-1}$ at R = 1 AU. This accounts for the fact that $\Sigma \ \nu_T \propto R^{-1/2}$ is everywhere except the disk center and edge. The temperature in the accretion envelope is estimated for L = 25 $L_\odot$. The plateau in curve 1 fits that part of the disk where its vertical structure is two-layered. A dust-free (owing to its high temperature) layer is located around the central plane. The mean opacity here is two orders lower (and the vertical temperature gradient is commensurately low), than in the higher, colder layer. This colder layer contains condensed dust grains. The vertical profile of this portion of the disk is represented in Figure 1. It is clear from Figure 1 that the condensation front has a curved shape. Computations have shown that for all $R \lesssim 10^{14}$ centimeters the inequality $T_o < 2T_s$ is fulfilled. It follows, in particular from this, that the vertical temperature gradient at the disk formation stage is noticeably lower than the adiabatic value calculated in many studies. There is, therefore no convection.

Vityazev and Pechernikova (1985) and Ruzmaikina and Safronov (1985) estimated the maximum for additional heating of matter as it falls onto the disk surface; in the shock wave (where it had been) and via aerodynamic friction of the dust. The effect is relatively insignificant in the area $R \gg R_k$, $R \lesssim 10^{14}$ centimeters. Thus, the dust temperature does not exceed 600 K for 1 AU with $J \sim 10^{52} \text{ g} \cdot \text{cm}^2 \cdot \text{s}^{-1}$. That is, it is much lower than the temperature inside the disk (Curve 1, Figure 2).

Solar radiation falling at a low inclination on its surface (Safronov 1969) is an external source of disk heating after accretion of the envelope to the protosun and disk is completed. It should be taken into account

together with turbulence (an internal source) when calculating temperature in the disk. T_s in (7) is determined at this stage by the formula $\sigma T_s^4 = D_1 + F_s$, where F_s denotes the solar energy flux absorbed by the disk surface. It follows from the computations of Makalkin and Dorofeyeva (1989) that $D_1 \gg F_s$ everywhere for $R \gtrsim 0.1$ AU. The vertical temperature gradient is only higher than the adiabatic where dust grains are not vaporized. Where there is no dust, low opacity $\overline{Ke} \lesssim 10^{-2} \mathrm{cm}^2 \cdot \mathrm{g}^{-1}$ generates a very low-temperature gradient. The temperature distribution is then close to isothermic. With this in mind, the authors calculated a combined (adiabat - isotherm) thermodynamic model of a protoplanetary disk. Figure 2 (Curves 2, 3, and 4) illustrate the temperature distributions in the central plane generated in this model for different points in time. Temperatures as per Lewis' cosmochemical model (1974) are a good fit with calculated temperatures for the time interval $(2-5) \cdot 10^5$ years.

This model of protoplanetary disk formation predicts a significantly different solid matter thermal history: vaporization and subsequent condensation of dust grains in the internal region of the solar system as it forms, and the conservation of interstellar dust (including organic compounds and ices) in the peripheral (greater) portion of the planetary system. These predictions are in qualitative agreement with cosmochemical data and, in particular, with the latest data on Halley's comet (Mukhin *et al.* 1989). According to estimates of thermal conditions in the disk, the radius of the zone of vaporization of silicate and iron dust grains (T $\gtrsim$ 1400 K) reached 1 - 2 AU. However, matter which passes through a vaporization region as the disk forms is dispersed to a larger region. This is a consequence of both the mean, outwardly-directed flux of matter in the internal portion of the disk (Ruzmaikina and Maeva 1986), and turbulent mixing.

REFERENCES

Adams, F.C., and F.H. Shu. 1985. Infrared emission from protostars. Astrophys. J. 296:655-669.

Adams, F.C., and F.H. Shu. 1986. Infrared spectra of rotating protostars. Astrophys. J. 308:836-853.

Adams, M.T., K.M. Strom., and S.E. Strom. 1983. The star-forming history of the young cluster NGC 2264. Astrophys. J. Suppl. 53:893-936.

Arquilla, R. 1984. The structure and angular momentum content of dark clouds. Dissertation, University of Massachusetts.

Aumann, H.H., F.C. Gillett, C.A. Beichman, T. de Jong, J.R. Houck, F. Low, G. Neugebauer, R.G. Walker, and P. Wesselius. 1984. Discovery of a shell around Alpha Lyrae. Astrophys. J. 278: L23-L27.

Backman, D. 1987. IRAS statistics on IR-excesses and models of circumstellar disks. IAU Coll. No. 99, Hungary, June 22-27.

Bertrout, C., G. Basri, and J. Bouier. 1988. Accretion disks around T Tauri stars. Astrophys. J. 330:350-373.

Bodenheimer, P. 1981. The effect of rotation during star formation. Pages 5-26. In: Sugimoto, D., D.Q. Lamb, and D.N. Schramm (eds). Fundamental Problems in the Theory of Stellar Evolution. D. Reidel, Dordrecht.

Boss, A.P. 1984. Protostellar formation in rotating interstellar clouds. IV. Nonisothermal collapse. Astrophys. J. 277:768-782.

Boss, A.P. 1985. Angular momentum transfer by gravitational torques and the evolution of binary protostars. Icarus 61:3-17.

Boss, A.P. 1987. Protostellar formation in rotating interstellar clouds. VI. Nonuniform initial conditions. Astrophys. J. 319:149-161.

Cabot, W., V.M. Canuto, O. Hubickyj, and J.B. Pollack. 1987. The role of turbulent convection in the primitive solar nebula I. Theory. II. Results. Icarus 69:387-442, 423-457.

Cameron, A.G.W. 1962. The formation of the sun and planets. Icarus 1:13-69.

Cameron, A.G.W., and M.B. Fegley. 1982. Nucleation and condensation in the primitive solar nebula. Icarus 52:1-13.

Campbell, B., G.A.H. Walker, S. Yang. 1988. A search for substellar companions to solar-type stars. Astrophys. J. 331: 902-921.

Cassen, P.M., and A. Moosman. 1981. On the formation of protostellar disks. Icarus 48:353-376.

Cassen, P.M., and A.L. Summers. 1983. Models of the formation of the solar nebula. Icarus 53:26-40.

Kraicheva, Z.G., E.I. Popova, A.V. Tutukov, and L.P. Yungelson. 1978. Some properties of spectrally binary stars. Astron. J. 55:1176-1189.

Larson, R.B. 1969. Numerical calculations of the dynamics of a collapsing protostar. Mon. Not. Roy. Astron. Soc. 168:271-295.

Levy, E.H., and C.P. Sonett. 1978. Meteorite magnetism and early solar-system magnetic fields. Pages 516-532. In: Gehrels, T. (ed). Protostars and Planets. University of Arizona Press, Tucson.

Lin, D.N.C., and J. Papaloizou. 1980. On the structure and evolution of the primordial solar nebula. Mon. Not. Roy. Astron. Soc. 191:37-48.

Lin, D.N.C., and J. Papaloizou. 1985. On the dynamical origin of the solar system. Pages 981-1007. In: Black, D.C. and M.S. Matthews. Protostars and Planets. II. University of Arizona Press, Tucson.

Lynden-Bell, D., and J.E. Pringle. 1974. The evolution of viscous disks and the origin of the nebular variables. Mon. Not. Roy. Astron. Soc. 168:603-637.

Lewis, J.S. 1974. The temperature gradient in the solar nebula. Science 186:440-443.

Myers, P.S. 1983. Dense cores in dark clouds. III. Subsonic turbulence. Astrophys. J. 257:620-632.

Myers, P.S., and P.J. Benson. 1983. Dense cores in dark clouds. II. NH$_3$ observations and star formation. Astrophys. J. 266: 309-320.

Makalkin, A.B. 1987. Protoplanetary disk thermics. Astron. vestnik 21:324-327.

Makalkin, A.B., and V.A. Dorofeeva. 1989. P-T conditions in a preplanetary gas-dust disk and dust component evolution. Pages 46-88. In: Magnitskiy, V.A. Planetary Cosmogony and Earth Science. Nauka, Moscow.

Marochnik, L.S., and L.M. Mukhin. 1988. "Vega" and "Jotto" missions: Does an invisible mass exist in the solar system? Letters to Astron. J. 14:564-568.

Mukhin, L.M., T.V. Ruzmaikina, and A.I. Grechinskiy. 1989. The nature of the dust component of Halley's comet. Kosmicheskye issledovania. 27:280-286.

Pollack, J.B., C.P. Mckay, and B.M. Christofferson. 1985. A calculation of a Rosseland mean opacity of dust grains in primordial solar system nebula. Icarus 64:471-492.

Ruden, S.P., and D.N.C. Lin. 1986. The global evolution of the primordial solar nebula. Astrophys. J. 308:883-901.

Ruzmaikina, T.V. 1981. On the role of magnetic fields and turbulence in the evolution of the presolar nebula. 23rd COSPAR Meeting, Budapest. Adv. Space Res. 1:49-53.

Ruzmaikina, T.V. 1981. The angular momentum of protostars that generate protoplanetary disks. Letters to Astron. J. 7:188-192.

Ruzmaikina, T.V. 1982. In: Volk, H. (ed.). Diskussions forum Ursprung des Sonne-systems. Mitt. Astron. Ges. 57:49-53.

Ruzmaikina, T.V. 1985. The magnetic field of a collapsing solar nebula. Astron. vestnik 18:101-112.

Ruzmaikina, T.V., and S.V. Maeva. 1986. Research on the process of protoplanetary disk formation. Astron. vestnik 20:212-227.

Ruzmaikina, T.V., and V.S. Safronov. 1985. Premature particles in the solar nebula. LPSC XVI (Abstracts):720-721.

Ruzmaikina, T.V., V.S. Safronov, and S.J. Weidenschilling. 1989. Radial mixing of material in the asteroid zone. University of Arizona Press, Tucson, in press.

Safronov, V.S. 1969. Preplanetary cloud evolution and the formation of Earth and the planets. Nauka, Moscow.

Safronov, V.S., and T.V. Ruzmaikina. 1978. On the angular momentum transfer and accumulation of solid bodies in the solar nebula. Pages 545-564. In: Gehrels, T. (ed). Protostars and Planets. University of Arizona Press, Tucson.

Safronov, V.S., and T.V. Ruzmaikina. 1985. Formation of the solar nebula and the planets. Pages 959-980. In: Black, D. and M.S. Matthews. Protostars and Planets. II. University of Arizona Press, Tucson.

Sargent, A.I., and S. Beckwith. 1987. Kinematics of the circumstellar gas of HL Tau and R Monocerotis. Astrophys. J. 323:294-.

Smith, B.A., and R.J. Terrile. 1984. A circumstellar disk around β. Pictoris. Science 226:1421-1424.

Tereby, S., F.H. Shu, and P.M. Cassen. 1984. The collapse of cores of slowly rotating isothermal clouds. Astrophys. J. 286: 529-55.

Schakura, N.I., and R.A. Sunyaev. 1973. Black holes in binary systems. Observational appearance. Astron. Astrophys. 24:337-355.

Stahler, S.W., F.H. Shu, and R.E. Taam. 1980. The evolution of protostars. I. Global formulation and results. Astrophys. J. 241:637-654.

Vityazev, A.V., and G.V. Pechernikova. 1985. On the evaporation of dust particles during the preplanetary disk formation. LPSC XVI Abstracts:885-886.

Weidenschilling, S.J. 1977. The distribution of mass in the planetary system and solar nebula. Astrophys. Space Sci. 51: 153-158.

Yuan, C., and P. Cassen. 1985. Protostellar angular momentum transport by spiral density waves. Icarus 64:435-447.

Physical-Chemical Processes in a Protoplanetary Cloud

AVGUSTA K. LAVRUKHINA
V.I. Vernadskiy Institute of Geochemistry and Analytic Chemistry

ABSTRACT

According to current views, the protosun and protoplanetary disk were formed during the collapse of a fragment of the cold, dense molecular interstellar cloud and subsequent accretion of its matter to a disk. One of the most critical cosmochemical issues in this regard is the identification of relics of such matter in the least altered bodies of the solar system: chondrites, comets, and interplanetary dust. The presence of deuterium-enriched, carbon-containing components in certain chondrites (Pillinger 1984) and radicals and ions in comets (Shulman 1987) is evidence that this area holds great promise. If a relationship is established between solar nebula and interstellar matter, we can then identify certain details, such as the interstellar cloud from which the Sun and the planets were formed. We can also come to a deeper understanding of the nature of physico-chemical processes in the protoplanetary cloud which yielded the tremendous diversity of the chemical and mineralogical compositions of the planets and their satellites, meteorites, and comets.

CHARACTERISTICS OF THE CHEMICAL COMPOSITION OF MOLECULAR INTERSTELLAR CLOUDS

One would expect that chemical compositions of interstellar clouds are significantly varied and are a function of such physical parameters as temperature and density. Moreover, one would expect that they are also dependent on the age of a cloud, its history, the impoverishment of

elements, the flux of the energy particles of cosmic rays and photons, and the flow of material emanating from stars located in or adjacent to a cloud.

Several objects have been studied in the greatest detail at this time (Irvine *et al.* 1987): (1) The core of the KL region of the Orion nebula. It contains at least four identified subsources with varying chemistry. (2) The central galactic cloud Sgr B2. It is the most massive of all the known gigantic molecular clouds and contains high-luminosity stars. (3) The cold, dark, low-mass clouds, TMC-1 and L134N. They have substructures with individual "lumps" with a mass of several $M_\odot$. It has been suggested that such areas correspond to locations where solar-type stars were formed. TMC-1, in particular, matches that portion of ring material which is predicted by the rotating molecular cloud collapse model. (4) Clouds in the spiral arm. These are sources of HII areas or are the remnants of supernova CaSA. (5) The expanding envelope of the evolving carbon star, IRC + 10216 (CWLeO). Table 1 provides data on 80 molecules which were discovered by 1988 in the interstellar medium. Of these, 13 are new and three are the cyclical molecules, C_3H_2, SiC_2 and c-C_3H. A maximum high deuterium-enrichment $D/H \approx 10^4 \times (D/H)_{cosm}$ (Bell *et al.* 1988) is characteristic for the first of these. The ratio $[C_3HD]/[C_3H_2]$ lies within the range 0.05-0.15 for 12 dark, cold (approximately 10K) clouds. The first phosphorus compound (PN) discovered in the interstellar medium is among the new molecules. Phosphorus nitride has been discovered in three gigantic molecular clouds and, in particular, in the Orion nebula. In the Orion nebula it associates with dark gas flowing from the infrared source IRc2. This is most likely a protostar. PN abundance is low, while the search for other phosphorus compounds (PH_3, HCP, and PO) has yet to meet with success. This may seem strange, because P abundance in diffuse clouds approaches cosmic levels. However, many metals and S are impoverished because they are among the constituents of dust grains.

Data relating to the discovery of NaCl, AlCl, KCl, and AlF in envelopes of evolving stars are of tremendous interest. The carbon star IRC +10216 is an example. The distributions of these compounds are in agreement with calculations of chemical equilibria for the atmospheres of carbon-rich stars (C/O > 1) in the area T = 1200-1500K. In addition to the molecules indicated above, SiH_4, CH_4, $H_2C=CH_2$, $CH=CH$, SiC_2 have been discovered in the envelopes of stars. The latter molecule is particularly interesting: it broadens the range of molecules which condense at C/O > 1. Data on the search for O_2 in six dark clouds are also evidence of the presence of dark clouds which characterize an oxygen insufficiency.

Two important consequences follow from the data contained in Table 1. (1) The obviously nonequilibrious nature of interstellar chemistry. Evidence of this is the presence of highly reactive ions and radicals with one or two unpaired electrons. (2) The presence of numerous molecules with

TABLE 1 Identification of Interstellar Molecules (Irvine 1988)

Simple hydrides, oxides, sulfides and other molecules

H_2	CO	NH_3	CS	$\underline{NaCl^x}$
HCl	SiO	SiH_4^x	SiS	$\underline{AlCl^x}$
H_2O	SO_2	CC	H_2S	$\underline{KCl^x}$
	OCS	CH_4^x	$\underline{PN}$	$\underline{AlF^x}$
	HNO ?			

Nitrides, derivative acetylenes, and other molecules

HCN	$HC{\equiv}C\text{-}CN$	$H_3C\text{-}C{\equiv}C\text{-}CN$	$H_3C\text{-}CH_2\text{-}CN$	$H_2C{=}CH_2^x$
H_3CCN	$H(C{\equiv}C)_2\text{-}CN$	$H_3C\text{-}C{\equiv}CH$	$H_2C{\equiv}CH\text{-}CN$	$HC{\equiv}CH^x$
CCCO	$H(C{\equiv}C)_3\text{-}CN$	$H_3C\text{-}(C{\equiv}C)_2\text{-}H$	HNC	
$\underline{CCCS}$	$H(C{\equiv}C)_4\text{-}CN$	$H_3C\text{-}(C{\equiv}C)_2\text{-}CN?$	HN=C=O	
HC=CCHO	$H(C{\equiv}C)_5\text{-}CN$		HN=C=S	
$\underline{H_3CNC}$				

Aldehydes, alcohols, ethers, ketones, amides, and other molecules

$H_2C{=}O$	H_3COH	$HO\text{-}CH{=}O$	H_2CNH
$H_2C{=}S$	$H_3C\text{-}CH_2\text{-}OH$	$H_3C\text{-}O\text{-}CH{=}O$	H_3CNH_2
$H_3C\text{-}CH{=}O$	H_3CSH	$H_3C\text{-}O\text{-}CH_3$	H_2NCN
$NH_2\text{-}CH{=}O$	$\underline{(CH_3)_2CO?}$	$H_2C{=}C{=}O$	

Cyclical molecules

C_3H_2	SiC_2^x	$\underline{c\text{-}C_3H}$

Ions

CH^+	HCO^+	$HCNH^+$	$H_3O^+?$
HN_2^+	$HOCO^+$	SO^+	$HOC^+?$
	HCS^+		$H_2D^+?$

Radicals

OH	C_3H	CN	HCO	$\underline{C_2S}$
CH	C_4H	C_3N	NO	NS
C_2H	C_5H	$\underline{H_2CCN}$	SO	
	$\underline{C_6H}$			

(a) New molecules discovered after 1986 are underlined.
(x) Present only in clouds of evolving stars.
(?) Not yet confirmed.

unsaturated bonds, despite the fact that hydrogen distribution rates are three to four orders greater than for C, N, and O. This is evidence of the predominance of kinetic over thermodynamic factors in chemical reactions in the interstellar medium and of the large contribution of energy from cosmic rays and UV radiation to these processes. Chemically saturated compounds such as CH_3CH_2CN are only present in the "warmer" sources (i.e., in Orion). $HNCO$, CH_3CN, HC_3N, C_2H_3CN, C_2H_5CN levels are higher in warm clouds, possibly owing to higher NH_3 parent molecule levels.

One interesting feature pertaining to the distribution rates for certain interstellar molecules is their uniformity for dark molecular clouds with wide variation in P and T parameters. Furthermore, an inverse dependence of the amount of gas molecules on dust density is absent. This would have been an expected consequence of molecules freezing into the ice mantle of particles. This is confirmed by data on the constancy of the CO/dust ratio in three clouds. It is further supported by the absence of a drop in H_2CO levels as dust density rises in dark molecular clouds. An indication that the efficiency rate of this in-freezing is not uniform for various molecules has also not been confirmed. A high degree of homogeneity of the $H^{13}CO/^{13}CO$ and $C_2H/^{13}CO$ ratios for many clouds has been found. Clearly, the processes involved in the breakdown of particle ice mantles are highly efficient. Their efficiency may be enhanced when dust grain density increases as the grains collide with each other.

Data on the distribution of different interstellar molecules are in general agreement with calculations in which ion-molecular reactions in gas are the primary process. However, there is a question as to the reliability of calculations with a value of CO < 1 in a gas phase. It has been found that the abundance values for many C-rich molecules and ions are extremely low in steady-state conditions. Despite the fact that various explanations of these facts have been offered, an alternative hypothesis suggests that C/O > 1 in the gas phase. Other facts were already indicated above which can be explained by such a composition of the gas phase. Enhanced carbon levels may be attributed to the fact that CH_4 (being a nonpolar molecular) is more easily volatilized from the surface of the grain mantle than NH_3 and H_2O.

Therefore, we can hypothesize that in certain dark molecular clouds or in different portions of them, the gas phase has a C/O ratio which departs from the cosmic value. This is fundamentally critical to understanding the processes in the early solar system. It has been found that many unique minerals of enstatite chondrites (including enstatite, silicon-containing kamasite, nainingerite, oldgamite, osbornite, and carbon) could only have been formed during condensation from gas with C/O > 1 (Petaev *et al.*

1986). SiC and other minerals, which were condensed in highly reducing conditions, have also been found in CM-type carbonaceous chondrites (Lavrukhina 1983).

PROPERTIES AND PHYSICO-CHEMICAL PROCESS IN THE GENESIS OF INTERSTELLAR DUST GRAINS

According to current thinking (Voshchinnikov 1986), the total sum of molecules in a "dense," not-too-hot gaseous medium of complex molecular composition precipitates into a solid phase, thereby forming embryos of dust grains. These grains then begin to grow through accretion of other molecular compounds or atoms. The grains may in turn act as catalysts for reactions to form new types of molecules on their surface. A portion of these remains as particles, and the rest converts to the gaseous phase.

Laminated interstellar grains are formed in this manner. Their cores are made up of refractory silicate compounds, metal iron, and carbon. The grain mantle is formed from a mixture of ices of water, ammonia, methane, and other low-temperature compounds with varying admixtures. Atomic carbon may also be adsorbed on the mantle surface at the low temperatures of dark molecular clouds. These dust grains are often aspherical. Their size is approximately 0.3 μm. Generation of the finest particles ($\lesssim$ 0.01 μm) also takes place. They have no mantle due to the increased temperature of these grains as a single photon is absorbed or a single molecule is formed. The dust grains are usually coalesced as a result of photoelectron emission and collisions with electrons and ions. Mean grain temperature is approximately 10 K.

The following data are evidence of the chemical composition of dust grains.

(1) IR- absorption band:

λ 3.1 μm – ice H_2O (NH_3),

λ λ 9.7 and 18 μm - amorphous silicates,

λ λ 4.61 and 4.67 μm – molecules with the groups CN and CO,

λ λ 3.3 $\div$ 35 μm – molecules with the groups CH_2- and -CH_3.

(2) Emission spectra:

λ 11.3 μm – SiC,

λ 30 μm – mixtures of MgS, CaS, FeS_2,

λ 3.5 μm – formaldehyde (H_2CO),

six emission bands with λ λ 3.28 $\div$ 11.2 μm $\div$ large organic molecules (N_C ~ 20)

(3) λ 2200Å $\div$ graphite (?), carbines (-C≡C-), amorphous and glassy carbon.

According to current views, at least part of the interstellar molecules is formed from reactions on dust grain surfaces. At low-grain surface temperatures and moderately high gas temperatures, atoms and molecules coming into contact with the surface may adhere to it. Van-der-Vaals effects determine a minimum binding energy value. However, significantly high values are also possible with chemical binding. Migration along the grain surface of affixed atoms generates favorable conditions for molecule formation. A portion of the released binding energy (Ec) of atoms in a molecule is taken up by the crystal grid of the grain surface. If the remaining portion of the molecule's energy is greater than Ec, the molecule "comes unglued" and is thrown into the gas phase. This process is accelerated when the dust grains are heated by cosmic rays. H_2, CH_4, NH_3, and H_2O molecules are formed in this manner. Since the binding energy of C, N, O, and other atoms is on the order of 800 K, they adhere to the grain surface where they enter into chemical reactions with hydrogen atoms. The aforementioned molecules are thus formed. Part of these then "comes unglued," such as the H_2 molecule. A portion freezes to the grain surface. The temperatures at which molecules freeze are equal to (K): H_2 — 2.5, N_2 — 13, CO — 14, CH_4 — 19, NH_3 — 60, and H_2O — 92. Hence, the formation of the mantle of interstellar dust grains and certain molecules in the gaseous phase of dense, gas-dust clouds occurs contemporaneously.

Table 2 lists certain data on the characteristics of the basic physical and chemical processes involved in the formation and subsequent evolution of interstellar dust grains and the astrophysical objects in which these processes occur. With these data we can evaluate the nature of processes occurring in the protoplanetary cloud during the collapse and subsequent evolution of the Sun. These basic processes triggered: 1) the breakdown and vaporization of dust grains under the impact of shock waves at collapse and the accretion of primordial cloud matter onto the protoplanetary disk; 2) the collision of particles; and 3) particle irradiation by solar wind ions. The role of these processes varied at different distances from the protosun. Yet the main outcome of the processes is that organic, gas-phase molecules and the cores of dust grains (surrounded by a film of high-temperature polymer organic matter) are present throughout the entire volume of the disk. They are obvious primary-starting material for the formation of a great variety of organics which are observed in carbonaceous chondrites (Lavrukhina 1983). Dust grains at great distances from the protosun (R $\gtrsim$ 2 AU) will be screened from the impact of high temperatures and solar radiation. They will therefore remain fairly cold in order to conserve water and other volatile molecules in the mantle composition. Comets, obviously, contain such primary interstellar dust grains.

TABLE 2   Characteristics of the Basic Physico-Chemical Processes of Interstellar Dust Grain Genesis

Process	Parameters	Proposed Chemical Compounds or Processes	Astrophysical Objects
1. Condensation of high temperature embryos	$T=1400$- $-1280K$	Amorphous silicates, mixes of oxides MgO, SiO, CaO, FeO, Fe, Ni-particles, SiC, carbines, graphite(?) amorphous & glassy carbon	1) Atmospheres of cold stars $(10^{10}$-$10^{11}cm^{-3})$, 2) Planetary nebulae 3) Envelopes of novae and supernovae, 4) Envelopes of red giants
2. Formation of mantle on embryos	$T=700$-$25K$	FeS, H_2O, $NH_3 \cdot H_2O$, CH_4 x H_2O Solid clatrates Ar, Kr, Xe. Carbines	1) Upper layers of cold stars and interstellar space, 2) Dispersed matter of old planetary nebulae, 3) GMC,[x] 4) Old supernovae envelopes
3. Coalescence of fine particles with formation of "sleeve" pooding type particles	$\bar{t} \sim 10yrs$	Ice with phenocrysts from silicates, metals graphite (?)	Turbulent gas of protostellar clouds
4. Destruction of dust grains (primarily in mantles)	Particle life-span: ice-$(10^7$-$5 \cdot 10^8)$ years, silicate-$(4 \cdot 10^8$-$2 \cdot 10^{10})$ years	1) Collision of particles with $V > 20km \cdot c^{-1}$, 2) Sublimation, 3) Physical and Chemical destruction, 4) Photodesorption	1) Envelopes of red giants and novae, 2) Shock waves from supernovae flash, 3) Irradiation by high velocity ions of stellar wind and by high energy cosmic rays in GMC[x] and planetary nebulae
5. Oxidation-reduction reactions on grain surface	Low-temperature Fe oxidation by monatomic oxygen	FeO, Fe_2O_3, Fe_3O_4	Diffusion interstellar medium
Formation of hydrides	$T=2.5$-5 K	FeH, FeH_2, hydrides of transitional metals	Dark GMC[x] zones
6. Formation of envelopes and dust from solid organic compounds	Radiation polymerization of organic compounds with $T \geq 4K$ on grains with subsequent breakdown into fragments	Tolines, hexamethylentetramine, cellulose, complex organic or prebiological compounds (PAC)	UV-radiation, cosmic rays, shock waves in GMC[x]

(x) Gigantic dark molecular interstellar clouds

THE ISOTOPE COMPOSITION OF VOLATILES IN
BODIES OF THE SOLAR SYSTEM

Investigation of the isotope composition of H, O, C, N, and the inert gases in meteorites, planets, and comets is extremely relevant as we attempt to understand the processes involved in the genesis of the preplanetary cloud. The majority of these elements had a high abundance in the interstellar gas and the gas-dust, initial protosolar cloud. Great variation in the isotope composition for various cosmic objects is also characteristic of these elements. Such variation has made it possible to refute outmoded views of the formation of the protosolar cloud from averaged interstellar material (Lavrukhina 1982; Shukolyukov 1988).

From detailed studies of meteorites, we have been able to discover a number of isotopically anomalous components and identify their carrier phases (Anders 1987). These studies have demonstrated that the pro-tomatter of the solar system was isotopically heterogeneous. For example, examination of the hydrogen isotope has shown that objects of the solar system can be subdivided into three groups in terms of the hydrogen isotopy (Eberhardt *et al.* 1987). (1) Deuterium-poor interstellar hydrogen, proto-solar gas, and the atmospheres of Jupiter and Saturn; (2) deuterium-rich interstellar molecules (HNC, HCN, and HCO^+) of dark molecular clouds of Orion A; and (3) the atmospheres of Earth, Titanus, and Uranus, the water of Halley's comet, interplanetary dust, and certain chondrite fractions of Orgueil Cl and Semarkona LL3 occupy an intermediary position. Clearly, the isotopic composition of hydrogen in these components is determined by the mixing of hydrogen from two sources: a deuterium-poor and a deuterium-rich source. A single gas reservoir is thus formed.

A similar situation has been found for oxygen. Two oxygen components have been discovered in meteorites: impoverished and enriched ^{16}O of nucleogenetic origin (Lavrukhina 1980). The relative abundance of oxygen isotopes in chondrules tells us that chondrite chondrules of all chemical groups are convergent in relation to a single oxygen reservoir, characterized by the values $\delta\,^{18}O = 3.6 \pm 0.2L\%$ and $\delta\,^{17}O = 1.7 \pm 0.2L\%$ (Lavrukhina 1987; Clayton *et al.* 1983). They are similar to the corresponding values for Earth, the Moon, achondrites, pallasites, and mesosiderites. On the basis of these data and the dual-component, isotopic composition of nitrogen, carbon, and the inert gasses (Levskiy 1980; Anders 1987), workers have raised the idea that protosolar matter was formed from several sources. For example, two reservoirs of various nucleosynthesis are proposed that differ in terms of their isotopic composition and the degree of mass fractionization (Lavrukhina 1982; Levskiy 1980). Shukolyukov (1988) proposes three sources: ordinary interstellar gas; material injected into the solar system by

an explosion of an adjacent supernova; and interstellar dust made up of a mixture of different stages of stellar nucleosynthesis.

The presence of at least two sources of matter in protoplanetary matter may be evidence of the need to reconsider the hypothesis of the contemporaneous formation of the Sun and the protoplanetary cloud from a single fragment of a gigantic molecular interstellar cloud.

REFERENCES

Anders, E.A. 1987. Local and exotic components of primitive meteorites and their origin. Trans. R. Soc. Lond. A323(1):287-304.

Bell, M.B., L.W. Avery, H.E. Matthews *et al.* 1988. A study of C_3HD in cold interstellar clouds. Astrophys. J. 326(2):924-930.

Clayton, R.N., N. Onuma, Y. Ikeda. 1983. Oxygen isotopic compositions of chondrules in Allende and ordinary chondrites. Pages 37-43. In: King, E.A. (ed.). Chondrules and Their Origins. Lunar Planet. Institute, Houston.

Eberhardt, P., R.R. Hodges, D. Krankowsky *et al.* 1987. The D/H and 18O/16O isotopic ratios in comet Halley. Lunar and Planet. Sci. XVIII:251-252.

Irvine, W.M. 1988. Observational astronomy: recent results. Page 14. Preprint, Five College Ratio Astronomy Observatory. University of Massachusetts, Amherst.

Irvine, W.M., P.F. Goldsmith, and Å. Hjalmarson. 1987. Chemical abundances in molecular clouds. Pages 561-609. In: Hollenbach, D.J, and H.A. Thronson (eds.). Interstellar Processes. D. Reidel Publishing Company, Dordrecht.

Lavrukhina, A.K. 1987. On the origin of chondrules. Pages 75-77. In: XX Nat. Meteorite Conference: Thes. Paper. GEOKHI AS USSR, Moscow.

Lavrukhina, A.K. 1983. On the genesis of carbonaceous chondrite matter. Geokhimiya 11:1535-1558.

Lavrukhina, A.K. 1982. On the nature of isotope anomalies in the early solar system. Meteoritika 41:78-92.

Levskiy, L.K. 1989. Isotopes of inert gases and an isotopically hetereogeneous solar system. Page 7. In: VIII Nat. Symposium on Stable Isotopes in Geochemistry: Thes. Paper. GEOKHI AS USSR, Moscow.

Petaev, M.I., A.K. Lavrukhina, and I.L. Khodakovskiy. 1986. On the genesis of minerals of carbonaceous chondrites. Geokhimiya 9:1219-1232.

Pillinger, C.T. 1984. Light element stable isotopes in meteorites—from grams to picograms. Geochim. et Cosmochim. Acta 48(12):2739-2766.

Shukolyukov, Yu.A. 1988. The solar system's isotopicnonuniformity: principles and consequences. Geokhimiya 2: 200-211.

Shulman, L.M. 1987. Comet Nuclei. Nauka, Moscow.

Voshchinnikov, N.V. 1986. Interstellar dust. Pages 98-202. In: Conclusions of Science and Technology. Space research, vol 25. VINITI, Moscow.

Magnetohydrodynamic Puzzles in the Protoplanetary Nebula

EUGENE H. LEVY
University of Arizona

ABSTRACT

Our knowledge of the basic physical processes that governed the dynamical state and behavior of the protoplanetary accretion disk remains incomplete. Many large-scale astrophysical systems are strongly magnetized and exhibit phenomena that are shaped by the dynamical behaviors of magnetic fields. Evidence and theoretical ideas point to the possibility that the protoplanetary nebula also might have had a strong magnetic field. This paper summarizes some of the evidence, some of the ideas, some of the implications, and some of the problems raised by the possible existence of a nebular magnetic field. The aim of this paper is to provoke consideration and speculation, rather than to try to present a balanced, complete analysis of all of the possibilities or to imagine that firm answers are yet in hand.

INTRODUCTION

Magnetic fields are present and dynamically important in a wide variety of astrophysical objects. There are at least three reasons why such magnetic fields provoke interest: 1) The presence of a magnetic field invites questions as to the conditions of its formation, either as a relict from some earlier generation process, carried in and reshaped, or as a product of contemporaneous generation; 2) the Lorentz stresses associated with magnetic fields are important to the structure and dynamical evolution of many systems; and 3) magnetic fields can store, and quickly release, prodigious quantities of energy in explosive flares. The possible presence of a magnetic field in the protoplanetary nebula raises questions in all three of these areas.

MAGNETIZATION OF METEORITES

Perhaps the most provocative, yet still puzzling and ambiguous, evidence for strong magnetic fields in the protoplanetary nebula comes from the remanent magnetization of primitive meteorites. Our ignorance of the detailed history of the meteorites themselves, including the processes of their accumulation, and ambiguities in the magnetic properties of the meteorite material, causes difficulties in interpreting the significance of meteorite magnetization. These latter ambiguities are especially confusing to neat interpretations of the absolute intensities of the magnetic fields in which the meteorites acquired their remanence. A further complication arises because the meteorites exhibit a diversity of magnetizations, blocked at different temperatures and in different directions on different scales. It seems clear that a thorough understanding of meteorite magnetization, and its unambiguous interpretation, has yet to be written (Wasilewski 1987).

A variety of primitive meteorite materials carry remanent magnetization (Suguira and Strangway 1988). Of the wide variety of characteristics of meteorite remanence that have been measured, two regularities in particular seem to be important. First, the intensities of the remanence-specific magnetic moments seem typically to be larger for small components of the meteorites (e.g., chondrules and inclusions) than for the aggregate rocks. Second, the small, intensely magnetized components are frequently disordered and oriented in random directions.

The inferred *model* magnetizing intensities for "whole rock" samples tend to fall in the range of 0.1—1 Gauss (Nagata and Suguira 1977; Nagata 1979). The connection of this "model" magnetizing intensity to a real physical magnetic field depends to some extent on the way in which the meteorite accumulated and subsequently evolved in the presence of a magnetic field, as well as on the prior magnetic history of the individual components. More provocatively, the inferred magnetizing fields for individual small components of meteorites (for example, chondrules) range to the order of 10 Gauss (Lanoix *et al.* 1978; Suguira *et al.* 1979; Suguira and Strangway 1983). Probably even these inferences, although based on measurements of some of the physically simplest and least heterogeneous of meteorite material, should be regarded as tentative, pending wider-ranging and deeper studies of the processes involved in the acquisition of meteorite remanence.

One interpretation of the measurements is that the meteorite parent bodies were assembled out of small components that were already intensely magnetized as individual, free objects *before* they were incorporated into larger assemblages (Suguira and Strangway 1985). This interpretation, while simple and consistent with the measurements, is probably not unique.

One might imagine, for example, that the randomization of the small components occurred after magnetization, from an internally generated field on a larger object, during the subsequent churning of a regolith. However, for the present we will accept as a tentative inference that magnetizing fields as high as 10 Gauss might have occurred in the protoplanetary nebula. Inasmuch as the meteorites seem to have formed at several astronomical units, of the order of three, from the Sun, this 10 Gauss magnetic field also seems likely to have existed at that distance, although other possibilities are not strictly ruled out.

THE POSSIBLE ORIGIN OF A NEBULAR MAGNETIC FIELD

Assuming the presence of such a nebular magnetic field, there are several ways, in principle, that it could have arisen. One possibility is that the field could have been generated in the Sun; another possibility is that the field could have been a manifestation of the interstellar magnetic field compressed by the collapse of the protosolar gas (Safronov and Ruzmaikina 1985). Consider, however, the possibility that the nebular magnetic field was rooted in the early Sun. This possibility is largely attractive in the case that the nebula itself was too poor an electrical conductor to have its own magnetohydrodynamic character. In that case, the solar magnetic field must fall off at least as fast as r^{-3} in the electrically nonconducting space outside of the Sun. (In fact, if there is some ionized gas outflow from the Sun, even outside of the disk, the field might fall off somewhat less rapidly with distance, but probably not so differently as to alter the general conclusion here.) Then for a 10 Gauss magnetic field at 5×10^{13} cm from a 10^{11} cm radius Sun, the solar surface magnetic field would have had to have been about 10^9 Gauss. Such a situation would have had a profound influence on the Sun, especially with respect to the dynamical equilibrium and stability of the system, since the energy associated with such a field would have been comparable to the gravitational binding energy of the Sun. Clearly, there are many other aspects of this that could be considered. However, the purpose here is not to rule out completely the possibility of an important nebular field arising from the Sun; rather it is to indicate that such an assumption does not lead to an easy and obvious solution of the problem of a nebular magnetic field.

Here we will presume a turbulent nebula in which the short mixing times and low electrical conductivity, and the consequent rapid dissipation of magnetic fields, quickly erase any memory of the nebular fluid's previous magnetization. In this case, if the nebula is to carry a large-scale magnetic field, then it must be contemporaneously generated, most likely by some sort of dynamo process (Parker 1979; Zel'dovich and Ruzmaikin 1987). The possible existence of conditions in the nebula that could have allowed

the generation of such a magnetic field raises substantial physical questions. However, in the spirit of the present discussion, we will assume whatever is necessary and come back to the implications in the end.

The ability of a fluid flow to generate a magnetic field through hydromagnetic dynamo action can, in simple cases, be parameterized by a dimensionless number called the dynamo number, N, where

$$N = \frac{\gamma \Gamma \delta^3}{\eta^2}. \tag{1}$$

Γ is a measure of the helical part of the convection while γ measures the strength of the fluid shear, and η is the magnetic diffusivity, $c^2/4\pi\sigma$, with electrical conductivity σ; δ is the scale length of the magnetic field. Following a simple analysis for accretion disks (Levy 1978; Levy and Sonett 1978),

$$\gamma \equiv \frac{\partial V_\phi}{\partial r} - \frac{V_\phi}{r} \approx \frac{3}{2}\Omega. \tag{2}$$

In a Keplerian disk, $\Omega = (GM_\odot/r^3)^{1/2}$, and to order of magnitude, $\Gamma \approx \ell\Omega$, where ℓ is the large scale of the turbulence. Numerically then, $\gamma \sim 1.7 \times 10^{13}r^{-3/2}$ sec^{-1} and, $\Gamma \sim 1.1 \times 10^{13}\ell r^{-3/2}$ cm sec^{-1}. Now, taking both the scale of the largest eddies, l, and the scale of the magnetic field, δ, to be of the order of the scale height of the disk gas, $\sim 5 \times 10^{12}$ cm, we find that $N \sim 1.3 \times 10^{35} \eta^{-2}$, when r ~ 3 AU. Recent detailed numerical calculations of magnetic field generation in Keplerian disks indicate (Stepinski and Levy 1988) that magnetic field generation occurs at $N \approx 10^2$. Putting these results together, we find that a regenerative dynamo can be expected to be effective in such a disk if the electrical conductivity exceeds about 500 sec^{-1}. We will return to this question in the end.

Now consider the strength that such a magnetic field might attain. Inasmuch as the essential regenerative character of a dynamo fluid motion is associated with the cyclonic or helical component of the motion, which results from the action of the Coriolis force, then one estimate of the possible maximum amplitude of a dynamo magnetic can be derived from balancing the Coriolis force and the Lorentz stress:

$$\rho V\Omega \sim \frac{< B_p B_\phi >}{4\pi\delta}. \tag{3}$$

B_p and B_ϕ represent the poloidal and toroidal parts of the magnetic field respectively. Taking $\rho \sim 10^{-9}$ gm cm^{-3} and $V \sim 0.1$ kilometers per second, then with the other values as above, we find $< B_p B_\phi >^{1/2} \sim 10$ Gauss, as a measure of the maximum magnetic field strength that might be produced in such a nebula. Other processes also can act to limit the

strength of the magnetic field. For example, with the low ionization level indicated above for the action of a nebular dynamo, the strength of the magnetic field can be limited by the differential motion of the neutral and ionized components of the gas: a phenomenon sometimes called ambipolar diffusion. Consideration of this dynamical constraint (Levy 1978) yields a limit on the magnetic field strength similar to the one just derived.

It is provocative that this estimate of the magnetic field strength possible in a protoplanetary nebula dynamo agrees so closely with the inferred intensity of the magnetizing fields to which primitive meteorites were exposed. At this point, the general estimates are sufficiently crude, and other questions sufficiently open, that this coincidence probably cannot be considered more than provocative.

THE POSSIBILITY OF MAGNETIC FLARES

One of the most intriguing puzzles posed by meteorites is the evidence that some components were exposed to very large and very rapid transient excursions away from thermodynamic equilibrium. Specifically, meteorite chondrules are millimeter-scale marbles of rock, which apparently were quickly melted by having their temperatures transiently raised to some 1700K and then quickly cooled. While there is some uncertainty about the time scales involved, the evidence suggests time scales of minutes to hours, though some workers have suggested even shorter melting events, of the order of seconds. Although a number of possible scenarios have been suggested for the chondrule-melting events, none seem to have been established in a convincing way (King 1983; Grossman 1988; Levy 1988). Here we will focus on the possibility that chondrules melted as a result of being exposed to energetic particles from magnetic nebular flares (Levy and Araki 1988).

In astrophysical systems, explosive restructuring of magnetic fields, associated with instabilities that relax the ideal hydromagnetic constraints and allow changes in field topology, seems to be among the most prevalent of phenomena responsible for energetic transient events. Such events are well studied in the Earth's magnetosphere (where they are involved in the dissipative interaction between the solar wind and the geomagnetic field and in geomagnetic activity) and in the solar corona (where they produce solar flares and other transient manifestations). It is thought that many other explosive outbursts in cosmical systems result from similar mechanisms.

To summarize the analysis given in Levy and Araki (1988), following the simple and basic analysis given by Petschek (1964), the energy emerging from a flare event is estimated at

$$F \sim \frac{B^3}{8\pi\sqrt{4\pi\rho}} \ \mathrm{erg} \ \mathrm{cm}^{-2}s^{-1}. \tag{4}$$

Physically, this corresponds to an energy density equivalent to the energy of the magnetic field flowing at the Alfvén speed. Levy and Araki conclude that, in order to deliver energy to nebular dust accumulations at a rate sufficient to melt to the silicate rock, the flares must occur in the disk's tenuous corona, with local mass density in the range of 10^{-18} gm cm^{-3}, and with a magnetic field intensity in the range of about 5 Gauss. Under these conditions, they estimate that much of the flare energy is likely to emerge in the form of 1 MeV particles, which are channeled down along the magnetic field; in much the same way that geomagnetic-tail-flare particles are channeled to the Earth's auroral ovals. Under these conditions, Levy and Araki find that the value to which a particle's temperature can be raised is given by

$$T = \left(T_o^4 + \frac{B^3}{16\pi^{5/2}}\sigma\sqrt{\rho} \right)^{1/4}, \tag{5}$$

where T_o is the ambient temperature into which the particle radiates, and which has no substantial influence on the result. From equation (5) it is found that the above cited conditions in the flare site, $\rho \sim 10^{-18}$ gm cm^{-3} and B $\sim$ 5-7 Gauss, yield flare energy outflows sufficient to melt chondrules; substantially weaker magnetic fields or higher ambient mass densities yield energy fluxes too low to account for chondrule melting. It is easy to see that the time scale constraints for rapid chondrule formation are easily met. At the equilibrium temperature given by equation (5), the rate of energy inflow is balanced by radiative energy loss. Thus the heating time scale is of the order of the radiative cooling time scale, second to minutes, depending on the physical structure of the precursor dust accumulation, and the temperature variation of the chondrule closely tracks the variation of the energy inflow.

The conclusion from this exploration is that chondrules might plausibly have been melted from magnetic flare energy in the protoplanetary nebula. Apparently, the most reasonable conditions under which flares could have accomplished this occur for flares in a low-density corona of the disk and with magnetic fields having intensities of around 5 Gauss or somewhat greater. It is provocative that these conditions are entirely consistent with inferences about the possible character of nebular magnetic fields that were summarized in the previous two sections.

If chondrules were made in this way, it is also necessary that the locale of chondrule formation was at moderately high altitudes above the nebular midplane: below the locale of the flares, but still high enough that the

matter intervening between the flare site and the chondrule-formation site was sufficiently tenuous to allow the passage of MeV protons. This implies that the dust accumulations would have to have been melted at an altitude of about one astronomical unit above the midplane. It is conceivable that dust accumulations might have been melted into chondrules during their inward travel from interstellar cloud to the nebula. It is perhaps more likely that dust accumulations were lofted from the nebula to high altitudes by gas motions. This latter possibility requires that the precursor dust assemblages were very loose, fluffy, fairy-castle-like structures, somewhat like the dust balls that accumulate under beds (Levy and Araki 1988). However, this is perhaps the most likely physical state of early dust assemblages in the protoplanetary nebula.

Because the energetic particles associated with the flares described here are likely to have had energies in the range of an MeV, it is possible that nuclear reactions might also be induced that could account for some isotopic anomalies measured in meteorites. However, this possibility requires further investigation.

EXTERNAL MANIFESTATIONS

It is especially instructive to estimate the gross energetics of the flares described in the previous section. Again, following Levy and Araki (1988), consider that the time scale of the flare is of the order of the heating time of the chondrules, somewhere in the range of 10^2 to 10^4 seconds. Crudely, the flare energy is derived from the collapse of a magnetic structure of some spatial scale L_f in a time τ_f. The rate of such collapse is expected to occur at a fraction of the Alfvén speed, say $\sim 0.1\ V_A$, so that $L_f \sim 0.1\ V_A \tau_f$; the volume of involved magnetic field is then about $(0.1\ V_A \tau_f)^3$. Thus the total flare energy should be of the order of

$$\epsilon_f \sim 10^{-3} \frac{B^5}{64 \pi^{5/2} \rho^{3/2}} \tau_f^3. \tag{6}$$

Taking $B \sim 5$ Gauss and $\rho \sim 10^{-18}$ gm cm^{-3}, then ϵ_f ranges from some 3 $\times$ 10^{30} to 3 $\times$ 10^{36} ergs per flare, as the flare time scale ranges from 100 to 10,000 seconds. For a flare time scale of one hour, equation (6) gives an energy of 1.3 $\times$ 10^{35} ergs. The total flare energy given by equation (6) is an especially sensitive function of the magnetic field strength: a 10 Gauss magnetic field would multiply all of the above energies by a factor of 32.

Now it is interesting to compare these results with the observations of flaring T Tauri stars. Such stars show diverse flaring activities over a range of time scales and intensities (Kuan 1976). Worden et al. (1981) suggest that 10-minute flares on T Tauri stars release at least 10^{34} ergs per event. This is in the range of flare energies given in the previous

paragraph. Although there is considerable uncertainty in the numbers and in the physical conditions, it is conceivable that at least some of the flares observed on T Tauri stars are the same phenomenon that we have described here as a possible energy source for chondrule melting.

In a possibly related development, Strelnitskij (1987) has interpreted observed linear-polarization rotation angles, in an H_2O maser around a "young star," to require the presence of an approximately 10 Gauss magnetic field at distances of 10 and more astronomical units from the central star. It is not clear whether this surprising result has any connection to the problems discussed here, but the observation is surely provocative in terms of our understanding of the environments of young stars and protostars.

DYNAMICAL EFFECTS OF THE MAGNETIC FIELD

A nebular magnetic field having the strength and distribution discussed in this paper would have had substantial effects on the structure and dynamical evolution of the system. The main effects would be of two kinds, deriving from pressure of the magnetic field and the ability of the field to transport angular momentum.

Consider that a 5 Gauss magnetic field exerts a pressure of just about 1 dyne/cm^2. Compare this with the nebular gas pressure, which, for the mass density $\rho \sim 10^{-9}$ gm/cm^2 and the gas temperature $T \sim 300K$, is about 10 dynes/cm^2. Thus the magnetic pressure is about 10% of the gas pressure. Although a 10% change in the effective gas pressure seems like a relatively small effect, within the context of the ideas discussed here, the overall effect of the magnetic field will be, in fact, much larger. The magnetic field constitutes a net expansive stress on the system, all of which must be confined in equilibrium by the gravity acting on the gas. This can be seen in a straightforward way from the magnetohydrodynamic virial theorem. To the extent that the low-mass-density corona also is permeated by a significant magnetic field, the expansive stress associated with the coronal fields must also be confined by the disk mass. Thus, to make a crude estimate, if the volume of coronal space filled with disk-generated magnetic field is, say, five times larger than the volume of the disk itself, suggested as a possibility in this discussion, then the effective expansive stress communicated to the disk gas is some five times larger. In that case the magnetic field becomes a major factor in the structure and dynamical balance of the disk, especially with respect to the vertical direction. Because the magnetic field acts much like a bouyant, zero-mass gas, the dynamical behavior of the gas and disk system would be expected to have similarities to what we observe in the solar photosphere-corona magnetic coupling. This is exactly the situation envisioned above in the speculative picture

of nebula-corona flares. In that case, the nebula would also be expected to exhibit behaviors similar to those described by Parker (1966) for the galactic disk.

The magnetic contribution to angular momentum transport could have similarly important effects with a magnetic field such as that considered in this paper. Consider the torque transmitted across a cylindrical surface aligned with disk axis and cutting the disk at a radius R:

$$T = \frac{<B_p B_\phi>}{4\pi} 2\pi R^2 (2\Lambda), \tag{7}$$

where we have included the torque between $z = \pm\Lambda$. Let τ_L be the time scale for angular momentum transport, then $\tau_L \sim L/T$, where L is a characteristic angular momentum of the system. Taking $L \sim \pi R^2(2\Lambda)\rho R^2\Omega$ we find that

$$\tau_L \sim \frac{2\pi\rho\sqrt{GMR}}{<B_p B_\phi>}. \tag{8}$$

Taking $\sqrt{<B_p B_\phi>} \sim$1-10 Gauss results in an evolutionary time scale for angular momentum transport of 10^2-10^4 years. Thus, the presence of such a nebular magnetic field would have a substantial impact on the angular momentum transport and on the radial evolution of the system.

This angular momentum transport rate is large in comparison with the time scales generally believed to characterize nebular evolution. In that respect, it is worth noting effects that could alter the simplest relationships between the 1-10 Gauss field strengths and the overall evolution time scale. First, we note that MHD dynamo modes in a disk are spatially localized (Stepinski and Levy 1988), so that such fast angular momentum transport may extend over only limited portions of the nebula at any one time. Second, detailed observations of the Sun show us that intense magnetic fields may be confined to thin flux ropes, with the intervening field strength being much weaker. Such a situation in the nebula, with a spatially intermittent magnetic field, might admit the most intense magnetic fields inferred from meteorite magnetization, while still producing an overall rate of angular momentum transfer much lower than that estimated above. Finally, angular momentum transport at the fast rate suggested in the previous paragraph might be expected to produce sporadic, temporally *intermittent* evolutionary behavior in the nebula over short time scales. One might imagine that the $10^2 - 10^4$ years magnetic timescale could represent rapidly fluctuating local *weather* episodes during the slower, long-term, large-scale evolution of the nebula.

THE PROBLEM OF IONIZATION

Perhaps the most difficult barrier to understanding the possible presence of a substantial magnetic field in the protoplanetary nebula is the question of electrical conductivity. Except near its very center, the nebula was a relatively high-density, dusty gas at relatively low temperatures. Under such conditions, the thermally induced ionization fraction and the electrical conductivity are very low. Significant levels of electrical conductivity require some nonthermal ionization source to produce mobile electrons. Consolmagno and Jokipii (1978) point out that ionization resulting from the decay of short-lived radioisotopes might have raised the electron fraction to the point at which the nebula gas was coupled to the magnetic field. Based on their preliminary analysis, Consolmagno and Jokipii suggested that an electron density of perhaps a few per cm^3 would have been produced with the then prevalent ideas about the abundance of ^{26}Al in the nebula. This is sufficient to produce the behaviors described above. Thus, although more complete calculations of nebular electrical conductivity are needed (and are underway) in light of new information about the cosmic abundance of ^{26}Al (Mahoney *et al.* 1984) and new information about the dominant ion reactions, it is at least possible that the nebula was a sufficiently good conductor of electricity to constitute a hydromagnetic system.

SUMMARY AND CONCLUSIONS

A coherent picture can be drawn of the possible magnetohydrodynamic character of the protoplanetary nebula. This picture is based on a mixture of evidence and speculation. From this picture emerges a reasonable explanation of meteorite magnetization, a possible source of transient energetic events to account for chondrule formation, a plausible picture of dynamo magnetic field generation and field strength in the nebula, and a possible connection to energetic outbursts observed in association with protostars. This picture has potentially significant implications for our understanding of the dynamical behavior and evolution of the protoplanetary nebula because a magnetic field having the implied strength and character discussed here would have exerted considerable stress on the system.

The primary unresolved question involves the electrical conductivity of the nebular gas. In order for the described picture to be real, the nebular gas must conduct electricity well enough to become a hydromagnetic fluid; this requires a nonthermal source of ionization. However, other questions also press at us. What is the real nature and genesis of meteorite magnetization? Are the present, simplest interpretations correct, or is something eluding us? Clearly important work remains to be done in this area. What were the natures and histories of meteorite parent bodies? Could a magnetizing

field have been internally generated? We are seriously in need of *in situ* investigation (with sample return) of comets and asteroids. What is the nature of protoplanetary environments? Astronomical studies are needed to ascertain the small-scale environments associated with star formation and protoplanetary disks.

From a broader point of view, it is possible that many things begin to fall into place if one presumes that the protoplanetary nebula did, in fact, have the characteristics described here. The protosolar system then takes on the aspect of a typical astrophysical system, which of course it was, with dynamical behaviors thought to be common in many such systems. In this case, it seems that a considerable conceptual gap separates the relatively simple and well-behaved nebula that emerges from our planetary system-based theoretical fantasies, and the energetic, violently active systems that we associate with protostars in the astrophysical sky. Some considerable work—theoretical, observational, and experimental—remains in order to close that gap.

ACKNOWLEDGEMENT

REFERENCES

Consolmagno, G.J., and J.R. Jokipii. 1978. ^{26}Al and the partial ionization of the solar nebula. The Moon and Planets 19:253.

Grossman, J.N. 1988. Formation of chondrules. Page 680. In: Kerridge, J.F., and M.S. Matthews (eds.). Meteorites and the Early Solar System. University of Arizona Press, Tucson.

King, E.A. 1983. (ed.). Chondrules and their Origins. Lunar and Planetary Institute, Houston.

Kuan, P. 1976. Photometric variations of T Tauri stars. Astrophysical Journal 210:129.

Lanoix, M., D.W. Strangway, and C.W. Pearce. 1978. The primordial magnetic field preserved in chondrules of the Allende meteorite. Geophysical Research Letters 5:73.

Levy, E.H. 1978. Magnetic field in the early solar system. Nature 276:481.

Levy, E.H. 1988. Energetics of chondrule formation. Page 697. In: Kerridge, J.F., and M.S. Matthews (eds.). Meteorites and the Early Solar System. University of Arizona Press, Tucson.

Levy, E.H., and S. Araki. 1988. Magnetic reconnection flares in the protoplanetary nebula and the possible origin of meteorite chondrules. Icarus, 81:74-91.

Levy, E.H., and C.P. Sonett. 1978. Meteorite magnetism and early solar system magnetic fields. Page 516. In: T. Gehrels (ed.). Protostars and Planets. University of Arizona Press, Tucson.

Mahoney, W.A., J.C. Ling, W.A. Wheaton, and A.S. Jacobson. 1984. HEAO-3 Discovery of ^{26}Al in the interstellar medium. Astrophys. J. 286:578-585.

Nagata, T. 1979. Meteorite magnetism and the early solar system magnetic field. Physics of the Earth and Planetary Interiors 20:324.

Nagata, T., and N. Suguira. 1977. Paleomagnetic field intensity derived from meteorite magnetization. Physics of the Earth and Planetary Interiors 13:373.

Parker, E.N. 1966. The dynamical state of the interstellar gas and field. Astrophysical Journal 145:811.

Parker, E.N. 1979. Cosmical Magnetic Fields. Their Origin and their Activity. Clarendon Press, Oxford.

Petschek, H.E. 1964. Magnetic field annihilation. Page 425-439. In: Hess, W.N. (ed.). NASA SP-50: AAS-NASA Symposium on the Physics of Solar Flares. NASA, Washington, D.C.

Safronov, V.S., and T.V. Ruzmaikina. 1985. Formation of the solar nebula and the planets. Page 959. In: Black, D.C., and M.S. Matthews (eds.). Protostars and Planets II. University of Arizona Press, Tucson.

Strelnitskij, V.S. 1987. On magnetic field strength in H_2O masers. Astronomical Circular 1490:4.

Stepinski, T., and E.H. Levy. 1988. Generation of dynamo magnetic fields in protoplanetary and other astrophysical accretion disks. Astrophysical Journal 331:416.

Suguira, N., and D.W. Strangway. 1983. A paleomagnetic conglomerate test using the Abee E4 meteorite. Earth and Planetary Science Letters 62:169.

Suguira, N., M. Lanoix, and D.W. Strangway. 1979. Magnetic fields of the solar nebula as recorded in chondrules from the Allende meteorite. Physics of the Earth and Planetary Interiors 20:342.

Suguira, N., and D.W. Strangway. 1985. NRM directions around a centimeter-sized dark inclusion in Allende. Proceedings of the 15th Lunar and Planetary Science Conference. Journal of Geophysical Research 90:C729.

Suguira, N., and D.W. Strangway. 1988. Magnetic studies of meteorites. Page 595. In: Kerridge, J.F. (ed.). Meteorites and the Early Solar System. University of Arizona Press, Tucson.

Wasilewski, P. 1987. Magnetic record in chondrite meteorites–microstructure, magnetism, and the FeNi phase diagram. Proceedings of the Meteoritical Society 50:183.

Worden, S.P., T.J. Schneeberger, J.R. Kuhn, and J.L. Africano. 1981. Flare activity on T Tauri stars. Astrophysical Journal 244:520.

Zel'dovich, Ya.B., and A.A. Ruzmaikin. 1987. The hydromagnetic dynamo as a source of planetary, solar, and galactic magnetism. Soviet Physics Uspekhi 30:494.

/

Formation of Planetesimals

STUART J. WEIDENSCHILLING
Planetary Science Institute

ABSTRACT

A widely accepted model for the formation of planetesimals is by gravitational instability of a dust layer in the central plane of the solar nebula. This mechanism does not eliminate the need for physical sticking of particles, despite published claims to that effect. Such a dust layer is extremely sensitive to turbulence, which would prevent gravitational instability unless coagulation forms bodies large enough to decouple from the gas (> meter-sized). Collisional accretion driven by differential motions due to gas drag may bypass gravitational instability completely. Previous models of coagulation assumed that aggregates were compact bodies with uniform density, but it is likely that early stages of grain coagulation produced fractal aggregates having densities that decreased with increasing size. Fractal structure, even if present only at sub-millimeter size, greatly slows the rate of coagulation due to differential settling and delays the concentration of solid matter to the central plane. Low-density aggregates also maintain higher opacity in the nebula than would result from compact particles. The size distribution of planetesimals and the time scale of their formation depend on poorly understood parameters, such as sticking mechanisms for individual grains, mechanical properties of aggregates, and the structure of the solar nebula.

INTRODUCTION

It is now generally accepted that the terrestrial planets and the cores of the giant planets were formed by accretion of smaller solid bodies

(planetesimals). The principal alternative, production of giant protoplanets by large-scale gravitational instabilities of the gaseous component of the solar nebula, has been abandoned (Cameron 1988). Planetesimals, with initial sizes of the order of kilometers, must have formed from much smaller particles, perhaps consisting of a mixture of surviving presolar grains and condensates from the nebular gas. Such grains were small, probably sub-μm in size, and their motions were controlled by the drag of the surrounding gas, rather than by gravitational forces. It is necessary to understand the aerodynamic processes that affected these small bodies in order to understand how planetesimals formed.

The most common assumption is that planetesimals resulted from localized gravitational instabilities within a dust layer in the central plane of the nebular disk. Such a layer is assumed to form by settling of grains through the quiescent gas due to the vertical component of solar gravity. If the dust layer becomes sufficiently thin, and thereby sufficiently dense, it is unstable with respect to density perturbations. This instability causes the layer to fragment into self-gravitating clumps. These eventually collapse into solid bodies, i.e., planetesimals. This process was described qualitatively as early as 1949 (Wetherill 1980). Quantitative expressions for the critical density and wavelength were derived by Safronov (1969) and independently by Goldreich and Ward (1973). The critical density is approximately the Roche density at the heliocentric distance a,

$$\delta_c \simeq 3M_\odot/2\pi a^3, \tag{1}$$

where $M_\odot$ is the solar mass. Perturbations grow in amplitude if they are smaller in size than a critical wavelength

$$\lambda_c \simeq 4\pi^2 G\sigma_s/\Omega^2, \tag{2}$$

where σ_s is the surface density of the dust layer, G the gravitational constant, and $\Omega = (GM_\odot/a^3)^{1/2}$ is the Kepler frequency. The characteristic mass of a condensation is $m_c \sim \sigma_s\lambda_c^2$, and depends only on the two parameters σ and a. The assumption that σ_s equals the heavy element content of a planet spread over a zone surrounding its orbit (Weidenschilling 1977a) implies $\sigma_s \sim 10\text{g cm}^{-2}$ and $m_c \sim 10^{18}\text{g}$ in the Earth's zone. Density perturbations grow rapidly, on the time scale of the orbital period, but condensations cannot collapse directly to solid bodies without first losing angular momentum. Goldreich and Ward (1973) estimated the contraction time to be $\sim 10^3$ years in Earth's zone, while Pechernikova and Vityazev (1988) estimate $\sim 10^5 - 10^6$ years.

The assumption that planetesimals formed in this manner has influenced the choice of starting conditions for numerical simulations of planetary accretion. Many workers (Greenberg *et al.* 1978; Nakagawa *et al.*

1983; Horedt 1985; Spaute *et al.* 1985; Wetherill and Stewart 1989) have assumed an initial swarm of roughly kilometer-sized bodies of uniform size or with a narrow size distribution. This assumption was also due to the lack of alternative models.

Another reason for popularity of the gravitational instability model was the belief that it requires no mechanism for physical sticking of grains, as was explicitly stated by Goldreich and Ward. Because the physical and chemical properties of grains in the solar nebula are poorly characterized, sticking mechanisms seem *ad hoc*. Still, there are reasons to believe that grains in the solar nebula did experience sticking. It is common experience in the laboratory (and Earth's atmosphere) that microscopic particles adhere on contact due to electrostatic or surface forces. Weidenschilling (1980) argued that for typical relative velocities due to settling in the solar nebula, van der Waals forces alone would allow aggregates to reach centimeter sizes. There are additional arguments that coagulation must have produced larger bodies, of meter size or larger, before gravitational instability could produce planetesimals. In order to understand these arguments, it is necessary to review the nature of the aerodynamic interactions between solid bodies and gas in the solar nebula.

NEBULAR STRUCTURE

We assume that the solar nebula is disk-shaped, has approximately Keplerian rotation, and is in hydrostatic equilibrium. If the mass of the disk is much less than the solar mass ($< 0.1 M_\odot$), then its self-gravity can be neglected compared with the vertical component of the Sun's attraction, $g_z = GM_\odot z/a^3 = \Omega^2 z$, where z is the distance from the central plane. The condition of hydrostatic equilibrium implies that the pressure at $z = 0$ is

$$P_c = \Omega \Sigma c/4, \tag{3}$$

where Σ is the surface density of the disk and c is the mean thermal velocity of the gas molecules. Equation (3) is strictly true only if the temperature is independent of z, but it is a good approximation even if the vertical structure is adiabatic. It can be shown that

$$P(z) = P_c \exp\left(-z^2/H^2\right), \tag{4}$$

where $H = \pi c/2\Omega$ is the characteristic half-thickness or scale height.

There is also a radial pressure gradient in the disk. It is plausible to assume that the temperature and density decrease with increasing heliocentric distance. Some accretion disk models of the nebula have Σ approximately constant, but equation (3) shows that even these will have a strong pressure gradient, because Ω is proportional to $a^{-3/2}$ for Keplerian

motion. The decrease in pressure is due primarily to the weakening of solar gravity at larger distances. Because the gas is partially supported by the pressure gradient in addition to the rotation of the disk, hydrostatic equilibrium requires that its velocity be less than Keplerian;

$$V_g^2 = V_k^2 + \frac{a}{\rho}\frac{\partial P}{\partial a}, \tag{5}$$

where ρ is the gas density. We define $\Delta V = V_k - V_g$ as the difference between the Kepler velocity and the gas velocity. One can show (Weidenschilling, 1977b) that if one assumes a power law for the pressure gradient, so that $P \propto a^{-n}$, then the fractional deviation of the gas from V_k is

$$\frac{\Delta V}{V_k} = \frac{nRT/\mu}{2GM_\odot/a}, \tag{6}$$

where R is the gas constant and μ the molecular weight. We see that $\Delta V/V_k$ is approximately the ratio of thermal and gravitational potential energies of the gas. For most nebular models, this quantity is only a few times 10^{-3}, but this is enough to have a significant effect on the dynamics of solid bodies embedded in the gas.

AERODYNAMICS OF THE SOLID BODIES IN THE NEBULA

The effects of gas on the motions of solid bodies in the solar nebula have been described in detail by Adachi *et al.* (1976) and Weidenschilling (1977b); here we summarize the most important points. The fundamental parameter that characterizes a particle is its response time to drag,

$$t_e = mV/F_D, \tag{7}$$

where m is the particle mass, V its velocity relative to the gas, and F_D is the drag force. The functional form of F_D depends on the Knudsen number (ratio of mean free path of gas molecules to particle radius) and Reynolds number (ratio of inertial to viscous forces). In the solar nebula, the mean free path is typically a few centimeters, so dust particles are in the free-molecular regime. In that case, a spherical particle of radius s, bulk density ρ_s, has

$$t_e = s\rho_s/\rho c. \tag{8}$$

The dynamical behavior of a particle depends on the ratio of t_e to its orbital period, or, more precisely, to the inverse of the Kepler frequency. A "small" particle, for which $\Omega t_e \ll 1$, is coupled to the gas, i.e., the drag force dominates over solar gravity. It tends to move at the angular velocity

of the gas. The residual radial component of solar gravity causes inward radial drift at a terminal velocity given by

$$V_r = -2\Omega \Delta V t_e. \tag{9}$$

Similarly, there is a drift velocity due to the vertical component of solar gravity,

$$V_z = -\Omega^2 z t_e. \tag{10}$$

For a "large" body with $\Omega t_e \gg 1$ gas drag is small compared with solar gravity. Such a body pursues a Kepler orbit. Because the gas moves more slowly, the body experiences a "headwind" of velocity ΔV. The drag force causes a gradual decay of its orbit at a rate

$$Vr = da/dt = -2\Delta V/\Omega t_e. \tag{11}$$

For plausible nebular parameters, and particle densities of a few g cm^{-3}, the transition between "small" and "large" regimes occurs at sizes of the order of one meter. The peak radial velocity is equal to ΔV when $\Omega t_e = 1$. The dependence of a radial and transverse velocities on particle size are shown for a typical case in Figure 1.

PROBLEMS WITH GRAVITATIONAL INSTABILITY

Could the gravitational instability mechanism completely eliminate the need for particle coagulation? We first consider the case in which there is no turbulence in the gas. The time scale for a particle to settle toward the central plane of the nebula is $\tau_z = z/V_z$. From equations (8) and (10),

$$\tau_z = \rho c/s \rho_s \Omega^2. \tag{12}$$

If we take $\rho \sim 10^{-10}$g cm^{-3}, $\tau_z \sim (10^2/s)$ years, where s is in cm, so if s $= 1$ μm, $\tau_z \sim 10^6$ years. This is merely the e-folding time for z to decrease. For an overall solids/gas mass ratio f $= 3 \times 10^{-3}$, corresponding to the cosmic abundance of metal plus silicates, and an initial scale height H $\sim c/\Omega$, the dust layer requires ~ 10 $\tau_z \sim 10^7$ years to become thin enough to be gravitationally unstable. This exceeds the probable lifetime of a circumstellar disk. Undifferentiated meteorites and interplanetary dust particles include sub-μm sized components. These presumably had to be incorporated into their parent bodies not as separate grains, but as larger aggregates, which settled more rapidly.

In addition to the problem of settling time scale, there is another argument for growth of particles by sticking. A very slight amount of turbulence in the gas would suffice to prevent gravitational instability.
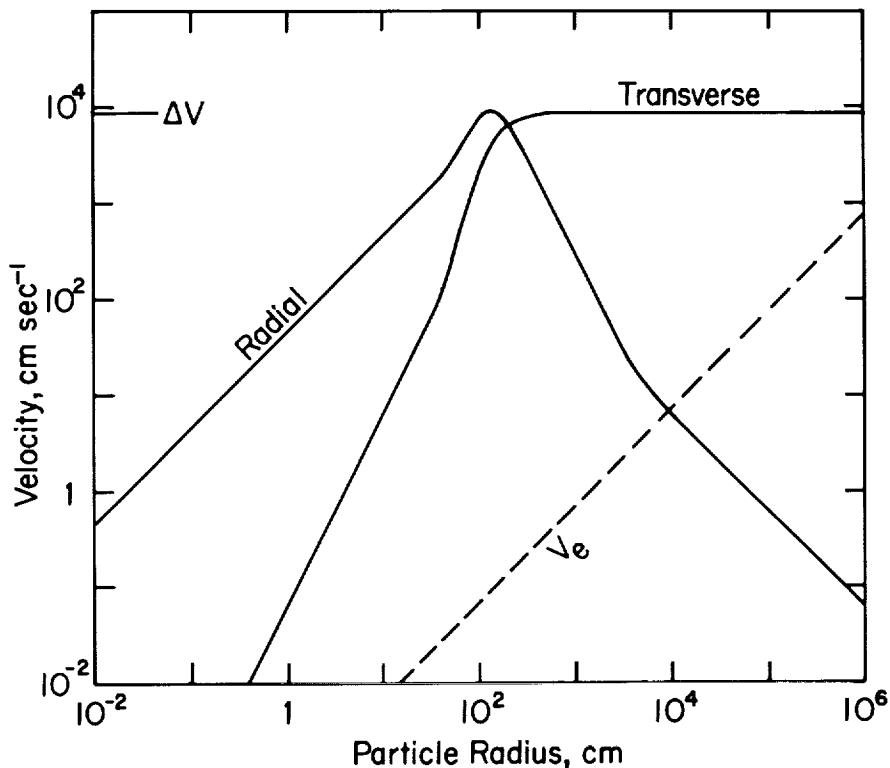
FIGURE 1 Radial and transverse velocities relative to the surrounding gas for a spherical body with density 1 g/cm^3. Also shown is the escape velocity from the body's surface, V_e. Numerical values are for the asteroid zone (Weidenschilling 1988) but behavior is similar for other parts of the nebula.

Particles respond to turbulent eddies that have lifetimes longer than $\sim t_e$. The largest eddies in a rotating system generally have timescales $\sim 1/\Omega$, so bodies that are "small" in the dynamical sense of $\Omega t_e < 1$, or less than about a meter in size, are coupled to turbulence. A particle would tend to settle toward the central plane until systematic settling velocity is of the same order as the turbulent velocity, V_t. From this condition, we can estimate the turbulent velocity that allows the dust layer to reach a particular density (Weidenschilling 1988). In order to reach a dust/gas ratio of unity

$$V_t \sim f\Omega s \rho_s/\rho. \qquad (13)$$

Typical parameters give $V_t \sim (s\rho_s)$ cm sec^{-1} when s and ρ_s are in cgs units, i.e., if particles have density of order unity, then the turbulent velocity of the gas must be no greater than about one particle diameter per second in order

for the space density of solids to exceed that of the gas. In order to reach the critical density for gravitational instability, the dust density must exceed that of the gas by about two orders of magnitude, with correspondingly smaller V_t, $\sim 10^{-2}$ particle diameters per second. For μm-sized grains, this would imply turbulent velocities of the order of a few meters *per year*, which seems implausible for any nebular model.

Even if the nebula as a whole was perfectly laminar, formation of a dense layer of particles would create turbulence. If the solids/gas ratio exceeds unity, and the particles are strongly coupled to the gas by drag forces, then the layer behaves as a unit, with gas and dust tending to move at the local Kepler velocity. There is then a velocity difference of magnitude ΔV between the dust layer and the gas on either side. Goldreich and Ward (1973) showed that density stratification in the region of shear would not suffice to stabilize it, and this boundary layer would be turbulent. Weidenschilling (1980) applied a similar analysis to the dust layer itself, and showed that the shear would make it turbulent as well. Empirical data on turbulence within boundary layers suggest that the eddy velocities within the dust layer would be a few percent of the shear velocity ΔV, or several meters per second. The preceding analysis argues that gravitational instability would be possible only if the effective particle size were of the order of a meter or larger. Bodies of this size must form by coagulation of the initial population of small dust grains.

PARTICLE GROWTH BY COAGULATION

If it is assumed that particles stick upon contact, then their rate of growth can be calculated. The rate of mass gain is proportional to the product of the number of particles per unit volume, their masses and relative velocities, and some collisional cross-section. We assume that the latter is simply the geometric cross-section, $\pi(s_1 + s_2)^2$ (although electrostatic or aerodynamic effects may alter this in some cases). For small particles (less than a few tens of μm), thermal motion dominates their relative velocities. The mean thermal velocity is $\bar{v} = (3kT/m)^{1/2}$, where T is the temperature and k is Boltzmann's constant. If we assume that all particles have the same radius s (a reasonable approximation, as thermal coagulation tends to produce a narrow size distribution), the number of particles per unit volume is $N = 3f\rho/4\pi\rho_s s^3$. The mean particle size increases with time as

$$s(t) = s_o + \left[15 f\rho \ (kT/8\pi p_s^3)^{1/2} t\right]^{2/5} \tag{14}$$

As a particle grows its thermal motion decreases, while its settling rate increases. When settling dominates, a particle may grow by sweeping up

smaller ones (particles of the same size, or t_e, have the same settling rate, and hence do not collide). If it is much larger than its neighbors, then the relative velocity is approximately the settling rate of the larger particle. It grows at the rate

$$\frac{ds}{dt} = f\rho V_z/4\rho_s = f\Omega^2 zs/4c, \tag{15}$$

giving

$$s(t) = s_o \exp\,(f\Omega^2 zt/4c). \tag{16}$$

A particle growing by this mechanism increases in size exponentially on a time scale $\tau_g = 4c/f\Omega^2 z$, or $\sim 4/f\Omega$ at $z \sim c/\Omega$. This time scale is a few hundred years at $a = 1$ AU, and is independent of the particle density or the gas density. The lack of dependence on particle density is due to the fact that a denser particle settles faster, but has a smaller cross-section, and can sweep up fewer grains; the two effects exactly cancel one another. Likewise, if the gas density is increased, the settling velocity decreases, but the number of accretable particles increases, provided f is constant. The time scale increases with heliocentric distance; $\tau_g \propto a^{3/2}$. A particle settling vertically from an initial height z_o and sweeping up all grains that it encounters can grow to a size $s(max) = f\rho z_o/4\pi_s$, typically a few cm for z_o $\sim c/\Omega$. Actually, vertical settling is accompanied by radial drift, so growth to larger sizes is possible.

The growth rates and settling rates mentioned above have been used (with considerable elaboration) to construct numerical models of particle evolution in a laminar nebula (Weidenschilling 1980; Nakagawa *et al.* 1981). The disk is divided into a series of discrete levels. In each level the rate of collisions between particles of different sizes is evaluated, and the changes in the size distribution during a timestep Δt is computed. Then particles are distributed to the next lower level at rates proportional to their settling velocities. A typical result of those simulations shows particle growth dominated by differential settling. Because the growth rate increases with z, large particles form in the higher levels first, and "rain out" toward the central plane through the lower levels. The size distribution remains broad, with the largest particles much larger than the mean size, justifying the assumptions used in deriving equation (16). After $\sim 10\ \tau_g$, typically a few thousand orbital periods, the largest bodies exceed one meter in size and the solids/gas ratio in the central plane exceeds unity. The settling is non-homologous, with a thin dense layer of large bodies containing $\sim$ 1-10% of the total surface density of solids, and the rest in the form of small aggregates distributed through the thickness of the disk.

The further evolution of such a model population has not yet been calculated. The main difficulty is the change in the nature of the interaction

between particles and gas when the solids/gas ratio exceeds unity. As mentioned previously, the particle layer begins to drag the gas with it. The relative velocities due to drag no longer depend only on t_e, as in equations (7-11), but on the local concentration of solids: there is also shear between different levels. Work is presently under way to account for these effects, at least approximately. A complete treatment may require the use of large computers using computational fluid dynamics codes.

PROPERTIES OF FRACTAL AGGREGATES

The earlier numerical simulations mentioned above assumed that all particles are spherical and have the same density, regardless of size. This is a good assumption for coagulation of the liquid drops, but it does not apply to solid particles. When two grains stick together, they retain their identities, and the combined particle is not spherical. Aggregates containing many grains have porous, fluffy structures. It has been shown (Mandelbrot 1982; Meakin 1984) that such aggregates have fractal structures. A characteristic of fractal aggregates is that the average number of particles found within a distance s of any arbitrary point inside the aggregate varies as $n(s) \propto s^D$, where D is the fractal dimension. The density varies with size according to the relation $\rho \propto s^{D-3}$. "Normal" objects have $D = 3$ and uniform density. Aggregates of particles generally have $D \simeq 2$, so that their density varies approximately inversely with size, i.e., they become more porous as they grow larger. For the most simple aggregation processes (Jullien and Botet 1986; Meakin 1988a,b) the fractal dimension lies in the range $1.7 < D < 2.2$, but more complex mechanisms can lead to values of D lying outside of this range. A hierarchical ballistic accretion in which clusters of similar size stick at their point of contact yields $D < 2$. Building up a cluster of successive accretion of single grains or small groups, or allowing compaction after contact, leads to $D > 2$. An example of this type is shown in Figure 2. This is a computer-generated model, but it corresponds closely to soot particles observed in the laboratory (Meakin and Donn 1988).

The structure of aggregates, which are very unlike uniform-density spheres, affects their aerodynamic behavior in the solar nebula. For a compact sphere, the response time is proportional to the size (equation 8). For a fractal aggregate, t_e increases much more slowly with size. Meakin has developed computer modeling procedures to determine the mean projected area of an aggregate, as viewed from a randomly selected direction (Meakin and Donn 1988; Meakin, unpublished). The variation of the projected area with the number of grains in the aggregate depends on the fractal dimension. If $D < 2$, aggregates become more open in structure at larger sizes, and are asymptotically "transparent," i.e., the ratio of mass to projected area never exceeds a certain limit. For $D > 2$, large aggregates

BALLISTIC CI-CI-3d
ONE STAGE RESTRUCTURING
S =11186

94 DIAMETERS
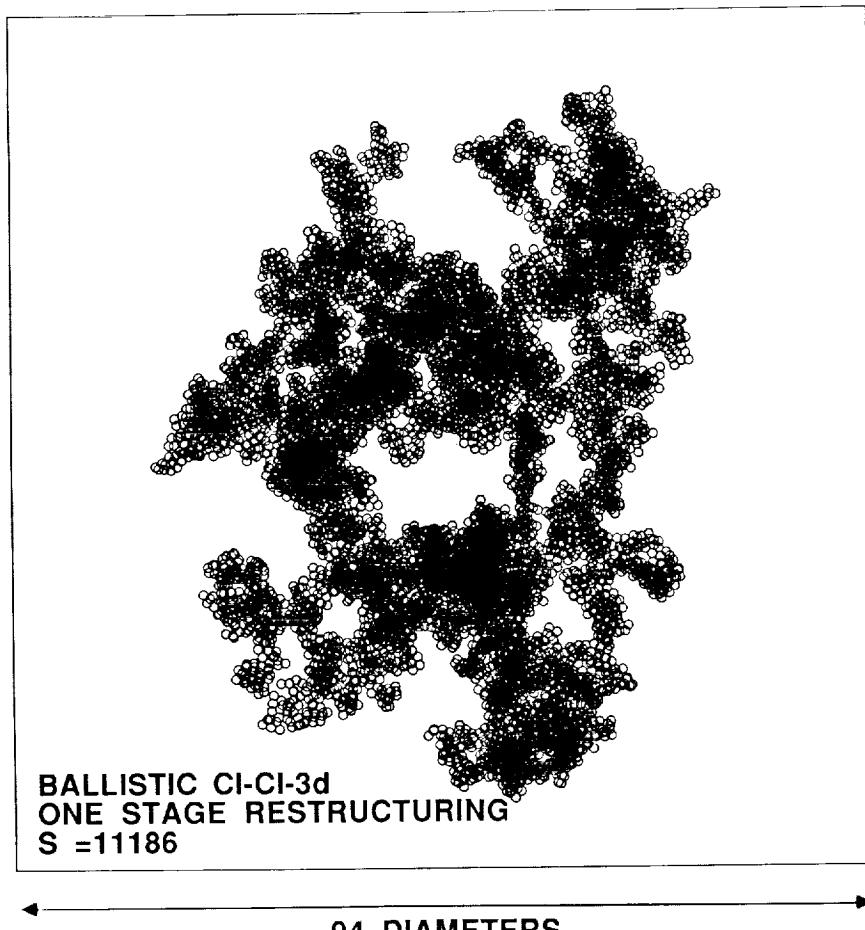
FIGURE 2 View of a computer-generated aggregate of fractal dimension $\simeq$ 2.11 and containing $\simeq$ 10^4 individual grains.

are opaque; the mass/area ratio increases without limit, although more slowly than for a uniform density object.

We assume that in the free molecular regime, when aggregates are smaller than the mean free path of a gas molecule (> cm in typical nebular models), t_e is proportional to the mass per unit projected area (m/A). From the case of a spherical particle in this regime, as in equation (8), we infer

$$t_e = 3(m/A)/4\rho c. \tag{17}$$

We can express t_e for aggregates conveniently in terms of the value
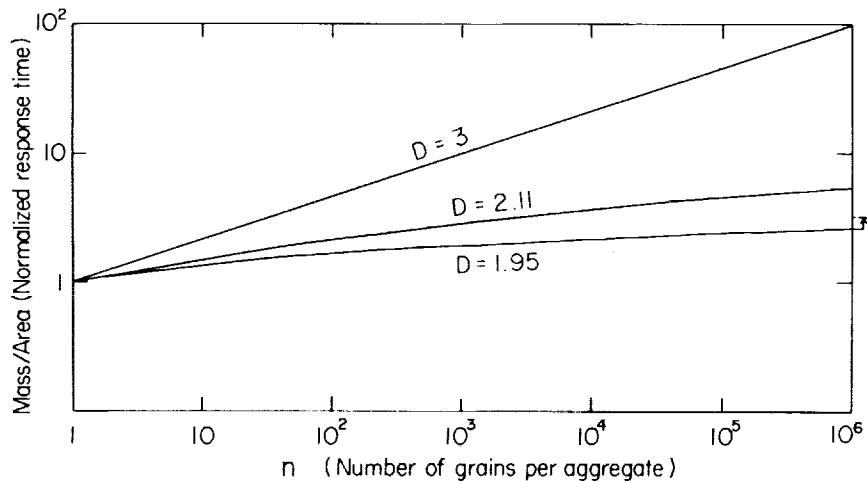
FIGURE 3 Mass/area ratio vs. number of grains for aggregates of different dimensions. $D = 3$ assumes coagulation produces a spherical body with the density of the separate components (liquid drop coalescence). For $D > 2$, m/A $\propto$ $m^{1-2/D}$, or $\propto$ $m^{1/3}$ for $D = 3$, and $\propto$ $m^{0.052}$ for $D = 2.11$. For $D < 2$, m/A approaches an asymptotic value, shown by the arrow.

for an individual grain, of size s_o and density ρ_o, $t_{eo} = s_o\rho_o/\rho c$. Fits to Meakin's data give

$$t_e = t_{eo}(0.343n^{-0.052} + 0.684n^{-0.262}) \qquad (18)$$

for $D = 2.11$, where n is the number of grains in the aggregate. This relation is plotted in Figure 3. For large aggregates of 10^6 grains, the mass/area ratio and t_e are $\sim$ 20-30 times smaller than for a compact particle of equal mass.

Fractal properties of aggregates also have implications for the opacity of the solar nebula. The dominant source of opacity is solid grains (Pollack *et al.* 1985). The usual assumption in computing opacity is that particles are spherical (Weidenschilling 1984). Their shapes are unimportant if they are smaller than the wavelength considered (Rayleigh limit). However, when sizes are comparable to the wavelength, the use of Mie scattering theory for spherical particles is inappropriate. In the limit of geometrical optics when particles are large compared with the wavelength, the opacity varies inversely with the mass/area ratio. From Figure 3 we see that if aggregates of 10^6 grains are in the geometrical optics regime, then the opacity is $\sim$ 20-30 times greater than for compact spherical particles of equal mass. Actually, if individual grains are below the Rayleigh limit, aggregates may not be in the geometrical optics regime, even if they are larger than the

wavelength. The optical properties of fractal aggregates need further study, but it is apparent that coagulation of grains is less effective for lowering the nebula's opacity than has been generally assumed.

COAGULATION AND SETTLING OF FRACTAL AGGREGATES

We have modeled numerically the evolution of a population of particles in the solar nebula with fractal dimension of 2.11, using the response time of equation (18). The modeling program is based on that of Weidenschilling (1980). The calculations assumed a laminar nebula with surface density of gas $3 \times 10^3 \text{g cm}^{-2}$, surface density of solids 10 g cm^{-2}, and temperature of 500K at a heliocentric distance of 1 AU. At $t = 0$ the dust was in the form of individual grains of diameter $1 \mu m$, uniformly mixed with the gas. With the assumption that coagulation produced spherical particles of dimension $D = 3$, or constant density (the actual value is unimportant; compare the discussion of equation (16)), "raining out" with growth of approximately 10-meter bodies in the central plane occurs in a few times 10^3 years.

For the case of fractal aggregates with $D = 2.11$, we assumed that an individual μm-sized grain ($s_o = 0.5 \mu m$) has a density $\rho_o = 3 \text{ g cm}^{-3}$. Aggregates of such grains have densities that decrease with size according to

$$\rho_s = \rho_o (s/s_o)^{D-3} \qquad (19)$$

until $s = 0.5$ mm, at which size $\rho_s \simeq 0.01 \text{ g cm}^{-3}$ (densities of this order are achieved by some aggregates under terrestrial conditions; Donn and Meakin 1988). The density is assumed constant at this value until $s = 1$ cm, and then increases approximately as s^2 to a final density of 2 g cm^{-3} for $s \geq 10$ cm. This variation is arbitrary, but reflects the plausible assumption that fractal structure eventually gives way to uniform density for sufficiently large bodies due to collisional compaction. In a laminar nebula, relative velocities may be low enough to allow fractal structure at larger sizes than assumed here, so this assumption may be conservative. Experimental data on the mechanical behavior of fractal aggregates are sorely needed.

Even the limited range of fractal behavior assumed here has a strong effect on the evolution of the particles in the nebula. Growth and settling are slowed greatly. After a model time of 2×10^4 years, the largest aggregates are < one millimeter in size. The highest levels of the disk, above one scale height, are slightly depleted in solids due to the assumption that the gas is laminar. The largest aggregates at this time have settling velocities ~ 1 cm sec^{-1} , so there would be no concentration toward the central plane if turbulence in the gas exceeded this value. Continuation of this calculation results in more rapid growth by differential settling beginning at about 2.5
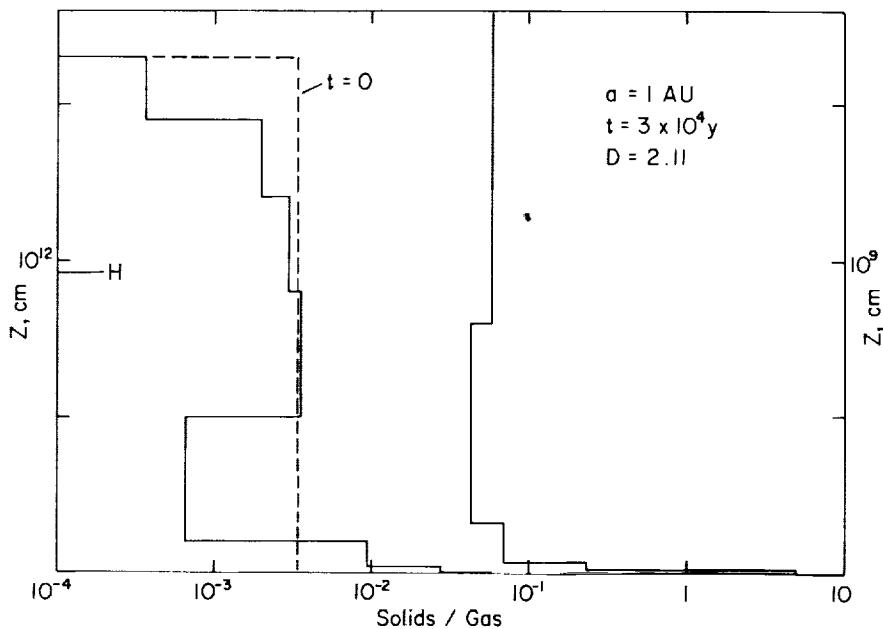
FIGURE 4 Outcome of a numerical simulation of coagulation and settling in the solar nebula at a = 1 AU. Aggregates are assumed to have fractal dimension 2.11 at sizes < 0.1 cm. AT t = 0, dust/gas ratio is uniform at 0.0034, corresponding to cosmic abundance of metal and silicates. At t = 3 × 10⁴y, dust/gas exceeds unity in a narrow region near the central plane (expanded scale at right).

× 10⁴ years. By 3 × 10⁴ years there is a high concentration of solids in a narrow zone at the central plane of the disk (see Figure 4). This layer contains approximately 1% of the total mass of solids, in the form of bodies several meters in diameter.

Evidently, the fractal nature of small aggregates greatly prolongs the stage of well-mixed gas and dust, before "rainout" to the central plane. The reason for this behavior is subtle. If we assume that the density varies as equation (19), then a generalization of the thermal coagulation growth rate of equation (14) gives

$$s(t) = s_o + \left[(3D/2 - 2)3f\rho(kT/2\pi\rho_0^3)^{1/2}s_o^{3(D-3)/2}t\right]^{1/(3D/2-2)} \tag{20}$$

For D = 2.11, this gives s ∝ t^{0.86}, vs. s ∝ t^{0.4} for D = 3. Thus, particle sizes increase *more* rapidly for thermal coagulation of fractal aggregates, due to their larger collisional cross-sections. For aggregates with D > 2, the behavior of t_e at large sizes is $t_e \simeq t_{eo}(s/s_o)^{D-2}$. Using this relation

in equation (15), D appears in both the numerator and denominator, leaving the growth rate unchanged. If thermal coagulation is faster for fractal aggregates and the growth rate due to settling is independent of D, then why is the evolution of the particle population slower? The answer is that the derivation of equation (15) assumes that the mass available to the larger aggregate is in much smaller particles, so that the relative velocity is essentially equal to the larger body's settling rate. Thermal coagulation tends to deplete the smallest particles most rapidly, creating a narrowly peaked size distribution, and rendering differential settling ineffective. Inserting the parameters used in our simulation into equation (20) predicts $s \simeq 0.1$ cm at $t = 2 \times 10^4$ years, in good agreement with the numerical results.

The qualitative behavior of this simulation, a long period of slow coagulation followed by rapid growth and "rainout," is to some degree an artifact of our assumption for the variation of particle density with size. A transition from fractal behavior to compact bodies without abrupt changes in slope would probably yield a more gradual onset of settling. It is likely that fractal structure could persist to sizes larger than the one millimeter we have assumed here, with correspondingly longer evolution time scales.

We have not yet modeled the evolution of the particle population after the solids/gas ratio exceeds unity in the central plane and can only speculate about possible outcomes. At the end of our simulations most of the mass in this level is in bodies approximately 10 meters in size. These should undergo further collisional growth, and in the absence of turbulence will settle into an extremely thin layer. Because these bodies are large enough to be nearly decoupled from the gas, it is conceivable that gravitational instability could occur in this layer. However, the surface density represented by these bodies is small, approximately 1% of the total surface density of solids. The critical wavelength, as in equation (2), is correspondingly smaller. If they grow to sizes >0.1 km before the critical density is reached, then their mean spacing is $\sim \lambda_c$, and gravitational instability is bypassed completely. In any case, the first-formed planetesimals would continue to accrete the smaller bodies that rain down to the central plane over a much longer time.

CONCLUSIONS

The settling of small particles to the central plane of the solar nebula is very sensitive to the presence of turbulence in the gaseous disk. It appears that a particle layer sufficiently dense to become gravitationally unstable cannot form unless the particles are large enough to decouple from the gas, i.e. > meter-sized. Thus, the simple model of formation of planetesimals directly from dust grains is not realistic; there must be an intermediate stage of particle coagulation into macroscopic aggregate bodies.

Coagulation of grains during settling alters the nature of the settling process. The central layer of particles forms by the "raining out" of larger aggregates that contain only a fraction of the total mass of solids. The surface density of this layer varies with time in a manner that depends on the nebular structure and particle properties. Thus, if planetesimals form by gravitational instability of the particle layer, the scale of instabilities and masses of planetesimals are not simply related to the total surface density of the nebula. It is possible that collisional coagulation due to drag-induced differential motions may be sufficiently rapid to prevent gravitational instability from occurring.

It is probable that dust aggregates in the solar nebula were low-density fractal structures. The time scale for settling to the central plane may have been one or two orders of magnitude greater than estimates which assumed compact particles. The inefficiency of settling by "raining out" suggests that a significant fraction of solids remained suspended in the form of small particles until the gas was dispersed; the solar nebula probably remained highly opaque. The mass of the nebula may have been greater than the value represented by the present masses of the planets.

ACKNOWLEDGEMENTS

REFERENCES

Adachi, I., C. Hayashi, and K. Nakazawa. 1976. Prog. Theor. Phys. 56:1756.

Cameron, A.G.W. 1988. Ann. Rev. Astron. Astrophys. 26:441.

Donn, B., and P. Meakin. 1989. Pages 577-580. Proc. Lunar Planet. Sci. Conf. 19th.

Goldreich, P., and W.R. Ward. 1973. Astrophys. J. 183:1051.

Greenberg, R., J. Wacker, W.K. Hartmann, and C.R. Chapman. 1978. Icarus 35:1.

Horedt, G.P. 1985. Icarus 64:448.

Jullien, R., and R. Botet. 1986. Aggregation and Fractal Aggregates, World Scientific, Singapore.

Mandelbrot, B.B. 1982. The Fractal Geometry of Nature. W.H. Freeman and Company, New York.

Meakin, P. 1984. Phys. Rev. A29:997.

Meakin, P. 1988a. Page 335. In: Domb, C., and J.L. Lebowitz (eds.). Phase Transitions and Critical Phenomena, vol. 12. Academic Press, New York.

Meakin, P. 1988b. Adv. Colloid and Interface Sci. 28:249.

Meakin, P., and B. Donn. 1988. Astrophys. J. 329:L39.

Nakagawa, Y., K. Nakazawa, and C. Hayashi. 1981. Icarus 45:517.

Nakagawa, Y., C. Hayashi, and K. Nakazawa. 1983. Icarus 54:361.

Pechernikova, G.V., and A.F. Vityazev. 1988. Astron. Zh. 65:58.

Pollack, J., C. McKay, and B. Christofferson. 1985. Icarus 64: 471.

Safronov, V.S. 1969. Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets. Nauka, Moscow (also NASA TTF-677).

Spaute, D., B. Lago, and A. Cazenave. 1985. Icarus 64:139.

Weidenschilling, S.J. 1977a. Astrophys. Space Sci. 51:153.

Weidenschilling, S.J. 1977b. Mon. Not. Roy. Astron. Soc. 180: 57.

Weidenschilling, S.J. 1980. Icarus 44:172.

Weidenschilling, S.J. 1984. Icarus 60:553.

Weidenschilling, S.J. 1988. Pages 348-371. In: Kerridge, J., and M. Matthews, (eds.). Meteorites and the Early Solar System. University of Arizona Press, Tucson.

Wetherill, G.W. 1980. Ann. Rev. Astron. Astrophys. 18:77.

Wetherill, G.W., and G. Stewart. 1989. Icarus. 77:330.

Formation of the Terrestrial Planets from Planetesimals

GEORGE W. WETHERILL
Carnegie Institution of Washington

ABSTRACT

Previous work on the formation of the terrestrial planets (e.g. Safronov 1969; Nakagawa *et al.* 1983; Wetherill 1980) involved a stage in which on a time scale of $\sim 10^6$ years, about 1000 embryos of approximately uniform size ($\sim 10^{25}$ g) formed, and then merged on a $10^7 - 10^8$ year time scale to form the final planets. Numerical simulations of this final merger showed that this stage of accumulation was marked by giant impacts ($10^{27} - 10^{28}$ g) that could be responsible for providing the angular momentum of the Earth-Moon system, removal of Mercury's silicate mantle, and the removal of primordial planetary atmospheres (Hartmann and Davis 1975; Cameron and Ward 1976; Wetherill 1985). Requirements of conservation of angular momentum, energy, and mass required that these embryos be confined to a narrow zone between about 0.7 and 1.0 AU. Failure of embryos to form at $1.5 - 2.0$ AU could be attributed to the longer ($\sim 10^7$ years) time scale for their initial stage of growth and the opportunity of effects associated with the growth of the giant planets to forestall that growth.

More recent work (Wetherill and Stewart 1988) indicates that the first stage of growth of embryos at 1 AU occurs by a rapid runaway on a much shorter $\sim 3 \times 10^4$ year time scale, as a consequence of dynamical friction, whereby equipartition of energy lowers the random velocities and thus increases the gravitational cross-section of the larger bodies. Formation of embryos at ~ 2 AU would occur in $< 10^6$ years, and it is more difficult to understand how their growth could be truncated by events in the outer solar system alone. Those physical processes included in this earlier work are not capable of removing the necessary mass, energy, and angular momentum

from the region between the Earth and the asteroid belt, at least on such a short time scale.

An investigation has been made of augmentation of outer solar system effects by spiral density waves produced by terrestrial planet embryos in the presence of nebular gas, as discussed by Ward (1986). This can cause removal of angular momentum and mass from the inner solar system. The theoretical numerical coefficients associated with the radial migration and eccentricity damping caused by this effect are at present uncertain. It is found that large values of these coefficients, compression of the planetesimal swarm by density wave drag, followed by resonance effects following the formation of Jupiter and Saturn, "clears" the region between Earth and the asteroid belt, and also leads to the formation of Earth and Venus with approximately their observed sizes and heliocentric distances. For smaller, and probably more plausible values of the coefficients, this mechanism will not solve the angular-momentum-energy problem. The final growth of the Earth on a $\sim 10^8$ year time scale is punctuated by giant impacts, up to twice the mass of Mars. Smaller bodies similar to Mercury and the Moon are vulnerable to collisional fragmentation. Other possibly important physical phenomena, such as gravitational resonances between the terrestrial planet embryos have not yet been considered.

INTRODUCTION

This article will describe recent and current development of theories in which the terrestrial planets formed by the accumulation of much smaller (one- to 10-kilometer diameter) planetesimals. The alternative of forming these planets from massive gaseous instabilities in the solar nebula has not received much attention during the past decade, has been discussed by Cameron *et al.* (1982), and will not be reviewed here.

In its qualitative form, the planetesimal, or "meteoric" theory of planet formation dates back at least to Chladni (1794) and was supported by numerous subsequent workers, among the most prominent of which were Chamberlain and Moulton (Chamberlain 1904). Its modern development into a quantitative theory began with the work of O.Yu. Schmidt and his followers, most notably V.S. Safronov. The publication in 1969 of his book "Evolutionary of the Protoplanetary Swarm" (Safronov 1969) and its publication in English translation in 1972 were milestones in the development of this subject, and most work since that time has consisted of extension of problems posed in that work.

The formation of the terrestrial planets from planetesimals can be conveniently divided into three stages:

(1) The formation of the planetesimals themselves from the dust of the solar nebula. The current status of this difficult question has been reviewed by Weidenschilling *et al.* (1988).

(2) The local accumulation of these one- to 10-kilometer planetesimals into $\sim 10^{25} - 10^{26}$g "planetary embryos" revolving about the sun in orbits of low eccentricity and inclination. Recent work on this problem has been summarized by Wetherill (1989a), and will be briefly reviewed in this article.

(3) The final merger of these embryos into the planets observed today. Fairly recent discussions of this stage of accumulation have been given by Wetherill (1986, 1988). This work needs to be updated in order to be consistent with progress in our understanding of stage (2). Particular attention will be given to that need in the present article.

FORMATION OF THE ORIGINAL PLANETESIMALS

The original solid material in the solar nebula was most likely concentrated in the micron size range, either as relic interstellar dust grains, as condensates from a cooling solar nebula, or a mixture of these types of material. The fundamental problem with the growth of larger bodies from such dust grains is their fragility with regard to collisional fragmentation, not only at the approximate kilometers per second sound speed velocities of a turbulent gaseous nebula, but even at the more modest ~ 60 m/sec differential velocities associated with the difference between the gas velocity and the Keplerian velocity of a non-turbulent nebula (Whipple 1973; Adachi *et al.* 1976; Weidenschilling 1977). Agglomeration under these conditions requires processes such as physical "stickiness," the imbedding of high-velocity projectiles into porous targets, or physical coherence of splash products following impact. Despite serious efforts to experimentally or theoretically treat this stage of planetary growth, our poor understanding of physical conditions in the solar nebula and other physical properties of these primordial aggregates make it very difficult.

Because of these difficulties, many workers have been attracted to the possibility that growth of bodies to one- to 10-kilometer diameters could be accomplished by gravitational instabilities in a central dust layer of the solar nebula (Edgeworth 1949; Safronov 1960; Goldreich and Ward 1973). Once bodies reach that size, it is plausible that their subsequent growth would be dominated by their gravitational interactions. Weidenschilling (1984) however has pointed out serious difficulties that are likely to preclude the development of the necessary high concentration and low relative velocity (approximately 10 centimeters per second) in a central dust layer. Therefore the question of how the earliest stage of planetesimal growth took place remains an open one that requires close attention.

GROWTH OF PLANETESIMALS INTO PLANETARY EMBRYOS

If somehow the primordial dust grains can agglomerate into one- to 10-kilometer diameter planetesimals, it is then necessary to understand the processes that govern their accumulation into larger bodies.

The present mass of the terrestrial planets is $\sim 10^{28}$g, therefore about 10^{10} 10km ($\sim 10^{18}$g) bodies are required for their formation. It is completely out of the question to consider the gravitationally controlled orbital evolution of such a large swarm of bodies by either the conventional methods of celestial mechanics, or by Monte Carlo approximations to these methods. Therefore all workers have in one way or another treated this second stage of planetary growth by methods based on gas dynamics, particularly by the molecular theory of gases, in which the planetesimals assume the role of the molecules in gas dynamics theory. This approach is similar to that taken by Chandrasekhar (1942) in stellar dynamics. Nevertheless, the fact that the planetesimals are moving in Keplerian orbits rather than in free space requires some modification of Chandrasekhar's theory.

The most simple approach to such a "gas dynamics" theory of planetesimals is to simply assume that a planetesimal grows in mass (M) by sweep up of smaller bodies in accordance with a simple growth equation:

$$\frac{dM}{dt} = \pi R^2 \rho_s V F_g, \tag{1}$$

where R is the physical radius of the growing planetesimal, ρ_s is the surface mass density of the material being swept up, V is their relative velocity, and F_g represents the enhancement of the physical cross-section by "gravitational focussing," given in the two-body approximation by

$$F_g = (1 + 2\theta), \tag{2}$$

where θ is the Safronov number, $\theta = \frac{V_e^2}{V^2}$, and V_e is the escape velocity of the growing body.

Although it is possible to gain considerable insight into planetesimal growth by simple use of equation (1), its dependence on velocity limits its usefulness unless a way is found to calculate the relative velocity. Safronov (1962) made a major contribution to this problem by recognition that this relative velocity is not a free parameter, but is determined by the mass distribution of bodies. The mass distribution is in turn determined by the growth of the bodies, which in turn is dependent on the relative velocities by equation (1). Thus the mass and velocity evolution are coupled.

Safronov made use of Chandrasekhar's relaxation time theory to develop expressions for the coupled growth of mass and velocity. He showed that a steady-state velocity distribution in the swarm was established as

a result of the balance between "gravitational stirring" that on the average increased the relative velocity, and collisional damping, that decreased their relative velocity. The result was that the velocity and mass evolution were coupled in such a way that the relative velocity of the bodies was self-regulated to remain in the proper range, i.e. neither too high to prevent growth by fragmentation, nor too low to cause premature isolation of the growing bodies as a result of the eccentricity becoming too low. In Safronov's work the effect of gas drag on the bodies was not included. Hayashi and his coworkers (Nakagawa *et al.* 1983) complemented the work of Safronov and his colleagues by including the effects of gas drag, but did not include collisional damping. Despite these differences, their results are similar. The growth of the planetesimals to bodies of $\sim 10^{25} - 10^{26}$g begins with a steep initial distribution of bodies of nearly equal mass. With the passage of time, the larger bodies of the swarm remain of similar size and constitute a "marching front" that diminishes in number as the mass of the bodies increases. Masses of $\sim 10^{25}$g are achieved in $\sim 10^6$ years.

An alternative mode of growth was proposed by Greenberg *et al.* (1978). They found that instead of the orderly "marching front," runaway growth caused a single body to grow to $\sim 10^{23}$g in 10^4 years, at which time almost all the mass of the system remained in the form of the original 10^{16}g planetesimals. It is now known (Patterson and Spaute 1988) that the runaway growth found by Greenberg *et al.* were the result of an inaccurate numerical procedure. Nevertheless, as discussed below, it now appears likely that similar runaways are expected when the problem is treated using a more complete physical theory and sufficiently accurate numerical procedures.

This recent development emerged from the work of Stewart and Kaula (1980) who applied Boltzmann and Fokker-Planck equations to the problem of the velocity distribution of a swarm of planetesimals, as determined by their mutual gravitational and collisional evolution. This work was extended by Stewart and Wetherill (1988) to develop equations describing the rate of change of the velocity of a body of mass m_1 and velocity V_1 as a result of collisional and gravitational interaction with a swarm of bodies with masses m_2 and velocities V_2. In contrast with earlier work, these equations for the gravitational interactions contain dynamical friction terms of the form

$$\frac{dV_1}{dt} \propto (m_2 V_2^2 - m_1 V_1^2). \tag{3}$$

These terms tend to equipartition energy between the larger and smaller members of the swarm. For equal values of V_1 and V_2, they cause the velocity of a larger mass m_1 to *decrease* with time. In earlier work, the gravitationally induced "stirring" was always positive-definite, as a result of using relaxation time expressions that ensured this result.

The dV/dt equations of Stewart and Wetherill have been used to develop a numerical procedure for studying the evolution of the mass and velocity distribution of a growing swarm of planetesimals at a given heliocentric distance, including gas drag, as well as gravitational and collisional interactions (Wetherill and Stewart 1988).

When approached in this way it becomes clear that the coupled nonlinear equations describing the velocity and size distribution of the swarm bifuricate into two general types of solutions. The first, orderly growth, was described by the Moscow and Kyoto workers. The second is "runaway" growth whereby within a local zone of the solar nebula (e.g. 02 AU in width) a single body grows much faster than its neighbors and causes the mass distribution to become discontinuous at its upper end. Whether or not the runaway branch is entered depends on the physical parameters assumed for the planetesimals. More important however, are the physical processes included in the equations. In particular, inclusion of the equipartition of energy terms causes the solutions to enter the runaway branch for a very broad range of physical parameters and initial conditions. When these terms are not included, the results of Safronov, Hayashi, and their coworkers are confirmed (see Figure 1). On the other hand, when these terms are included, runaway solutions represent the normal outcome of the calculations.

The origin of the runaway can be easily understood. For an initial swarm of planetesimals of equal or nearly equal mass, the mass distribution will quickly disperse as a result of stochastic differences in the collision rate and thereby the growth rate of a large number of small bodies. As a result of the equipartition of energy terms, this will quickly lead to a velocity dispersion, whereby the larger bodies have velocities, relative to a circular orbit, significantly lower than that of the more numerous smaller bodies of the swarm. The velocity of the smaller bodies is actually accelerated by the same equipartition terms that decrease the velocity of the larger bodies.

A simplified illustration of this effect is shown in Figure 2. This calculation is simplified in that effects associated with failure of the two-body approximation at low velocities and with fragmentation are not included.

After only $\sim 3 \times 10^4$ years, the velocities of the largest bodies relative to those of the smaller bodies has dropped by an order of magnitude. These lower velocities increased the gravitational cross-section of the larger bodies sufficiently to cause them to grow approximately 100 times larger than those bodies in which most of the mass of the swarm is located. This "midpoint mass" (m_p), defined by being the mass below which half the mass of the swarm is located, is indicated on Figure 2. For bodies of this mass the Safronov number θ is 0.6 when defined as
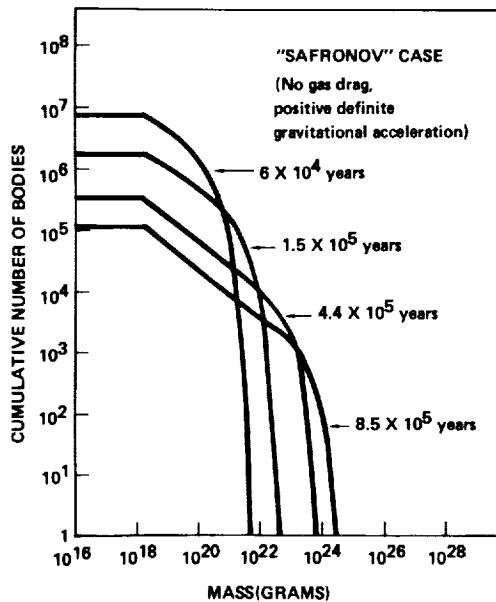
FIGURE 1a   Evolution of the mass distribution of a swarm of planetesimals distributed between 0.99 and 1.01 AU for which the velocity distribution is determined entirely by the balance between positive-definite gravitational "pumping up" of velocity and collisional damping. The growth is "orderly," i.e., it does not lead to a runaway, but rather to a mass distribution in which most of the mass is concentrated in $10^{24} - 10^{25}$ g bodies at the upper end of the mass distribution.

$$\theta = \frac{V_e^2(m_p)}{2V^2(m_p)} \tag{4}$$

i.e. its relative velocity is similar to its own escape velocity. In contrast, the value of θ_L calculated using the velocity of this body and that of the *largest* body of the swarm has a quite high value of 21.

At this early stage of evolution the growth is still orderly and continuous (Figure 3). However, by 1.3×10^5 years, the velocities of the largest bodies have become much lower than their escape velocities (Figure 2), and a bulge has developed at the upper end of the swarm as a result of their growing much faster than the smaller bodies in the swarm. At 2.6×10^5 years, a single discontinuously distributed body with a mass $\sim 10^{26}$ is found. At this time it has accumulated 13% of the swarm, and the next largest bodies are more than 100 times smaller. This runaway body will quickly capture all the residual material in the original accumulation zone, specified in this case to be 0.02 AU in width.

The orbit of the runaway body will be nearly circular, and it will be able
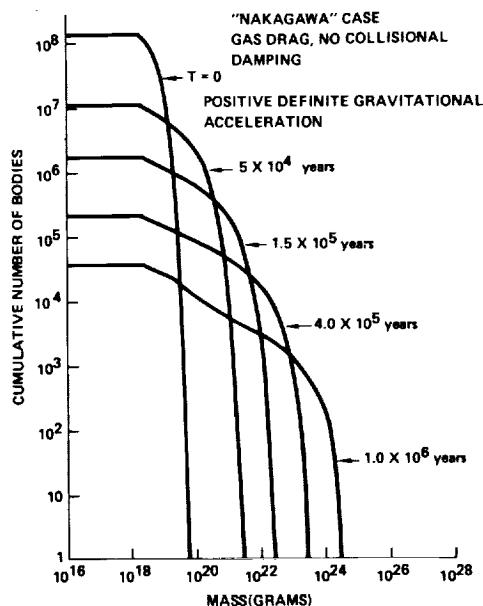
FIGURE 1b   Evolution of the mass distribution of a swarm in which the velocity damping is provided by gas drag, rather than by collisional damping. The resulting distribution is similar to that of Figure 1a.

to capture bodies approaching within several Hill sphere radii (Hill sphere radius = distance to colinear Lagrangian points). Even in the absence of competitors in neighboring zones, the runaway growth will probably self-terminate because additions to its mass (Δm) will be proportional to $(\Delta D)^2$, where ΔD is the change in planetesimal diameter, whereas the material available to be accumulated will be proportional to ΔD. Depending on the initial surface density, runaway growth of this kind can be expected to produce approximately 30 to 200 bodies in the terrestrial planet region with sizes ranging from that of the Moon to that of Mars.

There are a number of important physical processes that have not been included in this simplified model. These include the fragmentation of the smaller bodies of the swarm, the failure of the two-body approximation at low velocities, and the failure of the runaway body to be an effective perturber of small bodies that cross the orbit of only one runaway. These conditions are more difficult to model, but those calculations that have been made indicate that they all operate in the direction of increasing the rate of the runaway.
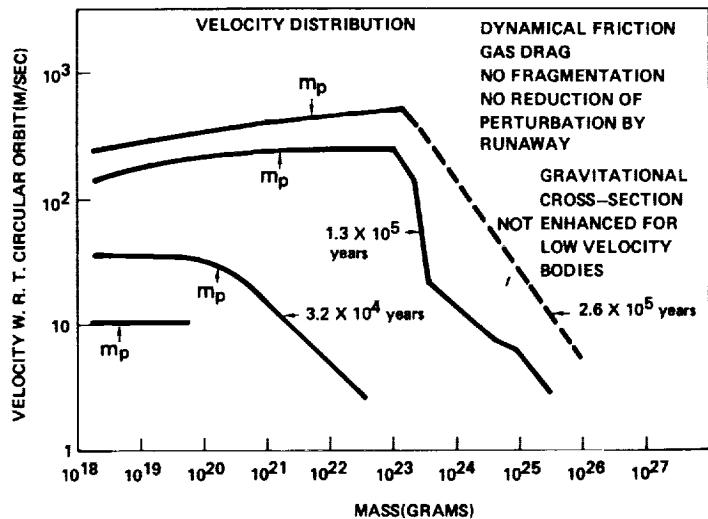
FIGURE 2 Velocity distribution corresponding to inclusion of equipartition of energy terms. After 3×10^4 years, the velocities of the largest bodies drop well below that of the midpoint mass m_p. This leads to a rapid growth of the largest bodies, and ultimately to a runaway, as described in the text.

GROWTH OF RUNAWAY PLANETARY EMBRYOS INTO TERRESTRIAL PLANETS

Because of the depletion of material in their vicinity, it seems most likely that the runaway bodies described above will only grow to masses in the range of 6×10^{25}g to 6×10^{26}g, and further accumulation of a number of these "planetary embryos" will be required to form bodies of the size of Earth and Venus.

Both two-dimensional and three-dimensional numerical simulations of this final accumulation of embryos into terrestrial planets have been reported. All of these simulations are in some sense "Monte Carlo" calculations, because even in the less demanding two-dimensional case, a complete numerical integration of several hundred bodies for the required number of orbital periods is computationally prohibitive. Even if such calculations were possible, the intrinsically chaotic nature of orbital evolution dominated by close encounters causes the final outcome to be so exquisitely sensitive to the initial conditions that the final outcome is essentially stochastic. Two-dimensional calculations have been reported by Cox and Lewis (1980); Wetherill (1980); Lecar and Aarseth (1986); and Ipatov (1981a). In some of these two-dimensional cases numerical integration was carried out during the close encounter.
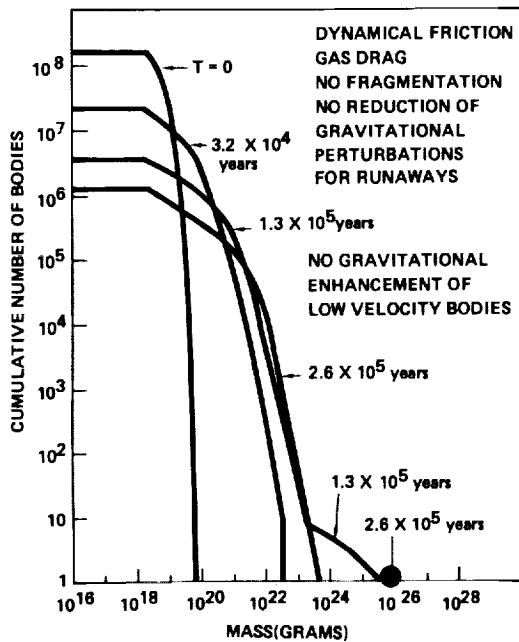
FIGURE 3 Effect of introducing equipartition of energy terms on the mass distribution. The tendency toward equipartition of energy results in a velocity dispersion (Figure 2) in which the velocity (with respect to a circular orbit) of the massive bodies falls below that of the swarm. After $\sim 10^5$ years, a "multiple runaway" appears as a bulge in the mass distribution in the mass range $10^{24} - 10^{25}$ g. After 2.6×10^5 years, the largest body has swept up these larger bodies, leading to a runaway in which the mass distribution is discontinuous. The largest body has a mass of $\sim 10^{26}$ g, whereas the other remaining bodies have masses $< 10^{24}$ g.

The three-dimensional calculations (Wetherill 1978, 1980, 1985, 1986, 1988) make use of a Monte Carlo technique based on the work of Öpik (1951) and Arnold (1965). In both the two- and three-dimensional calculations, the physical processes considered are mutual gravitational perturbations, physical collisions, and mergers, and in some cases collisional fragmentations and tidal disruption (Wetherill 1986, 1988).

In the work cited above it was necessary to initially confine the initial swarm to a region smaller than the space presently occupied by the observed terrestrial planets. This is necessary because a system of this kind nearly conserves mass, energy, and angular momentum. The terrestrial planets are so deep in the Sun's gravitational well that very little ($< 5\%$) of the material is perturbed into hyperbolic solar system escape orbits. The loss of mass, energy, and angular momentum by this route is therefore small.

Angular momentum is strictly conserved by gravitational perturbations and physical collisions. Some energy is radiated away as heat during collision and merger of planetesimals. A closed system of bodies, such as a stellar accretion disk, that conserves angular momentum and loses energy can only spread, not contract. Therefore the initial system of planetesimals must occupy a narrower range of heliocentric distance than the range of the present terrestrial planets. In particular (Wetherill 1978), it can be shown by simple calculations that a swarm that can evolve into the present system of the terrestrial planets must be initially confined to a narrow band extending from about 0.7 to 1.1 AU.

In theories in which planetesimals grow into embryos via the *orderly* branch of the bifurcation of the coupled velocity-size distribution equations, the time scale for growth of $\sim 10^{26}$g embryos at 1 AU is 1 to 2×10^6 years. If the surface density of material falls off as $a^{-3/2}$ beyond 1 AU, the time scale for similar growth at larger heliocentric distances will vary as a^{-3}, the additional $a^{-3/2}$ arising from the variation of orbital encounter frequency with orbital period. Thus at 2 AU, the comparable time scale for the growth of planetesimals into embryos would be 10-20 million years. Jupiter and Saturn must have formed while nebular gas was still abundant. Observations of pre-main-sequence stars, and theoretical calculations (Lissauer 1987; Wetherill 1989b) permits one to plausibly hypothesize that Jupiter and Saturn had already formed by the time terrestrial-type "rocky" planetesimals formed much beyond 1 AU. In some rather uncertain way it is usually supposed that the existence of these giant planets then not only cleared out the asteroid belt, aborted the growth of Mars, and also prevented the growth of planetesimals into embryos much beyond 1 AU. Interior to 0.7 AU, it can be hypothesized that high temperatures associated with proximity to the Sun restrained the formation or growth of planetesimals.

Subject to uncertainties associated with hypotheses of the kind discussed above, the published simulations of the final stages of planetary growth, show that an initial collection of several hundred embryos spontaneously evolve into two to five bodies in the general mass range of the present terrestrial planets. In some cases the size and distribution of the final bodies resemble rather remarkably those observed in the present solar system (Wetherill 1985). The process is highly stochastic, however, and more often an unfamiliar assemblage of final planets is found, e.g. $\sim$ three bodies, $\sim 4 \times 10^{27}$g of mass at 0.55, 1.0, and 1.4 AU.

Even when the initial embryos are quite small (i.e. as small as 1/6 lunar mass), it is found that the growth of these bodies into planets is characterized by giant impacts at rather high velocities ($>$ 10 kilometers per second). In the case of Earth and Venus, these impacting bodies may exceed the mass of the present planet Mars. These models of planetary accumulation thereby fit in well with theories of the formation of the Moon

accumulation thereby fit in well with theories of the formation of the Moon by "giant splashes" (Hartmann and Davis 1975; Cameron and Ward 1976), removal of Earth's atmosphere by giant impacts (Cameron 1983); and impact removal of Mercury's silicate mantle (Wetherill 1988; Benz *et al.* 1988).

Some rethinking of this discussion is required by the results of Wetherill and Stewart (1988). Now that it seems more likely that runaway growth of embryos at 1 AU took place on a much faster time scale, possibly as short as 3×10^4 years, it seems much less plausible that the formation of Jupiter and Saturn can explain the absence of one or more Earth-size planets beyond 1 AU. Growth rates in the asteroid belt may still be slow enough to be controlled by giant planet formation, but this will be more difficult interior to 2 AU.

The studies of terrestrial planet growth discussed earlier in which the only physical processes included are collisions and merger, are clearly inadequate to cause an extended swarm of embryos to evolve into the present compact group of terrestrial planets. Accomplishment of this will require inclusion of additional physical processes.

Three such processes are known, but their significance requires much better quantitative evaluation and understanding. These are:

(1) Loss of total and specific negative gravitational binding energy as well as angular momentum by exchange of these quantities to the gaseous nebula via spiral density waves (Ward 1986, 1988).

(2) Loss of material, with its associated energy and angular momentum from the complex of resonances in the vicinity of 2 AU (Figure 4). All of these resonances will not be present until after the formation of Jupiter and Saturn, and therefore are not likely to be able to prevent the formation of runaway bodies in this region. After approximately one million years however, they can facilitate removal of material from this region of the solar system by increasing the eccentricity of planetesimals into terrestrial planets and Jupiter-crossing orbits. Because bodies with larger semi-major axes are more vulnerable to being lost in this way, the effect will be to decrease the specific angular momentum of the swarm as well as cause the energy of the swarm to become less negative.

(3) Resonant interactions between the growing embryos. Studies of the orbital evolution of Earth-crossing asteroids (Milani *et al.* 1988) show that although the long-term orbital evolution of these bodies is likely to be dominated by close planetary encounters, more distant resonant interactions are also prominent. Similar phenomena are to be expected during the growth of embryos into planets. These have not been considered in a detailed way in the context of the mode of planetary growth outlined here, but relevant studies of resonant phenomena during planetary growth have
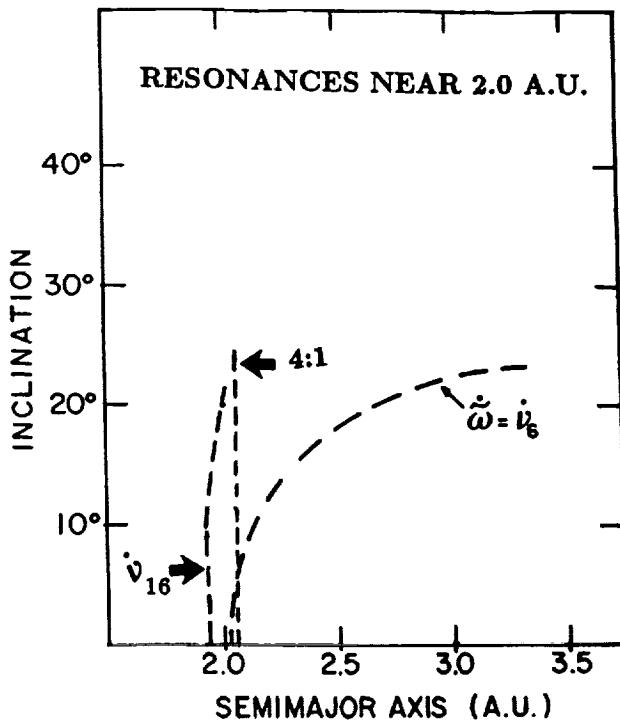
FIGURE 4   Complex of resonances in the present solar system in the vicinity of 2 AU. These resonances are likely responsible for forming a chaotic "giant Kirkwood gap" that defines the inner edge of the main asteroid belt.

been published (Ipatov 1981b; Weidenschilling and Davis 1985; Patterson 1987).

All of these phenomena represent real effects that undoubtedly were present in the early solar system and must be taken into consideration in any complete theory of terrestrial planet formation. Quantitative evaluation of their effect however, is difficult at present, and there is no good reason to believe they are adequate to the task.

A preliminary evaluation of the effect of the first two phenomena, spiral density waves and Jupiter-Saturn resonances near 2 AU have been carried out. Some of the result of this investigation are shown in Figures 5 and 6.

In Figure 5 the point marked "initial swarm" corresponds to the specific energy and angular momentum of an extended swarm of runaway planetesimals extending from 0.45 to 2.35 AU, with a surface density falling off as $1/a$. The size of the runaway embryos is prescribed by the condition
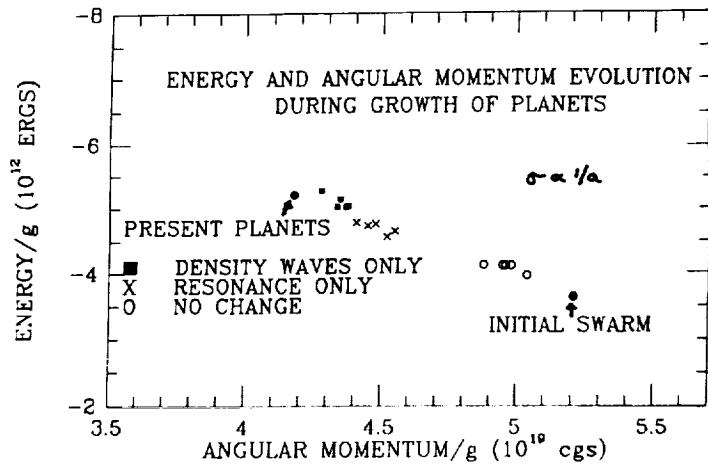
FIGURE 5 Energy and angular momentum evolution of an initial swarm with runaway planetesimal and gas surface density as 1/a between the 0.7 and 2.20 AU. Interior of 0.7 AU, it is assumed that the gas density is greatly reduced, possibly in an association bipolar outflow producing a "hole" in the center of the solar nebula. The surface density falls off exponentially between 2.20 and 2.35 AU and between 0.45 and 0.7 AU. The total initial mass of the swarm is 1.407 10^{28}g. The open circles represent simulations in which only gravitational perturbations and collisional damping are included. The crosses are simulations in which mass, angular momentum, and negative energy are lost by means of the resonances shown in Figure 3. The solid squares represent simulations in which angular momentum and losses result from inclusion of spiral density wave damping, as described by Ward (1986, 1988).

that they be separated by 4 Hill sphere radii from one another. As a result, the mass of the embryos is 2.3 × 10^{26}g at 2 AU, 0.8 × 10^{26}g at 1 AU, and 0.5 × 10^{26}g at 0.7 AU. The specific energy and angular momentum of the present solar system is indicated by the point so marked near the upper left of the Figure. The question posed here is whether inclusion of 2 AU resonances and Ward's equations for changes in semi-major axis and eccentricity of the swarm can cause the system to evolve from the initial point to the "goal" representing the present solar system. It is found that this is possible for sufficiently large values of the relevant parameters. More recent work (Ward, private communication 1989) indicates that the published parameters may be an order of magnitude too large. If so, this will greatly diminish the importance of this mechanism for angular momentum removal.

The open circles near the initial point show the results of five simulations in which neither of these effects were included. Some evolution toward the goal is achieved, nevertheless. This results from the more
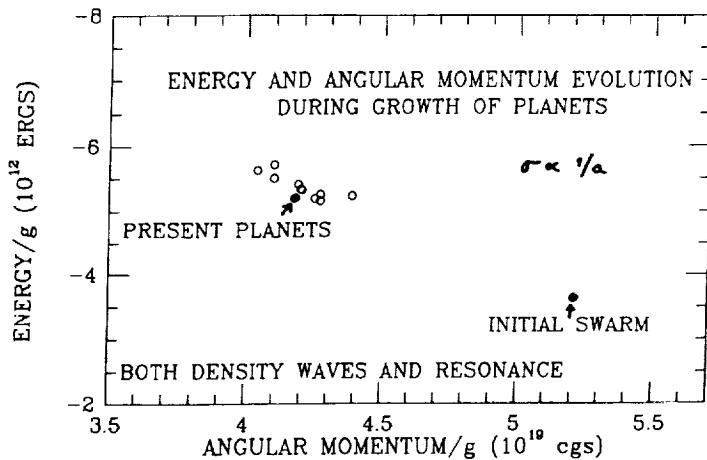
FIGURE 6   Energy and angular momentum loss when both the effects of resonances and spiral density are included. For the choice of parameters described in the text, the system evolves into a distribution matching that parameters described in the text, the system evolves into a distribution matching that observed.

distant members of extended swarm being more loosely bound than that employed in earlier studies, and consequently relatively more loss of bodies with higher angular momentum and with less negative energy.

This effect is enhanced when the resonances shown in Figure 4 are included (crosses in Figure 5). The effect of the resonances is introduced in a very approximate manner. If after a perturbation, the semi-major axis of a body is between 2.0 and 2.1 AU, its eccentricity is assigned a random value between 0.2 and 0.8. A similar displacement toward the specific angular momentum and energy of the present solar system is reached when da/dt and de/dt terms of the form given by Ward (1986, 1988) are included. The open squares in Figure 5 result from use of a coefficient having a value of 29 in Ward's equation for da/dt, and a value of 1 for de/dt. The value for da/dt is about twice that originally estimated by Ward (1986).

The effect of including both the resonances and the same values of spiral density wave damping are shown in Figure 6. The points lie quite near the values found for the terrestrial planets. Thus if one assumes appropriate values for these two phenomena, an initial swarm can evolve into one with specific energy and angular momentum about that found in the present solar system. The initial surface density can be adjusted to match the present total mass of the terrestrial planets, without disturbing the agreement with the observed energy and angular momentum. Similar agreement has been obtained in calculations in which both the gas and embryo surface densities varied as $a^{-3/2}$, instead of a^{-1} as used for the

data of Figure 6. Satisfactory matching has also been found for a swarm in which the gas surface density fell off as $a^{-3/2}$ whereas the embryo surface density decreased as a^{-1}.

Matching the mass, angular momentum, and energy of the present terrestrial planet system is a necessary, but not sufficient condition, for a proper model for the formation of the terrestrial planets. It is also necessary that to some degree the configuration (i.e. the number, position, mass, eccentricity, and inclination) of the bodies resemble those observed. Because a model of this kind is highly stochastic, and we have only one terrestrial planet system to observe, it is hard to know how exactly the configuration should match. As in the earlier work "good" matches are sometimes found, more often the total number of final planets with masses > 1/4 Earth mass is three, rather than the two observed bodies. It is possible this is a stochastic effect, but the author suspects it more likely that the differences result from the model being too simple. Neglecting such factors as the resonant interactions between the embryos as well as less obvious phenomena may be important.

Like the previous models in which the swarm was initially much more localized, the final stage of accumulation of these planets from embryos involves giant impacts. Typically, at least one body larger than the present planet Mars impacts the simulated "Earth," and impacts twice as large are not uncommon. Therefore, all of the effects related to such giant impacts, formation of the Moon, fragmentation of smaller planets, and impact loss of atmospheres are to be expected for terrestrial planet systems arising from the more extended initial embryo swarms of the kind considered here. Furthermore, the inward radial migration associated with density wave drag, as well as the acceleration in eccentricity caused by the resonances, augment the tendency for a widespread provenance of the embryos responsible for the chemical composition of the final planets.

ACKNOWLEDGMENTS

REFERENCES

Adachi, I., C. Hayashi, and K. Nakazawa. 1976. The gas drag effect on the elliptic motion of a solid body in the primordial solar nebula. Prog. Theor. Phys. 56:1756-1771.
Arnold, J.R. 1965. The origin of meteorites as small bodies. II. The model, III. General considerations. Astrophys. J. 141:1536-1556.

Benz, W., W.L. Slattery, and A.G.W. Cameron. 1988. Collisional stripping of Mercury's mantle. Icarus 74:516-528.

Cameron, A.G.W., and W.R. Ward. 1976. The origin of the Moon. Pages 120-122. In: Lunar Science VII. Lunar Science Institute, Houston.

Cameron, A.G.W. 1983. Origin of the atmospheres of the terrestrial planets. Icarus 56:195-201.

Cameron, A.G.W., W.M. DeCampli, and P.H. Bodenheimer. 1982. Evolution of giant gaseous protoplanets. Icarus 49:298-312.

Chamberlain, T.C. 1904. Fundamental problems of geology. Carnegie Inst. of Washington Year Book 3.

Chandrasekhar, S. 1942. Principles of stellar dynamics. University of Chicago Press, Chicago.

Chladni, E.F.F. 1794. Über den Ursprung der von PallasGefundenen und anderer ihr ähnlicher Eisen massen. J.F. Hartknoch, Riga. (Reprinted 1974 by the Meteoritical Society, with introduction by J.T. Wasson).

Cox, L.P., and J.S. Lewis. 1980. Numerical simulation of the final stages of terrestrial planet formation. Icarus 44: 706-721.

Edgeworth, K.E. 1949. The origin and evolution of the solar system. Mon. Not. Roy. Astron. Soc. 109:600-609.

Goldreich, P., and W.R. Ward. 1973. The formation of planetesimals.Astrophys. J. 183:1051-1061.

Greenberg, R., J. Wacker, C.R. Chapman, and W.K. Hartmann. 1978. Planetesimals to planets: numerical simulation of collisional evolution. Icarus 35:1-26.

Hartmann, W.K., and D.R. Davis. 1975. Satellite-sized planetesimals and lunar origin. Icarus 24:504-515.

Ipatov, S.I. 1981a. Computer modeling of the evolution of plane rings of gravitating particles moving about the Sun. Astron. Zh. 58:1085-1094.

Ipatov, S.I. 1981b. On the gravitational interaction of two planetesimals. Astron. Zh. 58:620-629.

LeCar, M., and S.J. Aarseth. 1986. A numerical simulation of the formation of the terrestrial planets. Astrophys. J. 305: 564-579.

Lissauer, J.J. 1987. Timescales for planetary accretion and the protoplanetary disk. Icarus 69:245-265.

Milani, A., G. Hahn, M. Carpino, and A.M. Nobili. 1988. Project Spaceguard: dynamics of planet-crossing asteroids. Classes of orbital behaviour. Icarus, in press.

Nakagawa, Y., C. Hayashi, and K. Nakazawa. 1983. Accumulation of planetesimals in the solar nebula. Icarus 54:361-376.

Öpik, E.J. 1951. Collision probabilities with the planets and the distribution of interplanetary matter. Proc. Roy. Irish Acad. 54A:165-199.

Patterson, C.W. 1987. Resonance capture and the evolution of the planets. Icarus 70:319-333.

Patterson, C., and D. Spaute. 1988. Self-termination of runaway growth in the asteroid belt. Abstr. Asteroids II Conf. March 8-11, 1988. Tucson, Arizona.

Safronov, V.S. 1960. Formation and evolution of protoplanetary dust condensations. Voprosy Kosmogonii 7:121-141.

Safronov, V.S. 1962. Velocity dispersion in rotating systems of gravitating bodies with inelastic collisions. Voprosy Kosmogonii 8:168.

Safronov, V.S. 1969. Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets. (Nauka, Moscow.) Translated for NASA and NSF by Israel Program for Scientific Translations, 1972. NASA TT F-677.

Stewart, G.R., and W.M. Kaula. 1980. Gravitational kinetic theory for planetesimals. Icarus 44:154-171.

Stewart, G.R., and G.W. Wetherill. 1988. Evolution of planetesimal velocities. Icarus 74:542-553.

Ward, W.R. 1986. Density waves in the solar nebula: differential Lindblad Torque. Icarus 67:164-180.

Ward, W.R. 1988. On disk-planet interactions and orbital eccentricities. Icarus 73:330-348.

Weidenschilling, S.J. 1977. Aerodynamics of solid bodies in the solar nebula. Mon. Not. Roy. Astron. Soc. 180:57-70.

Weidenschilling, S.J. 1984. Evolution of grains in a turbulent solar nebula. Icarus 60:553-567.

Weidenschilling, S.J., and D.R. Davis. 1985. Orbital resonances in the solar nebula: implications for planetary accretion. Icarus 62:16-29.

Weidenschilling, S.J., B. Donn and P. Meakin. 1988. Physics of planetesimal formation. In: Weaver, H. (ed.). The Formation and Evolution of Planetary Systems. Cambridge University Press, Cambridge, in press.

Wetherill, G.W. 1978. Accumulation of the terrestrial planets. Pages 565-598. In: Gehrels, T. (ed.). Protostars and Planets. University of Arizona Press, Tucson.

Wetherill, G.W. 1980. Formation of the terrestrial planets. Ann. Rev. Astron. Astrophys. 18:77-113.

Wetherill, G.W. 1985. Giant impacts during the growth of the terrestrial planets. Science 228:877-879.

Wetherill, G.W. 1986. Accumulation of the terrestrial planets and implications concerning lunar origin. Pages 519-550. In: Hartmann, W.K., R.J. Philips, and G.J. Taylor (eds.). Origin of the Moon. Lunar and Planetary Institute, Houston.

Wetherill, G.W. 1988. Accumulation of Mercury from planetesimals. In: Chapman, C., and F. Vilas (eds.). Mercury. University of Arizona Press, Tucson, in press.

Wetherill, G.W. 1989a. Formation of the solar system: consensus, alternatives, and missing factors. In: Weaver, H. (ed.). The Formation and Evolution of Planetary Systems. Cambridge University Press, Cambridge, in press.

Wetherill, G.W. 1989b. Origin of the asteroid belt. Pages 519-550. In: Binzel, R.P., T. Gehrels, and M.S. Matthews (eds.). Asteroids II. University of Arizona Press, Tucson.

Wetherill, G.W., and G.R. Stewart. 1988. Accumulation of a swarm of small planetesimals. Icarus, in press.

Whipple, F.L. 1973. Radial pressure in the solar nebula as affecting the motions of planetesimals. Pages 355-361. In: Hemenway, C.L., P.M. Millman, and A.F. Cook (eds.). Evolutionary and Physical Properties of Meteoroids. AA NASA SP-319.

The Rate of Planet Formation and the Solar System's Small Bodies

VIKTOR S. SAFRONOV
Schmidt Institute of the Physics of the Earth

ABSTRACT

The evolution of random velocities and the mass distribution of pre-planetary body at the early stage of accumulation are currently under review. Arguments have been presented for and against the view of an extremely rapid, runaway growth of the largest bodies at this stage with parameter values of $\Theta \gtrsim 10^3$. Difficulties are encountered assuming such a large Θ: (a) bodies of the Jovian zone penetrate the asteroid zone too late and do not have time to hinder the formation of a normal-sized planet in the astroidal zone and thereby remove a significant portion of the mass of solid matter and (b) Uranus and Neptune cannot eject bodies from the solar system into the cometary cloud. Therefore, the values $\Theta < 10^2$ appear to be preferable.

INTRODUCTION

By the beginning of the twentieth century, Ligondes, Chamberlain, and Multan had suggested the idea of planet formation via the combining (accumulation) of solid particles and bodies. However, it long remained forgotten. It was only by the 1940s that this idea was revived by O.Yu. Schmidt, the outstanding Soviet scientist and academician (1944, 1957). He initiated a systematic study of this problem and laid the groundwork for contemporary planet formation theory. He also suggested the first formula for the rate of mass increase of a planet which is absorbing all the bodies colliding with it. After it was amended and added to, this formula took on its present form (Safronov 1954, 1969):

116

$$\frac{dm}{dt} = \pi r^2 \rho v \left(1 + \frac{2Gm}{v^2 r}\right) = \frac{4\pi r 2\sigma(1 + 2\Theta)}{p}, \tag{1}$$

where m and r are the mass and radius of the accreting planet, P is its period of revolution around the Sun, ρ and σ are the volume and surface density of solid matter in a planet's zone, and Θ is the dimensionless parameter characterizing random velocities of bodies in a planet's zone (in relation to the Kepler circular velocity of a preplanetary swarm's rotation):

$$v = \left(\frac{Gm}{\Theta r}\right)^{1/2} \propto r\Theta^{-1/2}. \tag{2}$$

However, Schmidt did not consider the increase of a planet's collisional cross section as a result of focusing orbits via its gravitational field, and the factor $(1 + 2\Theta)$ in his formula was absent. For an independent feeding zone of a planet, the surface density σ at a point in time t is related to the initial surface density σ_o by the simple ratio:

$$\sigma = \sigma_o(1 - m/Q), \tag{3}$$

where Q is the total mass of matter in a feeding zone of m. It is proportional to the width of the zone $2\Delta R_f$ and is determined, when Θ is on the order of several units, by eccentricities of the orbits of bodies, that is, by their velocities v; when $\Theta \gg 1$, it is determined by the radius of the sphere of the planet's gravitational influence. In both cases it is proportional to the radius of an accreting planet r. (Schmidt took Q to be equal to the present mass of a planet.)

It is clear from (1) and (2) that relative velocities of bodies in the planet's zone are the most significant factor determining a planet's growth rate. The characteristic accumulation time scale is $\tau_a \propto \Theta^{-1} \propto v^2$. Investigations have shown that velocities of bodies are dependent primarily upon their distribution by mass. Velocities of bodies in the swarm rotating around the Sun increase as a result of the gravitational interactions of bodies and decrease as a result of their inelastic collisions. In a quasistationary state, opposing factors are balanced out and certain velocities are established in the system. If the bulk of the mass is concentrated in the larger bodies, expression (2) for the velocities is applicable for velocities with a parameter Θ on the order of several units. In the extreme case of bodies of equal mass, $\Theta \approx 1$. If the bulk of mass is contained in smaller bodies, an average gravitating mass is considerably less than the mass m of the largest body. The parameter Θ in expression (2) then increases significantly. The velocities of bodies, in turn, influence their mass distribution. Therefore, we need investigations of the coupled evolution of both distributions to produce a strict solution to the problem. This problem cannot be solved

analytically. We thus divided it into two parts: (1) relative velocities of bodies were estimated for a pre-assigned mass distribution, and (2) the mass spectrum was determined for pre-assigned velocities. At the same time, we conducted a qualitative study of the coagulation equation for preplanetary bodies. This effort yielded asymptotic solutions in the form of an inverse power law with an exponent q:

$$n(m) = cm^{-q}, \tag{4}$$

(1.5 < q < 2) which is valid for all values m except for the largest bodies. Stable and unstable solutions were found and disclosed an instability of solutions where q > 2 was noted. In this case, the system does not evolve in a steady-state manner. Then the velocities of bodies assumed a power the law of mass distribution (3) with q < 2. They are most conveniently expressed in the form (2). Then $\Theta \approx 3 \div 5$ was found for the system without the gas. In the presence of gas, the parameter Θ is two to three times greater for relatively small bodies.

The most lengthy stage was the final stage of accumulation at which the amount of matter left unaccreted was significantly reduced. There was almost no gas remaining at this stage in the terrestrial planet zone. We found from Expression (1) that with $\Theta = 3$ the Earth ($\sigma_o \approx 10$ g/cm^2) accreted about 99% of its present mass in a $\approx 10^8$ year period. The other terrestrial planets were also formed during approximately the same time scale. The time scale of this accumulation process has been repeatedly discussed, revised, and again confirmed for more than 20 years. It may seem strange, but this figure remains also the most probable at this time.

The situation in the region of the giant planets has proven much more complex. From eq.(1) we can find an approximate expression for the time scale T formation of the planet, assuming $\bar{\sigma} \approx \sigma_o/2$. Thus,

$$T \approx \frac{\delta r P}{\Theta \sigma_o}, \tag{5}$$

where δ is the planet density. Current masses of the outer planets (Uranus and Neptune) correspond to $\sigma_o \approx 0.3$, that is, to a value about 30 times less than in the Earth's zone. The periods of revolution of these planets around the Sun $P \propto R^{3/2}$ are two orders of magnitude greater than that of the Earth's. Therefore, with the same values for Θ (cited above), the growth time scale of Uranus and Neptune would be unacceptably high: $T \sim 10^{11}$ years. Of course, this kind of result is not proof that the theory is invalid. It does, however, indicate that some important factors have not been taken into account in that theory. In order to obtain a reasonable value for T, we must increase product $\Theta \sigma_o$ in (5) several dozen times. More careful consideration has shown that such an increase of $\Theta \sigma_o$ has real grounds.

According to (2), velocities of bodies should increase proportionally to the radius of the planet. It is easy to estimate that when Uranus and Neptune grew to about one half of the Earth's mass (i.e., a few percent of their present masses), velocities of bodies for Θ = const $\approx$ 5 should have already reached the third cosmic velocity, and the bodies would escape the solar system. Therefore, further growth of planet mass occurred with v = const, and consequently, according to (2), Θ must have increased. With the increase of m, the rate of ejection of bodies increased more rapidly than the growth rate of the planet. To the end of accumulation it exceeded the latter several times, the parameter Θ begin increased about an order of magnitude. It also follows from here that the initial amount of matter in the region of the giant planets (that is, σ_o) was several times greater than the amount entered into these planets. Therefore, the difficulty with the time scale for the formation of the outer planets proved surmountable (at least in the first approximation). Furthermore, the very discovery of this difficulty made it possible to discern an important, characteristic feature of the giant planet accumulation process: removal of bodies beyond the boundaries of the planetary system. Since they are not only removed from the solar system, but also to its outer region primarily, a source of bodies was thereby discovered which formed the cometary cloud.

The basic possibility of runaway growth, that is "runaway" in terms of the mass of the largest body from the general distribution of the mass of the remaining bodies in its feeding zone, has been demonstrated (Safronov 1969). Collisional cross-sections of the largest gravitating bodies are proportional to the fourth degree of their radii. Therefore, the ratio of the masses of the first largest body m (planet "embryo") to the mass m_1 of the second largest body, which is located in the feeding zone of m, increases with time. An upper limit for this ratio was found for the case of Θ = const: $\lim(m/m_1) \approx (2\Theta)^3$.

An independent estimate of m/m_1 based on the present inclinations of the planetary rotation axes (naturally considered as the result of large bodies falling at the final stage of accumulation) was in agreement with this maximum ratio with the values $\Theta \approx 3 \div 5$.

These results were a first approximation and, naturally, required further in-depth analysis. Several years later, workers began to critically review the results from opposing positions. Levin (1978) drew attention to the fact that as a runaway m occurred, parameter Θ must increase. The ratio m/m_1 will correspondingly increase. Assuming m_1 and not m is an effective perturbing body in expression (3), he concluded that the ratio m/m_1 may have unlimited growth. Another conclusion was reached in a model of many planet embryos (Safronov and Ruzmaikina 1978; Vityazev *et al.* 1978; Pechernikova and Vityazev 1979). At an early stage all the bodies were small. The zones of gravitational influence and feeding of the largest

bodies, proportional to their radii, were narrow. Each zone had its own leader (a potential planet embryo) and there were many such embryos in the entire zone of the planet (with the total mass m_p): $N_e \sim (m_p/m)^{1/3}$ for low values of Θ and several times greater when Θ was high. Leaders with no overlapping feeding zones were in relatively similar conditions. Their runaway growth in relation to the remaining bodies in their own zones was not runaway in relation to each other. There was only a slight difference in growth rates that was related to the change of σ/P with the distances from the Sun. This difference brought about variation in the masses of two adjacent embryos in the terrestrial planet region $m/m_1 \sim 1 + 2m/m_p$. As the leaders grew, their ring zones widened, adjacent zones overlapped, adjacent leaders appeared in the same zone and the largest of them began to grow faster than the smaller one which had ceased being the leader. Masses of the leader grew, but their number was decreased. Normal bodies, lagging far behind the leader in terms of mass, and former leaders m_1, m_2, ..., the largest of which had masses of $\sim 10^{-1}m$, were located in the zone of each leader m. Consequently, there was no substantial gap in the mass distribution of bodies. If the bulk of the mass in this distribution had been concentrated in the larger bodies, for example, if it had been compatible with the power law (4) with the exponent $q < 2$, the relative velocities of bodies could have been written in the form (2), with values of Θ within the first 10. The leaders in this model comprised a fraction μ_e of all the matter in the planet's zone, which for low values of Θ is equal to

$$\mu_e = N_e m/m_p \approx (m/m_p)^{2/3}, \tag{6}$$

while for large values of Θ, it is several times greater. Approximately the same mass is contained in former leaders. At the early stages of accumulation $m \ll m_p$ and $\mu_e \ll 1$. Thus, velocities of bodies are not controlled by leaders and former leaders, but by all the remaining bodies and are highly dependent upon the mass distribution of these bodies. In the case of the power law (4) with $q < 2$, runaway embryo growth leads only to a moderate increase in Θ to $\Theta_{max} \sim 10 \div 20$ at $M_e \sim 10^{-1}$, when control of velocities begins to be shifted to the embryos and Θ decreases to $1 \div 2$ at the end of accumulation (Safronov 1982). However, it has not proven possible, without numerical simulation of the process which takes into account main physical factors, to find the mass distribution that is established during the accumulation.

The first model calculation of the coupled evolution of the mass and velocity distributions of bodies at the early stage of accumulation (Greenberg *et al.* 1978) already led to interesting results. The authors found that the system, originally consisting of identical, kilometer-sized bodies, did not produce a steady-state mass distribution, like the inverse

power law with q < 2. Only several large bodies with r ~ 200 kilometers had formed in it within a brief time scale ($2 \cdot 10^4$ years.), while the predominate mass of matter continued to be held in small bodies. Therefore, velocities of bodies also remained low. Hence, Θ and not v increased in expression (2) as m increased. The algorithm did not allow for further extension of the calculations. If we approximate the mass distribution contained in Graph 4 of this paper by the power law (this is possible with r > 5 kilometers), we easily find that the indicator q decreases over time from q $\gtrsim$ 10 where t = 15,000 years to $\approx$ 3.5 where t = 22,000 years. It can be expected that further evolution of the system leads to q < 2.

Lissauer (1987) and Artymovich (1987) later considered the possibility of rapid protoplanet growth (of the largest body) at very low velocities of the surrounding bodies. The basic idea behind their research was the rapid accretion by a protoplanet m of bodies in its zone of gravitational influence. The width of this zone $\Delta R_g = k r_H$ is equal to several Hills sphere radii $r_H = (m/3M_\odot)^{1/3} R$. As m increases, zone ΔR_g expands and new bodies appear in it, which, according to the hypothesis, had virtually been in circular orbits prior to this. According to Artymovich, under the impact of perturbations of m, these bodies acquire the same random velocities as the remaining bodies of the zone of m over a time scale of approximately 20 synodical orbital periods. Assuming that any body entering the Hills sphere of m, falls onto m (or is forever entrapped in this sphere, for example, by a massive atmosphere or satellite swarm, and only then falls onto m), Artymovich obtains a very rapid runaway growth of m until the depletion of matter in zone ΔR_g within $3 \cdot 10^4$ years in Earth's zone and $4 \cdot 10^5$ years in Neptune's zone. Subsequent slow increase of ΔR_g and growth of m occurs as a result of the increase in the eccentricity of the orbit of m owing to perturbations of adjacent protoplanets. Lissauer estimates the growth rate of m using the usual formula (1). Assuming random velocities of bodies until their encounters with a protoplanet to be extremely low, he assumes that after the encounter they approximate the difference of Kepler circular velocities at a distance of $\Delta R = r_H$. That is, they are proportional to $R^{-1/2}$. From this he obtains $\Theta_{eff} \propto R$. Assuming further that these rates are equal to the escape velocity from m at a distance of r_H, he finds for R = 1 AU $v/v_e = (r/r_H)^{1/2} \approx 1/15$, where $v_e = (2Gm/r)^{1/2}$, and he produces the "upper limit" of $\Theta_{eff} \approx 400$ for this value of v/v_e using data from numerical integration (Wetherill and Cox 1985). To generalize the numerical results to varying R's, he proposes the expression

$$\Theta_{eff} \sim 400 R_{ae}(\delta/4)^{1/3}, \tag{7}$$

for a "maximum effective accretion cross-section" at the earliest stage of

accumulation, where R_{ae} denotes the distance from the Sun in astronomical units.

This approach has sparked great interest. First of all, it makes it possible to reconcile rapid accumulation with runaway of large bodies in computations relating to the early stage with slow accumulation in computations for the final stage. Runaway growth in the Earth's zone ends with m on the order of several lunar masses; in Jupiter's zone it is about ten Earth masses (within a time scale of less than 10^6 years). Secondly, an increase in Θ with a distance $\propto R$ from the Sun significantly accelerates the growth of giant planets and may help in removing difficulties stemming from the length of their formation process. Therefore, we need more detailed consideration of the plausibility of the initial assumptions of this work, and an assessment of the role of factors which were not taken into account, namely, the overlapping of ΔR_g zones of adjacent protoplanets and the encounters of bodies with more than one of these and repeated encounters of bodies with a protoplanet and collisions between bodies. This would allow an estimation of the extent to which the actual values of Θ may differ from Lissauer's value for Θ_{eff}, qualified by himself as the maximum value.

The idea of runaway planet growth has not been confirmed in several other studies. These include, for example, numerical simulation by Lecar and Aarseth (1986), Ipatov (1988), Hayakawa and Mizutani (1988), and analytical estimates by Pechernikova and Vityazev (1979). Wetherill notes that the runaway growth obtained by Greenberg et al. (1978) is related to defects in the computation program. Nevertheless, he feels that runaway growth may stem from other, as yet unaccounted factors. The first of these is the tendency toward an equiportion of energy of bodies of varying masses, which reduces velocity of the largest bodies and leads to acceleration of their growth (1990, in this volume).

Stevenson and Lunine (1988) recently proposed a mechanism for the volatile compound enrichment of the Jovian zone (primarily H_2O) as a result of turbulent transport of volatiles together with solar nebula gas from the region of the terrestrial planets. Vapor condensation in the narrow band of $\Delta R \approx 0.4$ AU at a distance of $R \approx 5$ AU may increase the surface density of solid matter (primarily ices) in it by an order of magnitude and significantly speed up the growth of bodies at an early stage. Using Lissauer's model of runaway growth, the authors have found a time scale of of $\sim 10^5 \div 10^6$ years for Jupiter's formation.

The possibility of the acceleration of Jupiter's growth by virtue of this mechanism is extremely tempting and merits further detailed study. At the same time, complications may also arise. If, for example, not one, but several large bodies of comparable size are formed in a dense ring, their mutual gravitational perturbations will increase velocities of bodies, the ring will expand, and accumulation will slow down. In view of this reasoning,

it must be noted that the question of the role of runaway growth in planet formation cannot be considered resolved, despite the great progress achieved in accumulation theory. It is, therefore, worthwhile to consider how the existence of asteroids and comets constrains the process.

CONDITIONS OF ASTEROID FORMATION

It is now widely recognized that there have never been normal-sized planets in the asteroid zone. Schmidt (1954) was convinced that the growth of preplanetary bodies in this zone originally occurred in the same way as in other zones, but was interrupted at a rather early stage because of its proximity to massive Jupiter. Jupiter had succeeded in growing earlier and its gravitational perturbations increased the relative velocities of the asteroid bodies. As a result, the process by which bodies merged in collisions was superseded by an inverse process: their destruction and breakdown. It was later found that Jovian perturbations may increase velocities and even expel asteroids from the outermost edge of the belt R > 3.5 AU, and resonance asteroids from "Kirkwood gaps," whose periods are commensurate with Jupiter's period of revolution around the Sun. The gaps are extremely narrow, while the mass of all of the asteroids is only one thousandth of the Earth's mass. Therefore, removal of 99.9% of the mass of primary matter from the asteroid zone is a more complex problem than the increase of relative velocities of the remaining asteroids, (up to five kilometers per second, on the average).

A higher density of solid matter in the Jovian zone, due to condensation of volatiles, triggered a more rapid growth of bodies in the zone, and correspondingly, the growth of random velocities of bodies and eccentricities of their orbits. With masses of the largest bodies $> 10^{27}$ g, the smaller bodies of the Jovian zone (JZB) began penetrating the asteroid zone (AZ) and "sweeping out" all the asteroidal bodies which stood in its way and were of significantly smaller size. It has been suggested that the bulk of bodies was removed from the asteroid zone in this way (Safronov 1969). Subsequent estimates have shown that the JZB could only have removed about one half of the initial mass of AZ matter (Safronov 1979). In 1973, Cameron and Pine proposed a resonance mechanism by which resonances scan the AZ during the dissipation of gas from solar nebulae. However, it was demonstrated (Torbett and Smoluchowski 1980) that for this to be true, a lost mass of gas must have exceeded the mass of the Sun. In a model of low-mass solar nebulae ($\lesssim 0.1M_\odot$), resonance displacement could have been more effective as Jupiter's distance from the Sun varied both during its accretion of gas and its removal of bodies from the solar system.

We can make a comparison of the growth rates of asteroids and JZB using (1), if we express it as

$$\frac{dr_a}{dr_j} = \frac{(1+2\Theta_a)\sigma_a\delta_j}{(1+2\Theta_j)\sigma_j\delta_a}\left(\frac{R_j}{R_a}\right)^{3/2}, \tag{8}$$

where the indices a and j, respectively, denote the zones under consideration. Assuming $\Theta_j \approx 2\Theta_a$, $\delta_j \approx 2\delta_a/3$, $\sigma_j/\sigma_a \approx 3(R_j/R_a)^{-n}$, that is, a threefold density increase in the Jovian zone due to the condensation of volatiles, we have $dr_a/dr_j \approx (R_j/R_a)^{3/2+n}/9 \approx 0.1 \cdot 2^{3/2+n}$. It is clear from this that the generally accepted value for the density $\sigma g(R)$ of a gaseous solar nebula of $n = 3/2$ yields $dr_a \approx dr_j$, that is, it does not cause asteroid growth to lag behind JZB. In order for JZB's to have effectively swept bodies out of the AZ, there would have to have been a slower drop of $\sigma g(R)$ with $n \approx 1/2$ (Ruzmaikin et al. 1989) and, correspondingly, for solid matter $\sigma j \approx 15 \div 20$ g/cm^2. This condition is conserved when there is runaway growth of the largest bodies in both zones. Expression (8) is now formulated not for the largest, but for the second largest bodies. The ratio $\Theta_j \approx 2\Theta_a$ is conserved. However, a large body rapidly grows in the asteroidal zone, creating the problem of how it is to be removed. The second condition for effective AZ purging is the timely appearance of JZB's in it. Eccentricities of their orbits must increase to 0.3 − 0.4, while random velocities of $v \approx ewR$ must rise to two to three kilometers per second. According to (2), the mass of Jupiter's "embryo" must have increased to a value of $m'_j \approx m_\oplus(0.1\cdot\Theta)^{3/2}$. It is clear from this that JZB penetrating the asteroidal zone at the stage of rapid runaway growth of Jupiter's core, according to Lissauer, (with $\Theta_{eff} \approx 10^3$ to $m_c \approx 15m_\oplus$) is ruled out entirely. In view of these considerations, the values of $\Theta_j \lesssim 20 \div 30$ and $\Theta_a \lesssim 10 \div 15$ are more preferable. But the time scale for the growth of Jupiter's core to its accretion of gas then reaches 10^7 years.

THE REMOVAL OF BODIES FROM THE SOLAR SYSTEM AND THE FORMATION OF THE COMETARY CLOUD

Large masses of giant planets inevitably lead to high velocities of bodies at the final stage of accumulation and, consequently, to the removal of bodies from the solar system at this stage. It follows from this that the initial mass of solid matter in the region of giant planets was significantly greater than the mass which these planets now contain. It also follows that part of the ejecta remained on the outskirts of the solar system and formed the cloud of comets. The condition, that the total angular momentum of all matter be conserved, imposes a constraint upon the expelled mass ($\sim 10^2 m_\oplus$). Since the overwhelming majority of comets unquestionably belongs to the solar system and could not have been captured from outside, and since in situ comet formation at distances of more than 100 AU from the

Sun are only possible with an unacceptably large mass of the solar nebula, the removal of bodies by giant planets appears to be a more realistic way of forming a comet cloud. The condition for this removal is of the same kind as the condition for the ejection of bodies into the asteroidal zone. It is more stringent for Jupiter: velocities of bodies must be twice as great. For the outer planets, the ejection condition can be expressed as $m_p/m_\oplus >$ $4 \ (\Theta/R_{ae})^{3/2}$. It follows from this that Neptune's removal of bodies is only possible where $\Theta < 10^2$; for Uranus it is only possible for an even lower Θ.

REFERENCES

Artymowicz, P. 1987. Self-regulating protoplanet growth. Icarus 70:303-318.

Greenberg, R., J. Wacker, W.K. Hartmann, and C.R. Chapman. 1978. Planetesimals to planets: Numerical simulation of collisional evolution. Icarus 35:1-26.

Hayakawa, M., and H. Mizutani. 1988. Numerical simulation of planetary accretion process. Lunar Planet. Sci. XIX:455-456.

Ipatov, S.I. 1988. Solid-state accumulation of the terrestrial planets. Astron. Vestnik. 21:207-215.

Lecar, M., and S.J. Aarseth. 1986. A numerical simulation of the formation of the terrestrial planets. Astrophys. J. 305: 564-579.

Levin, B.Yu. 1978. Several issues of planetary accumulation. Letters to AJ. 4:102-107.

Lissauer, J. 1987. Time scales for planetary accretion and the structure of the protoplanetary disk. Icarus 69:249-265.

Pechernikova, G.V., and A.V. Vityazev. 1979. Masses of largerbodies and dispersion of velocities during planet accumulation. Letters to AJ 5:(1):54-59.

Ruzmaikina, T.V., V.S. Safronov, and S.J. Weidenschilling. 1989. Radial mixing of material in the asteroidal zone. In: Binzel, R.P., and M.S. Matthews (eds.). Asteroids II. University of Arizona Press, Tucson.

Safronov, V.S. 1954. On the growth of planets in the protoplanetary cloud. Astron. J. 31:499-510.

Safronov, V.S. 1969. Evolution of the Preplanetary Cloud and the Formation of the Earth and the Planets. Nauka, Moscow.

Safronov, V.S. 1982. The present state of the theory on the Earth's origin. AS USSR Publ., Fizika Zemli 6:5-24.

Safronov, V.S., and T.V. Ruzmaikina. 1978. On the angular momentum transfer and accumulation of solid bodies in the solar nebula. Pages 545-564. In: Gehrels, T. (ed.). Protostars and Planets. University of Arizona Press, Tucson.

Schmidt, O.Yu. 1944. Meteorite theory on the origin of the Earth and the planets. DAS USSR 45(6):245-249.

Schmidt, O.Yu. 1957. Four Lectures on the Theory of the Earth's Origins. 3rd ed. AS USSR Publ., Moscow.

Stevenson, D.J., and J.I. Lunine. 1988. Rapid formation of Jupiter by diffusive redistribution of water vapor in the solar nebula. Icarus 75:146-155.

Torbett M., and R.V. Smoluchowski. 1980. Sweeping of the Jovian resonances and evolution of the asteroids. Icarus 44:722-729.

Vityazev, A.V., G.V. Pechernikova, and V.S. Safronov. 1978. Maximum masses, distances and accumulation time scales for terrestrial planets. Astron. J. 55(1):107-112.

Wetherill, G.W. 1989. Formation of the terrestrial planets from planetesimals, this volume.

Wetherill, G.W., and L.P. Cox. 1985. The range of validity of the two-body approximation in models of terrestrial planet accumulation. II. Gravitational cross-sections and runaway accretion. Icarus 63:290-303.

Astrophysical Dust Grains in Stars, the Interstellar Medium, and the Solar System

Robert D. Gehrz
University of Minnesota

ABSTRACT

Studies of astrophysical dust grains in circumstellar shells, the interstellar medium (ISM), and the solar system, may provide information about stellar evolution and about physical conditions in the primitive solar nebula. Infrared observations give information about the mineral composition and size distribution of the grains. Grain materials identified in sources external to the solar system include silicates, silicon carbide, amorphous carbon, and possibly hydrocarbon compounds. The nucleation and growth of astrophysical carbon grains has been documented by infrared observations of classical novae. In the solar system, dust is known to be a major constituent of comet nuclei, and infrared spectroscopy of comets during perihelion passage has shown that the ablated material contains silicates, amorphous carbon, and hydrocarbons. Cometary grains resemble extra-solar-system grains in some ways, but there is evidence for additional processing of the grain materials in comets. Comets are discussed as a possible source for zodiacal dust.

Solar system grain materials have been sampled by the collection of micrometeorites and by isolating microscopic inclusions in meteorites. Meteorite inclusions exhibit several chemical abundance anomalies that are similar to those predicted to be produced in the explosive nucleosynthesis that accompanies novae and supernovae. The possible connections between extra-solar-system astrophysical dust grains and the grains in the solar system are explored. A recent suggestion that grains are rapidly destroyed in the interstellar medium by supernova shocks is discussed. Experiments to establish the relationships between extra-solar-system astrophysical grains

and solar system grains, and between cometary dust and the zodiacal dust are suggested. Among the most promising are sample return missions and improved high-resolution infrared spectroscopic information.

THE CYCLING OF DUST IN STELLAR EVOLUTION AND THE FORMATION OF PLANETARY SYSTEMS

Small refractory dust grains are present in circumstellar shells around many different classes of stars, in the interstellar medium (ISM), and in comets and the zodiacal cloud in the solar system. The mineral composition and size distribution of the grains in all three environments have similarities, but there are also distinct differences. Condensable elements produced in stars during nucleosynthesis presumably condense into grains during the final mass-loss stages of stellar evolution in aging stars like M-type giants and supergiants (Gehrz 1989), or during nova and supernova eruptions (Clayton 1982; Gehrz 1988). These grains can be expelled into the ISM where they may be processed further by supernova shocks and in molecular clouds. The grains can eventually be incorporated into young stars and planetary systems during star formation in the clouds (Gehrz *et al.* 1984).

Dusty and rocky solids that may derive from remnants of the formative phase are present in our own solar system and around some other main-sequence stars. Grains could therefore be a significant reservoir for condensable elements, effecting the transportation of these elements from their sites of production in stars into new stellar and planetary systems.

Studies of astrophysical grains can provide significant information about stellar evolution, physical conditions in circumstellar environments, and processes that occur during star formation. In particular, grains in the solar system may contain evidence about conditions in the early solar nebula. A major issue is whether grains made by stars actually survive intact in significant quantities as constituents of mature planetary systems, or whether most of the dust we see in the ISM and the solar system represents a re-accretion of condensables following the destruction of circumstellar or interstellar grains. There is theoretical evidence that grains can be destroyed in both the ISM (Seab 1987) and during the formation of planetary systems (Boss 1988). A paucity of gas-phase condensables in the ISM (Jenkins 1987), a low input rate to the ISM of dust from evolved stars (Gehrz 1989), and evidence for a hydrocarbon component in ISM grains (Allamandola *et al.* 1987) all suggest that grains can accrete material in molecular clouds. If the grains can survive the interstellar and star formation environments intact, then studies of the elemental abundances and minerology of solar system grains may provide fundamental information about stellar nucleosynthesis and evolution. If the grains are substantially processed, or evaporate and recondense after leaving the circumstellar

environment, then their chemical history may be much more difficult to evaluate, and their current composition may reflect relatively recent events. This review discusses the observed characteristics of astrophysical grains, compares the properties of grains in the solar and extra-solar environments, and suggests investigations to address the question of whether stardust is an important constituent of solar system solids.

ASTROPHYSICAL DUST GRAINS IN
CIRCUMSTELLAR ENVIRONMENTS

Infrared observations made more than two decades ago provided the first convincing evidence that dust grains are present in the winds of late-type giant and supergiant stars. Most of this circumstellar dust is refractory material. Woolf and Ney (1969) were the first to recognize that silicate grains, similar in mineral composition to materials in the Earth's crust and mantle, were a major constituent of the dust in oxygen-rich stars.

Circumstellar stardust around carbon-rich stars is composed primarily of amorphous carbon and silicon carbide (Gehrz 1989). The identification of the silicate and SiC material comes from 10 and 20μm emission features caused by the Si-O/Si-C stretching and O-Si-O bending molecular vibrational modes (see Figures 1 and 2). Silicates, having the triatomic molecule SiO_2 in their structure, show both the features. The diatomic molecule SiC cannot bend, and therefore exhibits only the μm stretching feature. The 10 and 20μm emission features in stellar objects are broad and generally devoid of the structure generally diagnostic of crystaline silicate minerals like Olivine and Enstatite (Rose 1979; Campins and Tokunaga 1987). This suggests that the silicateous minerals in stardust are amorphous, and that the grains probably have a considerable spread in size distribution.

Some astrophysical sources exhibit near infrared emission or absorption features in the 3.1 to 3.4μm spectral region (see Figure 3) that have been attributed to stretching vibrations in C-H molecular bonds associated with various hydrocarbon compounds (Allamandola 1984; Sakata et al. 1984; Allamandola et al. 1987; Allen and Wickramasinghe 1987). The hydrocarbon grain materials proposed to account for these features include polycyclic aromatic hydrocarbons (PAH's), hydrogenated amorphous carbon (HAC's), and quenched carbonaceous composites (QCC's). There is evidence for the presence of hydrocarbon grains in the near infrared spectra of a handful of stellar objects (see de Muizon et al. 1986; Gehrz 1989).

Observations have confirmed the existence of dust in circumstellar environments other than those associated with late-type stars. Dust is known to have condensed around novae (Gehrz 1988) and probably can
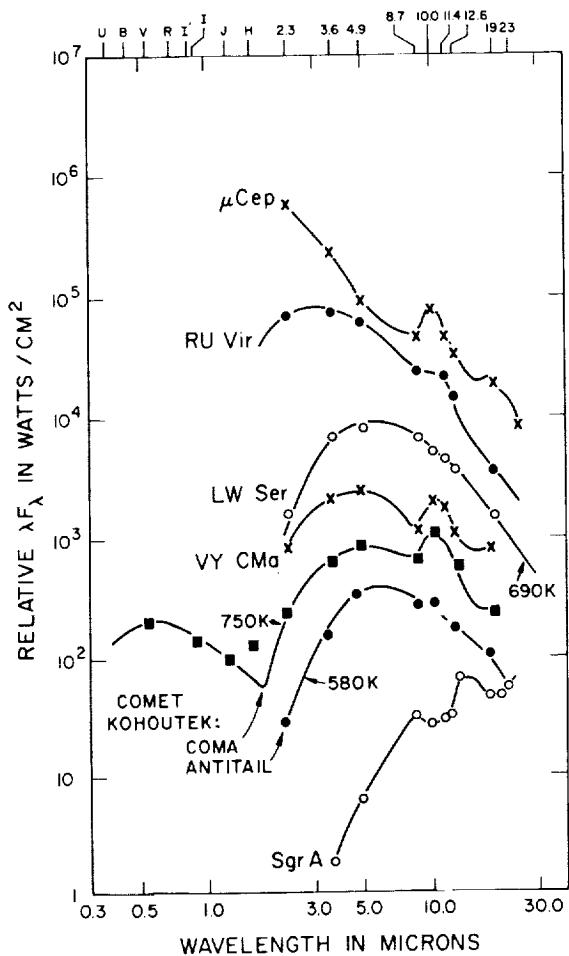
FIGURE 1 The infrared energy distributions of various objects that show emission or absorption due to small astrophysical dust grains. VY CMa and μ Cep are M supergiants (oxygen-rich stars) that have strong 10 and 20μm silicate emission features (Gehrz 1972). These same emission features appear in the comae of comets (Comet Kohoutek) where the grains are small (Ney 1974). The superheat in the coma dust continuum shows that the grains, probably amorphous carbon, producing this continuum are also small. The feature is weak in the Kouhoutek antitail because the grains are large (Ney 1974). General interstellar silicate absorption at 10 and 20μm is evident in the nonthermal spectrum of the Galactic Center source Sgr A (data from Hackwell *et al.* 1970). Carbon stars (RU Vir) often show a 11.3μm emission feature caused by SiC and a near infrared thermal continuum due to amorphous carbon (Gehrz *et al.* 1984). Novae (LW Ser) form carbon dust in their ejecta (Gehrz *et al.* 1980). The carbon produces a grey, featureless continuum from 2 to 23μm. The superficial similarities in the spectra of these objects is striking.

form in the ejecta from supernovae, though there is as yet no unambiguous observational evidence for supernova dust formation (Gehrz and Ney 1987).[1]

A small amount of dust is also present in the winds of planetary nebulae (Gehrz 1989) and Wolf-Rayet (WR) stars (Hackwell *et al.* 1979; Gehrz 1989). The Infrared Astronomical Satellite (IRAS) provided far infrared data that show evidence for the existence of faint, extended circumstellar dust shells around some main sequence (MS) stars (Aumann *et al.* 1984; Paresce and Burrows 1987). These shells, typified by those discovered around Vega (α Lyr) and β Pic, are disk-like structures believed to be fossil remnants of the star formation process. Although the material detected by IRAS around MS stars is most likely in the form of small and large grains, the presence of planets within the disks cannot be ruled out. The existing data are not spatially or spectrally detailed enough to lead to definitive conclusions about the mineral composition and size distribution of these fossil remnants of star/planetary system formation. There are large amounts of dust present in the circumstellar regions of many young stellar objects (YSO's), often confined in disk-like structures that are associated with strong bipolar outflows (Lada 1985). In the case of YSO's, it is unclear whether the dust is condensing in the wind or remains from the material involved in the collapse phase.

Most main-sequence stars and older YSO's do not have strong infrared excesses from dust shells, nor do they show evidence for visible extinction that would be associated with such shells. It is tempting to conclude that the shells in these objects have been cleared away in the early stages of stellar evolution by stellar winds, by Poynting-Robertson drag, or by the rapid formation of planets (see the contribution by Strom *et al.* in these proceedings). Rapid clearing of the circumstellar material poses a problem for the rather long time scale apparently required for the formation of giant planets (see the contribution by Stevenson in these proceedings). An alternative possibility is that dust grains grow to submillimeter or centimeter sizes (radii from 100 microns to 10 centimeters) during the contraction of the core to the Zero Age Main Sequence (ZAMS). Such grains will produce negligible extinction and thermal emission compared to an equal mass of the 0.1-10 micron grains that are believed to make up most of the material in circumstellar shells that reradiate a substantial fraction of the energy released by the central star. It can be shown that the opacity of a circumstellar shell of mass $M = N4\pi\rho a^3/3$ (where N is the number of grains in the shell, ρ is the grain density, and a is the grain radius) is

---

[1] SN 1987a is believed to have condensed dust grains about 400 days after its eruption. The dust formation is discussed in an analysis of recent infrared data by R.D. Gehrz and E.P. Ney (1990. Proc. Natl. Acad. Sci. 87:4354-4357).
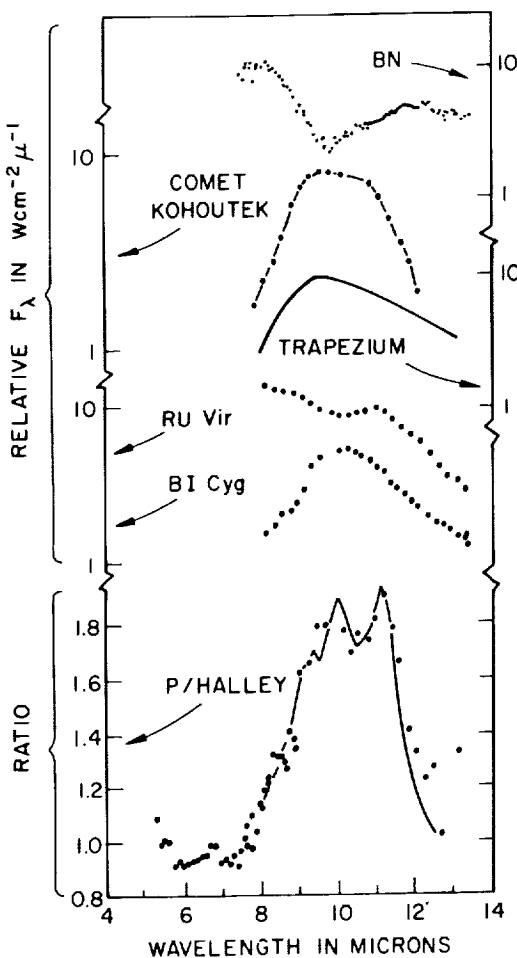
FIGURE 2 High-resolution infrared spectra of the 7-14 micron emission and absorption features of different astrophysical sources illustrating some basic differences between extra-solar-system sources and comets. The classical astrophysical 10μm silicate emission feature, typified by the Trapezium emissivity profile (Gillett *et al.* 1975) and the M-supergiant BI Cyg (Gehrz *et al.* 1984), peaks at 9.7μm. It is broad and without structure, suggesting that the grains are amorphous with a wide range of grain sizes. The feature appears in absorption in compact sources deeply embedded in molecular cloud cores such as the BN (Becklin-Neugebauer) object in Orion (Gillett *et al.* 1975). The carbon star RU Vir exhibits a classical 11.3μm SiC emission feature (Gehrz *et al.* 1984). Comet Kohoutek data are from Merrill (1974) as presented by Rose (1979). The 10μ emission feature of Kohoutec is similar to the classical astrophysical silicate feature. P/Halley's emission feature (Bregman *et al.* 1987) shows detailed structure suggesting that the grain mixture contains significant quantities of crystaline anhydrous silicate minerals. The solid line to the P/Halley data is a fit based on spectra of IDP's with Olivene and Pyroxene being the dominant components (Sandford and Walker 1985). The feature at 6.8μm may be due to carbonates or hydrocarbons.

inversely proportional to the grain radius if M is held constant. A shell of 0.1 micron grains would be reduced in opacity by a factor of 10^6 if the grains were accreted into 10 cm planetesimals while the total circumstellar dust mass remains constant. A 10 $L_\odot$ star would require the age of the solar system to clear 10 cm planetesimals from a circumstellar radius of 10 AU by Poynting-Robertson drag, and the sweeping effects of radiation pressure on 10 cm grains would be negligible. It would appear possible to postulate scenarios for the accretion of large circumstellar bodies that are consistent with both the relatively rapid disappearance of observable circumstellar infrared emission and the long time scales required for giant planet formation.

CIRCUMSTELLAR GRAIN FORMATION AND MASS LOSS

The observations described above suggest that many classes of evolved stellar objects are undergoing steady-state mass loss that injects stardust of various compositions into the interstellar medium. Gehrz (1989) has estimated the rates at which various grain materials are ejected into the ISM by different classes of stars. Stardust formation in most stars is a steady-state process, and the detailed physics of the grain formation is exceedingly difficult to resolve with current observational capabilities. Infrared studies of objects exhibiting outbursts that lead to transient episodes of dust formation, on the other hand, have revealed much about the formation of stardust and its ejection into the ISM. The long-term infrared temporal development of a single outburst is governed by the evolution of the grains in the outflow. Observations have shown that it is possible in principle to determine when and under what conditions the grains nucleate, to follow the condensation process as grains grow to large sizes, to record the conditions when grain growth ceases, and to observe behavior of the grains as the outflow carries them into the ISM.

The primary examples of transient circumstellar dust formation have been recorded in classical nova systems (Gehrz 1988) and WR Stars (Hackwell *et al.* 1979). In both cases, grains nucleate and grow on a time scale of 100 to 200 days, and the grains are carried into the ISM in the high-velocity outflow. The dust formation episodes apparently occur as frequently as every five years in WR stars and about once per 100-10,000 years in classical novae. About 10^{-6} to 10^{-5} solar masses of dust form in each episode, and the grains can grow as large as 0.1 to 0.3 microns. There is evidence that the grains formed in nova ejecta are evaporated or sputtered to much smaller sizes before they eventually reach the ISM. Novae have been observed to produce oxygen silicates, silicon carbide (SiC), amorphous carbon, and perhaps hydrocarbons (Gehrz 1988; Hyland and MacGregor 1989); WR stars apparently condense iron or amorphous carbon (Hackwell *et al.* 1979).
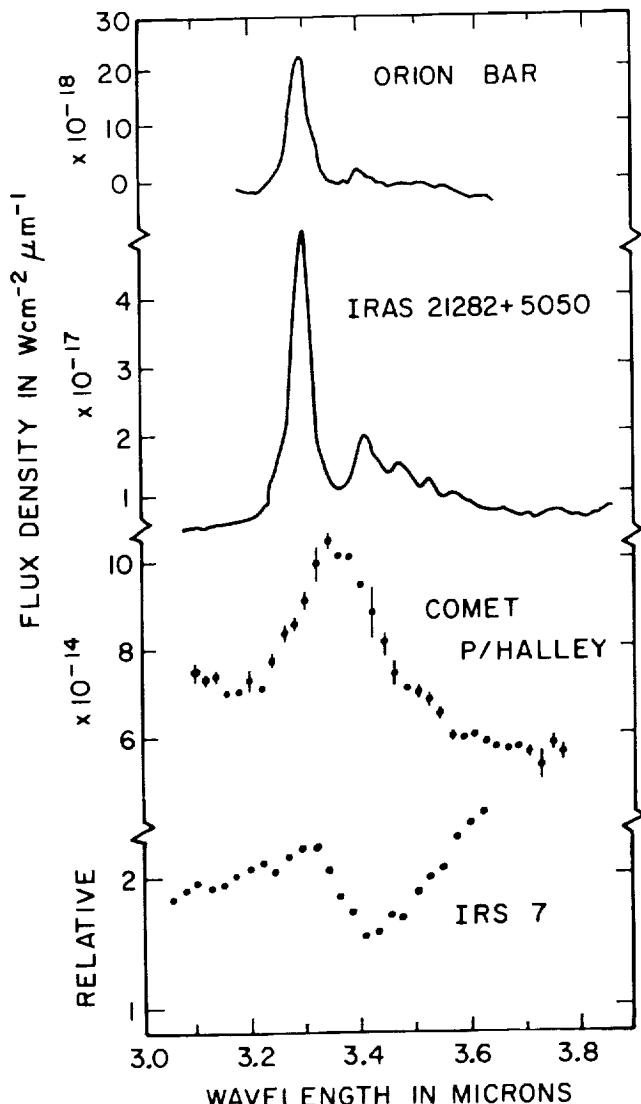
FIGURE 3 High-resolution infrared spectra of the 3.3-3.4μ hydrocarbon emission and absorption bands in three extra-solar-systems sources and Comet Halley. The Orion Bar is a shocked emission region in a molecular cloud; the Orion Bar curve is drawn after data from Bregman *et al.* (1986, in preparation) as shown in Figure 1 of Allamandola *et al.* (1987). IRAS 21282+5050 is a compact star-like object of undetermined nature (after data from de Muizon *et al.* 1986). Comet P/Halley (Knacke *et al.* 1986) has features that are broader and peak at longer wavelengths than the features of the comparison objects. Bottom curve shows a 3.5μm interstellar absorption feature in the spectrum of IRS 7, a highly reddened source near the Galactic Center (Jones *et al.* 1983).

Classical novae typify the episodic circumstellar dust formation process. Their infrared temporal development progresses in several identifiable stages. The initial eruption results from a thermonuclear runaway on the surface of a white dwarf that has been accreting matter from a companion star in a close binary system. Hot gas expelled in the explosion is initially seen as an expanding pseudophotosphere, or "fireball." Free-free and line emission are observed when the expanding fireball becomes optically thin. A dust condensation phase, characterized by rising infrared emission, occurs in many novae within 50 to 200 days following the eruption. The infrared emission continues to rise as the grains grow to a maximum radius. Grain growth is terminated by decreasing density in the expanding shell. The infrared emission then declines as the mature grains are dispersed by the outflow into the ISM. The rate of decline of the infrared radiation and the temporal development of the grain temperature suggest that the grain radius decreases either by evaporation or sputtering during their dispersal. Existing observations are consistent with the hypothesis that the nova grains could be processed to interstellar grain sizes before they reach the ISM.

Hydrocarbon molecules described above (PAH's, HAC's, QCC's) may produce some of the infrared emission features observed in planetary nebulae, comets, and molecular cloud cores (see Figure 3 and Allamandola *et al.* 1987). Although Hyland and MacGregor (1989) have reported possible hydrocarbon emission from a recent nova, and Gerbault and Goebel (1989) have argued that hydrocarbons may produce anomalous infrared emission from some carbon stars, there is currently no compelling evidence that hydrocarbon grains are an abundant constituent of the dust that is expelled into the ISM in stellar outflows (Gehrz 1989). Generally, circumstellar hydrocarbon emission is observed only in sources with high-excitation nebular conditions (Gehrz 1989). There is circumstantial evidence that grains condensed in the ejecta of novae, supernovae, and WR stars may contain chemical abundance anomalies similar to those in solar system meteorite inclusions (see below and Clayton 1982; Truran 1985; Gehrz 1988).

INTERSTELLAR DUST GRAINS

Hackwell *et al.* (1970) showed that the same 10- and 20-meter silicate features responsible for emission in M-stars were present in absorption in the infrared spectrum of the non-thermal Galactic Center source Sgr A. They concluded that the absorption was caused by interstellar silicate grains in the general ISM and that the grains are similar to those seen in M-stars. Interstellar silicate absorption has since been confirmed for a variety of other objects that are obscured by either interstellar dust or cold dust in molecular clouds (see, for example, the BN object in Figure 2 and the other objects embedded in compact HII regions discussed by Gillett *et al.* 1975).

There is roughly 0.04 mag of silicate absorption per magnitude of visual extinction to Sgr A. Observations of stellar sources with interstellar silicate absorption along other lines of sight in the Galaxy yield similar results.

Continuum reddening by interstellar dust in the general ISM has been measured by a number of investigators who compared optical/infrared colors of reddened luminous stars with the intrinsic colors exhibited by their unreddened counterparts (Sneden *et al.* 1978; Rieke and Lebofsky 1985). Interstellar dust also polarizes starlight (Serkowski *et al.* 1975). An estimate of the grain size distribution for the grains causing this so-called "general" interstellar extinction can be made given the wavelength dependence of the reddening and polarization curves. The same reddening and polarization laws appear to hold in all directions in the galaxy that are not selectively affected by extinction by dense molecular clouds. The reddening and polarization curves observed for stars deeply embedded in dark and bright molecular clouds lead to the conclusion that the grains in clouds are substantially larger than those causing the general interstellar extinction (Breger *et al.* 1981).

The shape of the general interstellar extinction curve as determined by optical/infrared measurements is consistent with the assumption that the ISM contains carbon grains of very small radii (0.01-0.03μm) and a silicate grain component (Mathis *et al.* 1977; Willner 1984; Draine 1985). Both components are also observed in emission in dense molecular cloud cores where the dust is heated by radiation from embedded luminous young stars (Gehrz *et al.* 1984). The ISM dust in dense clouds contains a probable hydrocarbon grain component that causes the 3.2-3.4μm emission features (see Figure 3), and several other "unidentified" infrared emission features that are seen in the 6-14μm thermal infrared spectra of some HII regions, molecular cloud cores, and young stellar objects (Allamandola 1984; Allamandola *et al.* 1987). Jones *et al.* (1983) showed that a 3.4μm C-H stretch absorption feature is present in the spectrum of the highly reddened source IRS7 towards the Galactic Center (see Figure 3). SiC has not been observed in either the general extinction or in the extinction/emission by molecular cloud grains, but its presence may be obscured by the strong silicate features.

At least some of the dust present in the ISM must be stardust produced by the processes described above. Gehrz (1989) has reviewed the probable sources for the production of the dust that is observed to permeate the ISM. These include condensation in winds of evolved stars, condensation in ejecta from nova and supernovae, and accretion in dark clouds. Most of the silicates come from M stars and radio luminous OH/IR (RLOH/IR) stars; carbon stars produce the carbon and SiC. Some stars, novae, and supernovae may eject dust with chemical anomalies. Since there is apparently

not a substantial stellar source of hydrocarbon grains, the ISM hydrocarbon component may be produced during grain processing or growth in molecular clouds. The observation that there are large grains in molecular clouds provides additional evidence that grains can grow efficiently in these environments.

COMET DUST AND THE ZODIACAL CLOUD

The thermal infrared energy distributions of the comae and dust tails of most comets (see Figures 1 and 2) show the characteristic near infrared continuum dust emission that is probably caused by small iron or carbon grains, and prominent 10 and 20m emission features characteristic of silicate grains (Ney 1974; Gehrz and Ney 1986). The cometary silicate features, first discovered in Comet Bennett by Maas et al. (1970), suggest that comets contain silicate materials similar to those observed in the circumstellar shells of stars and in the ISM. Determinations of the composition of Halley's coma grains by the Giotto PIA/PUMA mass spectrometers appear to confirm the silicate grain hypothesis (Kissell et al. 1986). There are some basic differences however, between the cometary 10m emission features and their stellar/interstellar counterparts (see Figure 2). As discussed above, the latter are broad and structureless suggesting a range of sizes and an amorphous grain structure. The 10μm feature in P/Halley, on the other hand, shows definite structure that suggests the presence of a grain mixture containing 90% crystalline silicates (55% olivines, 35% pyroxene) and only 10% lattice-layer silicates (Sandford and Walker 1985; Sandford 1987; Gehrz and Hanner 1987). This observation would imply that the grains in some comets may have undergone considerable high-temperature processing compared to extra-solar-system grains. Some pristine comets, like Kohoutek, show a 10m feature more like the stellar feature (see Figure 2 and Rose 1979). In the case of Kohoutek, the model fits indicate that the mineral composition is almost entirely low-temperature hydrated amorphous silicates (Rose 1979; Campins and Tokunaga 1987; Hanner and Gehrz 1987; Brownlee 1987; and Sanford 1987).

The 3 to 8μm thermal continuum radiation in the comae and Type II dust tails of most comets are often hotter than the blackbody temperature for the comet's heliocentric distance (Ney 1982). This "superheat" suggests that the grains are smaller than about 1 micron in radius. ISM grains may be 10 to 100 times smaller than this. The antitail of Kohoutek was cold with only weak silicate emission showing that comets also have much larger grains frozen in their nuclei (Ney 1974).

The 3.3-3.4μm feature (see Figure 3) in P/Halley suggests the presence of hydrocarbon grains in the ablated material. It is obvious that the Halley emission feature differs substantially both in width and effective wavelength

from the 3.3-3.4μm emission seen in other astrophysical sources. The implication is that the material in P/Halley has been processed in some way compared with the material observed in extra-solar-system objects.

Comets are presumably a Rosetta Stone for the formation of the solar system because the contents of their nuclei were frozen in the very early stages of the accretion of the solar nebula. The grains contained therein may be indicative of interstellar material in the primitive solar system, or may represent material processed significantly during the early collapse of the solar system. These materials are ablated from comet nuclei during perihelion passage. Cometary particles injected into the interplanetary medium by ablation probably produce the shower meteors and the zodiacal cloud. Studies of zodiacal dust particles may therefore provide important information about the properties of cometary dust grains. No cometary or zodiacal particles have yet been collected *in situ*. However, Brownlee (1978) and his recent collaborators have collected grains believed to be interplanetary dust particles (IDP's) from the stratosphere, on the Greenland glaciers, and off the ocean floor. These particles yield infrared absorption spectra showing that they are composed primarily of crystalline pyroxene and olivines, and layer-lattice silicates (Sandford and Walker 1985). Composite spectra modeled by combinations of these particles have been shown to match the Halley and Kohoutek data reasonably well (Sandford and Walker 1985; Sandford 1987). There is no evidence for 3.4μm hydrocarbon features in the laboratory spectra of IDP's, but there are spectral pecularities that are associated with carbonaceous minerals. Walker (1987) has analyzed the minerology of IDP's and concludes that they are comprised largely of materials that were formed in the solar system and contain only a small fraction of ISM material in the form of very small grains. The comparisons between IDP's and comets are, of course, only circumstantial at present. It is therefore crucial to contemplate experiments to collect zodiacal particles and cometary grains to establish the connections between zodiacal, cometary, and interstellar/circumstellar particles.

THE SURVIVAL OF DUST GRAINS DURING STELLAR EVOLUTION

An intriguing question is whether significant numbers of stardust grains can survive from the time that they condense in circumstellar outflows until they are accreted into the cold solid bodies in primitive planetary systems. While there is evidence that grains are rapidly destroyed in the ISM by supernova shocks (Seab 1987) and are heated above the melting point in the nebular phases of collapsing stellar systems (Boss 1988), there is equally compelling evidence for the existence of solar-system grains that are probably unaltered today from when they condensed long ago in circumstellar outflows (Clayton 1982). If stardust cannot survive long

outside the circumstellar environment, then most of the grains in evolved planetary systems must be those that grew in ISM molecular clouds and/or those that condensed in the collapsing stellar systems.

It is difficult to escape the conclusion that grains can be destroyed in the ISM. Seab (1987) has reviewed the possibility that ISM grains are rapidly destroyed by supernova shock waves on very short time scales (1.7-5 × 10^7 yrs). He notes however, that the extreme depletion of refractory grain materials in the ISM is difficult to reconcile with the destruction hypothesis. Assuming the gas mass in the Galactic ISM is 5 × 10^9 $M_\odot$ with a gas to dust ratio of 100/1, shocks will process 10-30 $M_\odot$ yr^{-1} in gas and destroy 0.1-0.3 $M_\odot$ yr^{-1} in dust.

Despite the rapid destruction that may befall ISM grains, it appears likely that they can reform and grow to large sizes in molecular clouds and star formation regions. Gehrz (1989) examined the galactic dust ecology by comparing circumstellar dust injection with depletion by star formation and supernova shocks and found that dust grains may be produced by accretion in molecular clouds at one to five times the circumstellar production rate. There is currently no observational evidence for a substantial stellar source of hydrocarbon grains. If this is the case, then the production of hydrocarbon grains and hydrocarbon mantles on stardust may be primarily an ISM process. Seab (1987) has predicted rapid accretion rates for grains in molecular clouds ($\approx 10^6$ – 10^7 years) which implies that such clouds could be efficient sources for the production and/or growth of dust in the ISM. Depletions in the ISM of heavy elements associated with dust (Jenkins 1987) are a strong indication that any gas-phase heavy elements ejected from stars are efficiently condensed onto ISM dust grains. Since early-type stars (WR stars, Of stars), and supernovae can eject a significant amount of gas-phase condensable matter on a galactic scale, it seems highly likely that these materials must be incorporated into dust in the ISM itself.

There is substantial evidence that there were high temperatures (1700-2000K) in the early solar nebula within a few AU of the sun (Boss 1988) so that most small refractory grains in this region would have melted or vaporized. On the other hand, microscopic inclusions in solar system meteorites exhibit abundance anomalies that may have been produced by the condensation of grain materials in the immediate vicinity of sources of explosive nucleosynthesis such as novae or supernovae (Clayton 1982). For example, ^{22}Ne can be made by the reaction ^{22}Na$(\beta^{+\nu})^{22}$Ne which has a half-life of only 2.7 years, so that ^{22}Ne now found in meteorite inclusions (the so-called Ne-E anomalie) must have been frozen into grains shortly after the production of ^{22}Na in a nova eruption (Truran 1985). Another anomaly seen in meteorite inclusions that can be produced in nova grains is excess ^{26}Mg from ^{26}Al$(\beta^{+\nu})^{26}$Mg which has a half-life of 7.3 × 10^5 years. Infrared observations have now revealed several novae in which

forbidden fine structure line emission has provided evidence for chemical abundance anomalies that would be associated with the production of ^{22}Na and ^{26}Al (Gehrz 1988). Xe can be produced by a modified r-process in nucleosynthesis supernovae and trapped in grains that condense in the ejecta (Black 1975). Some of these anomalies also imply that the grains survived intact from their sites of circumstellar condensation until their accretion into the body of the meteorite. High-temperature processing might be expected to drive off volatiles such as ^{22}Ne and Xe which are highly overabundant in some inclusions.

ESTABLISHING CONNECTIONS BETWEEN STARDUST AND DUST IN THE SOLAR SYSTEM

The investigation of the properties of astrophysical dust grains is an area that can benefit from studies that use the techniques of both astrophysics and planetary science. It is now possible to conduct both remote sensing and *in situ* experiments to determine with certainty the mineral composition and size distribution of the dust in the solar system. The Vega and Giotto flybys of Comet P/Halley produced some tantalizing results that demand confirmation. A sample return mission to a comet or asteroid is of the highest priority. Any mission that returns a package to Earth after a substantial voyage through the solar system should contain experiments to collect interplanetary dust particles. It will be important to establish whether IDP's that are collected from space resemble those collected on the Earth and whether they have chemical anomalies that are similar to those seen in meteorite inclusions. Examples of missions now planned that could be modified to include IDP dust collection are Comet Rendezvous and Asteroid Flyby (CRAF) and the MARS SAMPLE RETURN missions. The mineral composition of the zodiacal cloud remains uncertain. Infrared satellite experiments to measure the spectrum of the cloud can provide significant diagnostic information. Near infrared reflectance spectroscopy can reveal the presence, mineral composition, and size distribution of various types of silicate grains. These features have already been observed in the spectra of asteroids. Emission spectroscopy can determine whether silicates and silicon carbide are present. The contrast of the 10-20μm emission features are related to the size distribution (Rose 1979). Ground-based studies of the mineralogy of stardust and solar system dust also require high signal to noise high-resolution spectroscopy of the emission features shown in Figures 1 and 3 in a wide variety of sources. New improvements in infrared area detectors should make achievement of this objective realistic within the coming decade.

ACKNOWLEDGEMENTS

REFERENCES

Allamandola, L.J. 1984. Absorption and emission characteristics of interstellar dust. Pages 5-35. In: Kessler, M.F., and J.P. Phillips (eds.). Galactic and Extragalactic Infrared Spectroscopy. Reidel, Dordrecht.

Allamandola, L.J., A.G.G.M. Tielens, and J.R. Barker. 1987. Infrared absorption and emission characteristics of interstellar PAH's. Pages 471-489. In: Hollenbach, D.J., and H. A. Thronson, Jr. (eds.). Interstellar Processes. Reidel, Dordrecht.

Allen, D.A., and D.T. Wickramasinghe. 1987. Discovery of organic grains in comet Wilson. Nature 329:615.

Aumann, H.H., F.C. Gillett, C.A. Beichman, T. de Jong, J.R. Houck, F.J. Low, G. Neugebauer, R.G. Walker, and P.R. Wesselius. 1984. Discovery of a shell around Alpha Lyrae. Ap. J. (Letters) 278:L23-27.

Boss, A.P. 1988. High temperatures in the early solar nebula. Science 241:565-567.

Breger, M., R.D. Gehrz, and J.A. Hackwell. 1981. Interstellar grain size II: infrared photometry and polarization in Orion. Ap. J. 248:963-976.

Bregman, J.D., H. Campins, F.C. Witteborn, D.H. Wooden, D.M. Rank, L.J. Allamandola, M. Cohen, and A.G.G.M. Tielens. 1987. Airborne and ground-based spectrophotometry of comet P/Halley from 1-13 micrometers. Astron. Astrophys. 187:616-620.

Brownlee, D.E. 1978. Interplanetary dust: possible implications for comets and pre-solar interstellar grains. Pages 134-150. In: Gehrels, T. (ed.). Protostars and Planets. University of Arizona Press, Tucson.

Brownlee, D.E. 1987. A comparison of Halley dust with meteorites, interplanetary dust, and interstellar grains. Pages 66-67. In: Hanner, M. (ed.). Infrared Observations of Comets Halley and Wilson and Properties of the Grains. NASA Conf. Pub. 3004.

Campins, H., and A. Tokunaga. 1987. Infrared observations of the dust coma. Pages 1-15. In: Hanner, M. (ed.). Infrared Observations of Comets Halley and Wilson and Properties of the Grains. NASA Conf. Pub. 3004.

Clayton, D.D. 1982. Cosmic Chemical Memory. Q. J. R. Astron. Soc. 23:174-212.

de Muizon, M., T.R. Geballe, L.B. d'Hendecourt, and F. Baas. 1986. New emission features in the infrared spectra of two IRAS sources. Ap. J. (Letters) 306:L105-L108.

Draine, B.T. 1985. Tabulated Optical Properties of Graphite and Silicate Grains. Ap. J. Suppl. 87:587-594.

Gehrz, R.D. 1972. Infrared radiation from RV Tauri Stars I. An infrared survey of RV Tauri stars and related objects. Ap. J. 178:715-725.

Gehrz, R.D. 1988. The infrared temporal development of classical novae. Ann. Rev. Astron. Astrophys. 26: 377-412.

Gehrz, R.D. 1989. Overview of dust formation: sources of stardust in the galaxy. In: Allamandola, L., and A.G.G.M. Tielens (eds.). Interstellar Dust: Proceedings of IAU Symposium No. 135. Reidel, Dordrecht, in press.

Gehrz, R.D., and M. Hanner. 1987. Cometary dust composition.Pages 50-52. In: Hanner, M. (ed.). Infrared Observations of Comets and Halley and Wilson and Properties of Grains. NASA Pub. 3004.

Gehrz, R.D., and E.P. Ney. 1986. Infrared temporal development of P/Halley. Pages 101-105. In: Exploration of Halley's Comet. ESA SP-250 European Space Agency, Noordwijk.

Gehrz, R.D., and E.P. Ney. 1987. On the possibility of dust condensation in the ejecta of Supernova 1987a. Proc. Nat. Acad. Sci 84:6961-6964.

Gehrz, R.D., and N.J. Woolf. 1971. Mass loss from M stars. Ap. J. 165:285-294.

Gehrz, R.D., D.C. Black, and P.M. Solomon. 1984. The formation of stellar systems from interstellar molecular clouds. Science 224:823-830.

Gehrz, R.D., G.L. Grasdalen, and J.A. Hackwell. 1987. Infrared astronomy. Pages 53-80. In: Encyclopedia of Physical Science and Technology, vol. 2. Academic Press, New York.

Gehrz, R.D., G.L. Grasdalen, J.A. Hackwell, and E.P. Ney. 1980. The evolution of the dust shell of Nova Serpentis 1978. Ap. J. 237:855-865.

Gehrz, R.D., E.P. Ney, G.L. Grasdalen, J.A. Hackwell, and H.A. Thronson. 1984. The mysterious 10 micron emission feature of Nova Aquilae 1982. Ap. J. 281:303-312.

Gerbault, F., and J.H. Goebel. 1989. Carbon stars with Alpha-C:H emission. In: Allamandola, L., and A.G.G.M. Tielens (eds.). Interstellar Dust: Contributed Papers. Reidel, Dordrecht. NASA Conference Publication, in press.

Gillett, F.C., W.J. Forrest, K.M. Merrill, R.W. Capps, and B.T. Soifer. 1975. The 8-13 spectra of compact HII regions. Ap. J. 200:609-620.

Gilman, R.C. 1974. Planck mean cross-sections for four grain materials. Ap. J. Suppl. 28:397-403.

Hackwell, J.A., R.D. Gehrz, and G.L. Grasdalen. 1979. Dust formation around the Wolf-Rayet Star HD193793. Ap. J. 234:133-139.

Hackwell, J.A., R.D. Gehrz, and N.J. Woolf. 1979. Interstellar silicate absorption bands. Nature 227:822-823.

Hanner, M. 1987. Grain optical properties. Pages 22-48. In: Hanner, M. (ed.). Infrared Observations of Comets Halley and Wilson and Properties of the Grains. NASA Pub. 3004.

Hyland, A.R., and P.J. MacGregor. 1989. PAH emission from Nova Cen 1986. In: Allamandola, L., and A.G.G.M. Tielens (eds.). Interstellar Dust: Contributed Papers. Reidel, Dordrecht. NASA Conference Publication, in press.

Jenkins, E.B. 1987. Element abundances in the interstellar atomic material. Pages 533-559. In: Hollenbach, D.J., and H.A. Thronson, Jr. (eds.). Interstellar Processes. Reidel, Dordrecht.

Jones, T.J., A.R. Hyland, and D.A. Allen. 1983. 3m Spectroscopy of IRS7 towards the galactic centre. M.N.R.A.S. 205:187-190.

Kissel, J. et al. 1986. Composition of Comet Halley dust particles from Giotto observations. Nature 321:336-337.

Knacke, R.F., T.Y. Brooke, and R.R. Joyce. 1986. Observations of 3.2-3.6 micron emission features in Comet Halley. Ap. J. (Letters) 310:L49-L53.

Lada, C.J. 1985. Cold outflows, energetic winds, and enigmatic jets around young stellar objects. Ann. Rev. Astron. Astrophys. 23:267-317.

Maas, R.W., E.P. Ney, and N.J. Woolf. 1970. The 10-micron emission peak of Comet Bennett 1969i. Ap. J. (Letters) 160:L101-L104.

Mathis, J.S., W. Rumple, and K.H. Nordsieck. 1977. The size distribution of interstellar grains. Ap. J. 217:425-433.

Ney, E.P. 1974. Infrared Observations of Comet Kohoutek Near Perihelion. Ap. J. (Letters) 189:L141-L143.

Ney, E.P. 1982. Infrared observations of comets. Pages 323-340. In: Wilkening, L. (ed.). Comets. University of Arizona Press, Tucson.

Paresce, F., and C. Burrows. 1987. Broad-band imaging of the Beta Pictoris circumstellar disk. Ap. J. (Letters) 319: L23-25.

Rieke, G.H., and M.J. Lebofsky. 1985. The interstellar extinction law from 1 to 13 Microns. Ap. J. 288:618-621.

Rose, L.A. 1979. Laboratory simulation of infrared astrophysical features. Astrophys. and Space Sci. 65: 47-67.

Sakata, A., S. Wada, T. Tanabe, and T. Onaka. 1984. Infrared spectrum of the laboratory-synthesized quenched carbonaceous composite (QCC): comparison with the infrared unidentified emission bands. Ap. J. (Letters) 287:L51-54.

Sandford, S.A. 1987. The spectral properties of interplanetary dust particles. Pages 68-72. In: Hanner, M. (ed.). Infrared Observations of Comets Halley and Wilson and Properties of the Grains. NASA Pub. 3004.

Sandford, S.A., and R.M. Walker. 1985. Laboratory infrared transmission spectra of individual interplanetary dust particles from 2.5 to 25 microns. Ap. J. 291:838-851.

Seab, C.G. 1987. Grain destruction, formation, and evolution. Pages 491-512. In: Hollenbach, D.J., and H.A. Thronson, Jr. (eds.). Interstellar Processes. Reidel, Dordrecht.

Serkowski, K., D.S. Mathewson, and V.L. Ford. 1975. Wavelength dependence of interstellar polarization and ratio of total to selective extinction. Ap. J. 196:261-290.

Sneden, C., R.D. Gehrz, J.A. Hackwell, D.G. York, and T.P. Snow. 1978. Infrared colors and the diffuse interstellar bands. Ap. J. 223:168-179.

Thuran, J.W. 1985. Pages 292-306. Nucleosynthesis in novae. In: Arnett, W.D., and J.W. Thuran (eds.). Nucleosynthesis: Challenges and New Developments. University of Chicago Press, Chicago.

Walker, R.M. 1987. Comparison of laboratory determined properties of interplanetary dust with those of Comet Halley particles: What are comets made of? Pages 53-63. In: Hanner, M. (ed.). Infrared Observations of Comets Halley and Wilson and Properties of the Grains. NASA Conf. Pub. 3004.

Wickramasinghe, D.T., and D.A. Allen. 1986. Discovery of organic grains in comet Halley. Nature 323:44.

Willner S.P. 1984. Observed spectral features of dust. Pages 37-57. In: Kessler, M.F., and J.P. Phillips (eds.). Galactic and Extragalactic Infrared Spectroscopy. Reidel, Dordrecht.

Woolf, N.J., and E. P. Ney. 1969. Circumstellar emission from cool stars. Ap. J. (Letters) 155:L181-L184.

Late Stages of Accumulation and Early Evolution of the Planets

ANDREY V. VITYAZEV AND G.V. PECHERNIKOVA
Schmidt Institute of the Physics of the Earth

ABSTRACT

This article briefly discusses recently developed solutions of problems that were traditionally considered fundamental in classical solar system cosmogony: determination of planetary orbit distribution patterns, values for mean eccentricity and orbital inclinations of the planets, and rotation periods and rotation axis inclinations of the planets. We will examine two important cosmochemical aspects of accumulation: the time scale for gas loss from the terrestrial planet zone, and the composition of the planets in terms of isotope data. We conclude that the early beginning of planet differentiation is a function of the heating of protoplanets during collisions with large (thousands of kilometers) bodies. This paper considers energetics, heat mass transfer processes, and characteristic time scales of these processes at the early stage of planet evolution.

INTRODUCTION

Using the theory of preplanetary cloud evolution and planet formation, which is based on the ideas of Schmidt, Gurevich, and Lebedinskiy, and which was developed in the works of Safronov (1969, 1982), we can estimate a number of significant parameters for the dynamics of bodies which accumulate in the planets. However, it long proved impossible to solve a number of problems in classical solar system cosmogony which were traditionally considered fundamental. Such problems include the theoretical derivation of patterns of planetary and satellite orbit distributions (the so-called Titius-Bode law), theoretical estimations of the value for

mean orbital eccentricities and inclinations, planet rotation periods and rotation axis inclinations, and other characteristics of the present structure of the Sun's planetary system. In addition, certain consequences of the theory and, most importantly, a conclusion on the relatively cold initial Earth and the late beginning of its evolution, clashed with data on geo- and cosmochemistry. These data are evidence of the existence of planet heating during the process of growth and very early differentiation. This paper briefly discusses a modified version of the theory which the authors developed in the 1970's and 1980's. Using this new version we were able to provide a fundamental solution to several key problems of planetary cosmogony and generate a number of new findings. The most significant of these appear to be an estimate of the composition of an accumulating Earth with data incorporated on oxygen isotopy and a conclusion on the early beginning of differentiation in growing planets.

FORMATION OF THE PLANETARY SYSTEM

Despite promising data on the existence of circumstellar disks, we have yet to discover an analogue to a circumsolar gas-dust disk. Nor have calculations (Ruzmaikina and Maeva 1986; Cassen and Summers 1983) produced a satisfactory picture of circumsolar disk formation, or a reliable estimate of its mass M and characteristic initial dimensions R*. Nevertheless, a model of a low-mass disk (M < 0.1 $M_\odot$), with moderate turbulence, a hot circumsolar zone and cold periphery, has received the widest recognition in the works of a majority of authors.

STANDARD AND MK—DISK MODELS

A reconstruction of surface density distribution $\sigma(R)$ according to Weidenshilling (1977) is shown in Figure 1. Mass of the disk, generated by adding on to produce the cosmic (solar) composition of present-day planet matter, is

$$(0.01 - 0.07)M_\odot, \text{with } \sigma(R > R_o = 1AU) \propto (R/R_o)^{-3/2}.$$

It is usually supposed that by the time the Sun achieved main sequence, its luminosity L_* did not greatly differ from the present $L_\odot$, and the temperature in the disk's central plane (z = 0) was on the order of a black body $T \simeq 300(R_o/R)^{1/2}K$. Estimations of the degree of ionization $(N_e/N \sim 10^{-11})$ and gas conductivity ($\lesssim 10^3$CGSE) in the central plane are insignificant, and the impact of the magnetic field is usually neglected in considering subsequent disk evolution. Using the hydrostatic equations
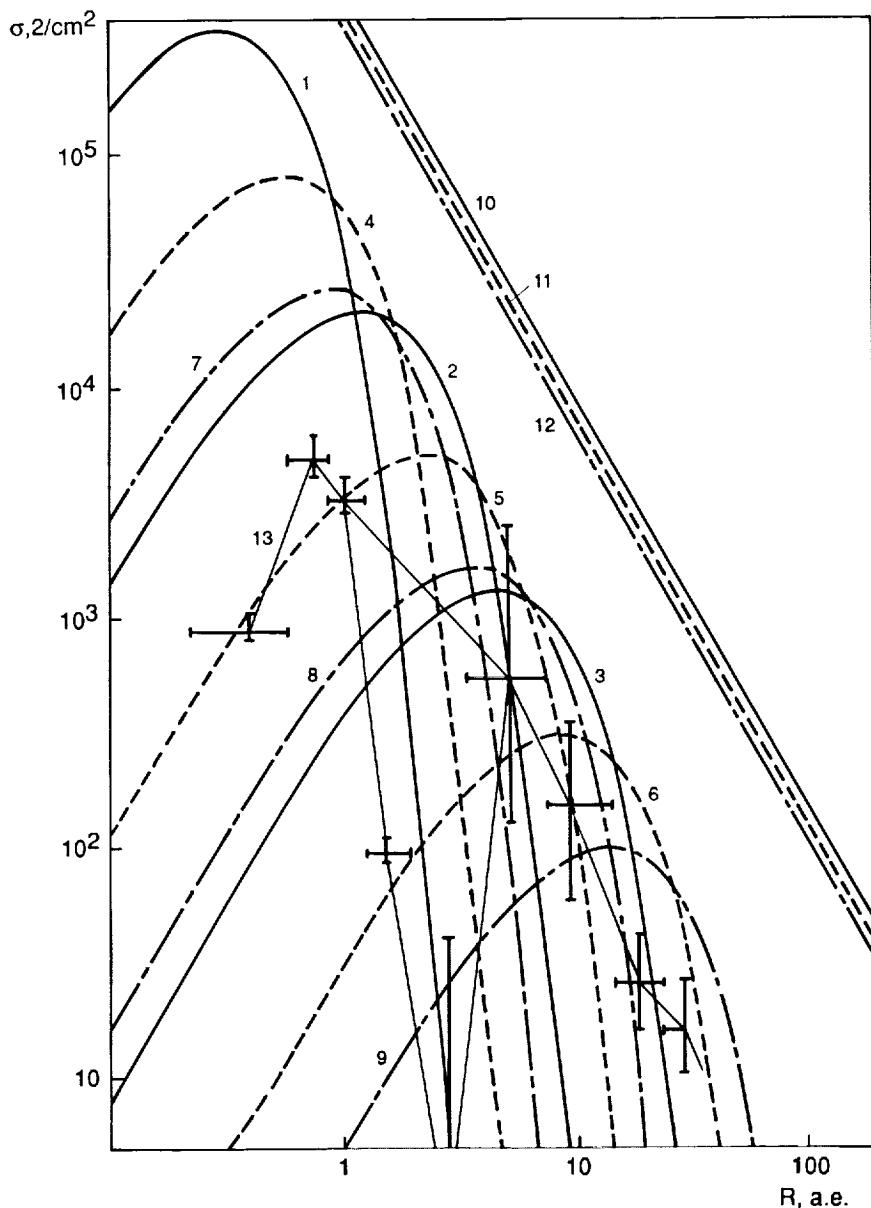
FIGURE 1 Surface density distribution $\sigma(R)$ in low-mass, circumstellar gaseous disks. The solid line indicates models of disks near F5 class stars: 1: m/k = 1; 2: m/k = 1/2; 3: m/k = 1/4; the dashed line indicates models of disks near G0 class stars: 4: m/k = 1; 5: m/k =1/2; 6: m/k =1/4; the dotted and dashed line shows disks near G5 class stars: 7: m/k = 1; 8: m/k = 1/2; 9: m/k = 1/4; the straight lines are critical density distributions for the corresponding classes of stars: 10: σ_{cr}(F5); 11: σ_{cr}(G0); 12: σ_{cr}(G5). Disk mass in models 1-12 is equal to M = $5 \cdot 10^{-2} M_{\odot}$. 13 is the standard model.

and the equation of the state of ideal gas at a temperature which is not dependent on z, we yield a density distribution for z:

$$\rho(z) \approx \rho_o(R)\exp(-z^2/h^2), \quad h^2 = 2kTR^3/GM_\odot\mu, \tag{1}$$

where $\mu(\approx 2.3)$ is the mean molecular mass, and k is the Boltsman constant. The model which has been termed standard is derived, by taking into account $\sigma(R) \simeq \sqrt{\pi}\,\rho_o(R)h(R)$:

$$\rho_o(R) \propto R^{\alpha-\beta}, P(R) \propto R^{-\beta}, T(R) \propto R^{-\alpha}, \alpha \lesssim 1, \beta \simeq 3. \tag{2}$$

Flattening of the disk is high: $\gamma = h/R \propto (RT)^{1/2}$, $\gamma \lesssim 0.1$. Disk rotation is differential and differs little from Kepler's:

$$\omega = \omega_k(1+\epsilon)^{1/2}, \quad \omega_k = V_k/R = \sqrt{GM_\odot/R^3}, \tag{3}$$

$$\epsilon \approx (c_s^2/V_k^2)(\text{d ln P/d ln R}) \lesssim 0.1, \quad c_s^2 = kT/\mu.$$

Here c_s is the speed of sound. In the standard model $V_k^2 \gg c_s^2 \gg v_A^2$, v_A is the Alfven velocity. Quasiequilibrium disk models are constructed by Vityazev and Pechernikova (1982) which do not use the contemporary distribution $\sigma(R)$. They were called MK-models since they are only defined by a mass M and a moment K of a disk which is rotating around a star with mass M_* and luminosity L_*. Expressions for densities $\rho(R,z)$ and $\sigma(R)$ were obtained by resolving the system of hydrodynamic equations with additional conditions (low level of viscous impulse transport and minimum level of dissipative function). The distribution of surface density in disks was found to be:

$$\sigma(R) \approx 1.55 \cdot 10^5 \tilde{m} \left(\frac{\tilde{m}}{k}\right)^{7.5} \left(\frac{M_*}{M_\odot}\right)^{12.25}$$

$$R^{1.75}\exp[-2.8\left(\frac{\tilde{m}}{k}\right)^2 \left(\frac{M_*}{M_\odot}\right)^3 R]g/cm^2, \tag{4}$$

where M is $\tilde{m}\cdot 10^{-2}M_\oplus$, K is $k\cdot 10^{51}g\ cm^2s^{-1}$, and R is in AU. By varying $\tilde{m}$ and k, star mass M_*, and luminosity L_*, we can generate a set of models of quasiequilibrium circumstellar disks (see Figure 1). We will note that the distributions of σ (R) in the standard and MK-models are qualitatively similar. However, there is a noticeable excess of matter in the remote zone in the first model in comparison to the latter ones. This excess may be related to diffusion spread of the planetesimal swarm during planet accumulation. Therefore, present planetary system dimensions may
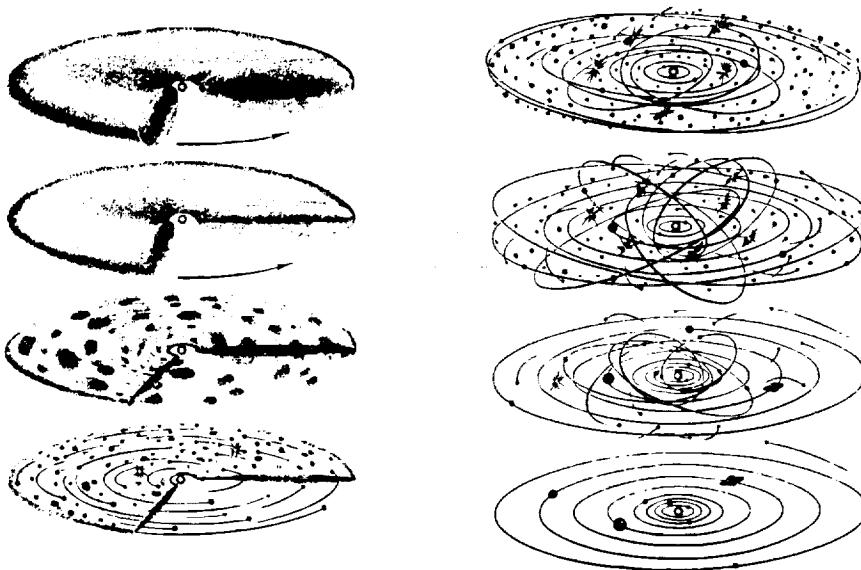
FIGURE 2 Evolution of the preplanetary disk. The left side shows flattening of the dust subdisk and the formation of the swarm of planetesimals. The right side illustrates plantesimals joining together to form planets (Levin 1964).

primary stages of its evolution and of planet formation (Figure 2) within the framework of the aforementioned low-mass disk models.

PLANETESIMAL MASS SPECTRUM N(M,T) AND MATTER REDISTRIBUTION

After dust settles on the central plane and dust clusters are formed due to gravitational instability, there occurs the growth and compacting of some clusters, and the breakdown and absorption of others. This process is described in detail by Pechernikova and Vityazev (1988). We later briefly touch upon the specific features of the final stages of accumulation of sufficiently large bodies, when a stabilization effect develops for the orbits of the largest bodies (Vityazev *et al.* 1990). In the coagulation equation

$$\frac{\partial n(m,t)}{\partial t} = \frac{1}{2} \int_0^m A(m^1, m - m^1) n(m^1) n(m - m^1) dm^1 - n(m,t)$$

$$\int_0^m A^{max}(m, m^1) n(m^1) dm^1, \tag{5}$$

the subintegral kernel $A(m,m^v)$, describing collision efficiency, must take into account the less efficient diffusion of large bodies. Let us write $A(m,m^v) = A^*B/(A^* + B)$, where

$$A^* = \tau_1(r + r^1)^2[1 + 2G(m + m^1)/(r + r^1)V^2]V = v_{ik}r_{ik},$$

is the usual coefficient which characterizes the collision frequency of gravitating bodies m and m^v with good mixing,

$$B = [D(m) + D(m^v)][\Delta R(m) + \Delta R(m^v)] = D_{ik}\Delta R_{ik},$$

is the coefficient accounting for diffusion, $D(m) = e^2(m)R^2/\tau_E(m)$, e is the eccentricity, τ_E is the characteristic Chandrasekhar relaxation time scale, $\Delta R = eR$. Estimates demonstrate that at the initial stages $(m \ll 0.1\ m_\oplus)$ B $\gg A^*$ and $A \rightarrow A^*$, while at the later stages $(m \gtrsim 0.1m_\oplus)$ we have B(i $\sim$ k) < A^*(i $\sim$ k) and in the limit A(i $\sim$ k) $\rightarrow$ B(i $\sim$ k). Pechernikova (1987) showed that if $A \propto (m^\alpha + m^{1\alpha}) \equiv m_{ik}{}^\alpha$ then the coagulation formula has an asymptotic solution that can be expressed as

$$n(m) \propto m^{-q}, q = 1 + \alpha/2. \tag{6}$$

For the initial stages $\alpha = 2/3 - 4/3$ and $q_* = 4/3 - 5/3$. For the later stages B(i $\sim$ k) $\propto m_{ik}{}^{-5/3}$ and $q^* \sim 1/6$, $q^* < q < q_*$, that is, the gently sloping power law spectrum for large bodies. With this finding we can understand the relative regularity of mass distribution in the planetary system. Unlike the findings of numerical experiments (Isaacman and Sagan 1977), only the low mass in the asteroid belt and a small Mars is an example of significant fluctuation. The authors, using numerical integration of equations, such as

$$\sigma_d = \frac{-1}{R}\frac{\partial}{\partial R}[Rv_R\sigma_d] + \frac{3}{R}\frac{\partial}{\partial R}\left[[\sqrt{R}\frac{\partial}{\partial R}\sqrt{R}D\sigma_d\right], \tag{7}$$

also examined the overall process of solid matter surface density redistribution $\sigma_d(R,t)$ $(\sigma_d \sim 10^{-2}\sigma)$ resulting from planetesimal diffusion. The spread effect of a disk of preplanet bodies proved significant for the later stages and was primarily manifested for the outer areas (Vityazev *et al.* 1990). The effect is less appreciable for the zone of the terrestrial planet group, and we shall forego detailed discussion of it here.

RELATIVE VELOCITY SPECTRUM

Wetherill's numerical calculations for the terrestrial planet zone (1980) confirmed the order of value of relative velocities that had been estimated earlier by Safronov (1969). Similar calculations for the zone of outer planets

are preliminary. In particular, there is vagueness in the growth time scales for the outer planets. By simplifying the problem for an analytical approach, we can look separately at the problem of mean relative velocities $\bar{v}$ (i.e. $\bar{e}$, $\bar{i}$) of the planetesimals and the problem of eccentricities $\bar{e}$ and inclination $\bar{i}$ of the orbits of accreting planets. The second problem is considered below. We will discuss here an effect which is important in the area of giant planets. As planetesimal masses (m) grow, their relative velocities (v) also increases. At a sufficiently high mean relative velocity ($\bar{v} \sim 1/3 \cdot V_k$) part of the bodies from the "high velocity Maxwell tail" may depart the system, carrying away a certain amount of energy and momentum. With this scenario, the formula for the mean relative velocity (Vityazev *et al.* 1990; Safronov 1969) must appear as follows:

$$\frac{1}{v^2}\frac{dv^2}{dt} = \frac{\beta_1 - \nu_E}{\tau_E} - \frac{1}{\tau_g} - \frac{\beta_2}{\tau_s}, \tag{8}$$

where τ_E, τ_g and τ_s are correspondingly the characteristic Chandrasekhar relaxation time scales, gas deceleration and the characteristic time scale between collisions, $\beta_1 1 = 0.05-0.13$ (Safronov 1969; Stewart and Kaula 1980), $\beta_2 = 0.5$ (Safronov 1969) and ν_E is part of the amount of energy removed by the "rapid particles"

$$\nu_E \approx \frac{1}{6}\int_{v_{cr}}^{\infty} v^2 n(v)dv / \int_0^{\infty} v^2 n(v)dv = \Gamma(5/2, b)/6\Gamma(5/2) \approx \tag{9}$$

$$\approx e^{-b}\left(\frac{1}{b^{1/2} + \sqrt{2+b}} + b^{1/2} + \frac{2}{3}b^{3/2}\right)/3\sqrt{\pi}, b = \frac{3}{2}\frac{\overline{v_{cr}^2}}{\overline{v}^2} \approx \frac{3}{2}\frac{(\sqrt{2}-1)^2\overline{v_k^2}}{\overline{v}^2}.$$

For now $e \sim i \sim v/V_k \ll 1$, the usual expressions for $\bar{v}$ follow from (8), in particular, with $\tau_g \gg \tau_E$, τ_S in a system of bodies of equal mass m we have

$$\bar{v} = \sqrt{Gm/\Theta r}, \text{where } \Theta \sim 1. \tag{10}$$

Despite continuing growth of the mass, as e approaches $e_{cr} \approx 0.3 - 0.4$, ν_E becomes comparable to β_1, the relative velocity $\bar{v}$ in the system of remaining bodies discontinues growth. In other words, with $e \approx e_{cr}$, the parameter Θ in expression (10) grows with m proportionally to $m^{2/3}$, reaching in the outer zone values $\sim 10^2$. This effect gives us acceptable time scales for outer planet growth. Because of low surface density, accreting planets in the terrestrial group zone cannot attain a mass sufficient with a build up of $\bar{e}$ to e_{cr}. Therefore the mean eccentricities do not exceed the values $e_{max} = 0.2 - 0.25$.

MASSES OF THE LARGEST BODIES IN
A PLANET'S FEEDING ZONE

Studies on accumulation theory previously assumed that in terms of mass a significant runaway of planet embryos from the remaining bodies in the future planet's feeding zone occured at an extremely early stage. According to Safronov's well-known estimates (1969), the mass of the largest body (after the embryo) m_1 was $10^{-2}-10^{-3}$ of the mass m of the accreting planet. Pechernikova and Vityazev (1979) proposed a model for expanding and overlapping feeding zones, and they considered growth of the largest bodies. The half-width of a body's feeding zone is determined by mean eccentricity $\bar{e}$ of orbits of the bulk of bodies at a given distance from the Sun:

$$\Delta R(t) \simeq e(t)R \simeq \bar{v}(t)R/V_k, \tag{11}$$

The characteristic mixing time scale for R in this zone virtually coincides with the characteristic time scale for the transfer of regular energy motion to chaotic energy motion, and the characteristic time scale for energy exchange between bodies. It is clear from (10) and (11) that

$$\bar{v}(t) \propto r(t), \bar{e}(t) \propto r(t), \Delta R(t) \propto r(t). \tag{12}$$

The mass of matter in the expanding feeding zone

$$Q(R,t) = 2\pi \int_{R-\Delta R}^{R+\Delta R} \sigma_d R \, d R, \tag{13}$$

will also grow with a time scale $\propto$ r(t), while disregarding the difference in matter diffusion fluxes across zone boundaries. When the mass of a larger body m(t) begins to equal an appreciable amount Q(t) with the flow of time (see Figure 3), the growth of the feeding zone decelerates. At this stage the larger bodies of bordering zones begin to leave behind, in terms of their mass, the remaining bodies in their zones. However, (as seen in Figure 3), this runaway by mass is much less than was supposed in earlier studies.

MASSES, RELATIVE DISTANCES, AND THE NUMBER OF PLANETS

Mass increase of the largest body in a feeding zone is represented by the well-known formula:

$$dm/dt = \pi r^2(1 + 2\Theta)\rho_d\bar{v} = 2(1 + 2\Theta)r^2 W_k \sigma_d, \tag{14}$$

where the surface density of condensed matter σ_d can be considered a sufficiently smooth function R and it can be assumed that $\sigma_d(t=0) =$
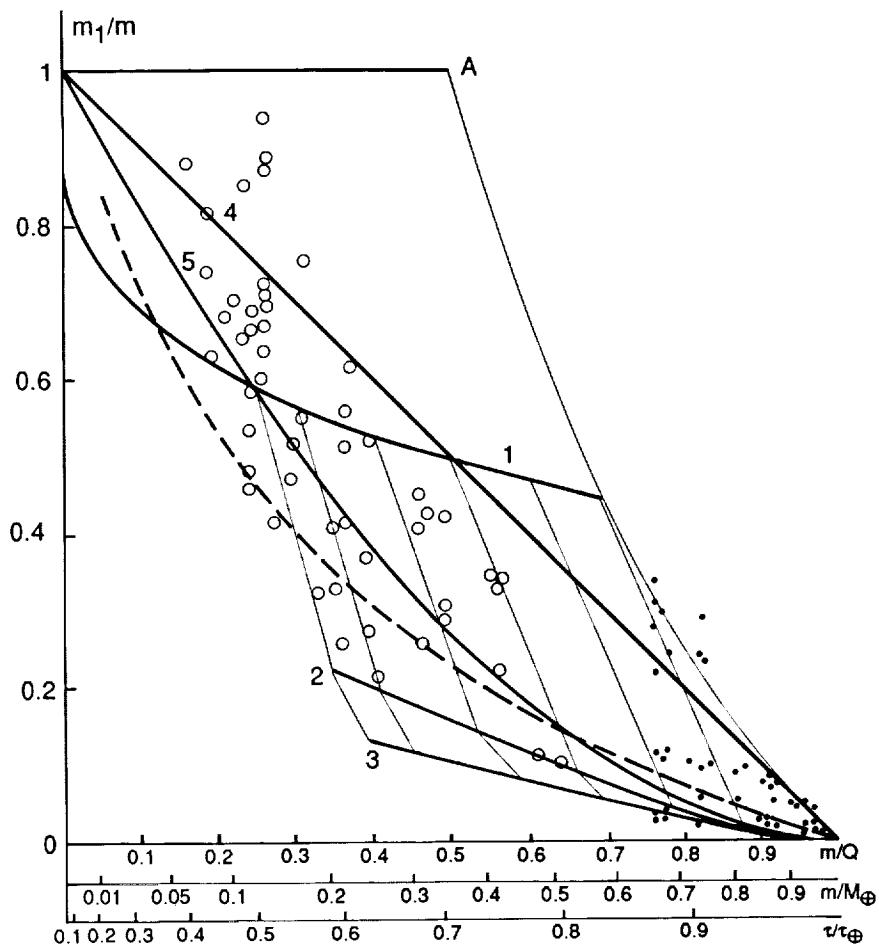
FIGURE 3 The region of determination and model distributions for m_1/m as the ratio of the mass, m_1, of the largest body in the feeding zone of a growing planet to the mass of a planet m: 1: $m_1/m = 1 - 0.62 \, (m/Q)^{0.3}$, which corresponds to the growth of bodies in the expanding feeding zone; 2: m_2/m; 3: m_3/m; 4: $m_1/m = 1 - m/Q$; 5: $m_1/m = (1 - m/Q)^2$. The circles and dots indicate the results of numerical simulation of the process of terrestrial planet accumulation (Ipatov 1987; Wetherill 1985).

$\sigma_o(R_o/R)^\nu$ with its values σ_o and v in each zone. As bodies precipitate to the largest body in a given zone m, a growing portion of matter is concentrated in m and the corresponding decrease in surface density is written as:

$$\sigma_d(t) = \sigma_d(t = 0)[1 - m(t)/Q(t)]. \tag{15}$$

From (10) through (15) for the growth rate of a planet's radius we have (with $\nu \neq 2$):

$$\frac{dr}{dt} = \frac{(1 + 2\Theta)\sigma_o W_k}{2\pi\delta} \left(\frac{R_o}{R}\right)^v$$

$$\left\{1 - \frac{2\delta(2 + v)r^3(t)(R/R_o)^v}{3\sigma_o R^2[1 + \bar{e})^{2+v} - (1 - \bar{e})^{2+v}]}\right\}. \tag{16}$$

It follows from (16) that the largest body which is not absorbed by the other bodies (a planet) ceases growing when it reaches a certain maximum radius (mass). The value of this radius is only determined by the parameters of the preplanetary disk. If we put a zero value in the bracket in (16), in the first approximation for $\bar{e}$, we can yield max r, max m, max $\bar{e}$, and max ΔR. In particular,

$$\text{max r} = \left(\frac{5\pi}{\Theta\delta M_\odot}\right)^{1/4} \cdot \sqrt{2\sigma_d(R,O)}r^{5/4}, \text{cm};$$

$$\text{max } \bar{e} = \frac{2}{3}\left(\frac{5\pi\delta R}{\Theta M_\odot}\right)^{1/2} \cdot \text{max r}. \tag{17}$$

The growth time scale is an integration (16). It is close to the one generated by Safronov (1969) and Wetherill (1980) and is on the order of 65-90 million years for accreting 80-90% of the mass.

One can state the following for distances between two accreting planets:

$$R_{n+1} - R_n \approx \Delta_n R + \Delta_{n+1} R = \bar{e}_n R_n + \bar{e}_{n+1} R_{n+1}, \tag{18}$$

hence

$$R_{n+1}/R_n \simeq (1 + \bar{e}_n)/(1 - \bar{e}_{n+1}) = b. \tag{19}$$

In view of (17) the following theoretical estimate can be made for terrestrial planets: b(max $\bar{e}$ = 0.2–0.25) = 1.5–1.67. For the zone of outer planets b(max $\bar{e}$ = e_{cr} = 0.32–0.35) = 1.85–2.3. The real values b in the present solar system are cited in Table 1. The theory that was developed not only explains the physical meaning of the Titius-Bode law, but also provides a satisfactory estimate of parameter b. The partial overlapping of zones,

TABLE 1 The real values b in the present solar system

	Venus- Mercury	Earth- Venus	Mars- Earth	Asteroids (Ceres)- Mars	Jupiter- Ceres	Saturn- Jupiter	Uranus- Saturn	Neptune- Uranus
b	1.87	1.38	1.52	1.77	1.88	1.83	2.01	1.57

embryo drift, and the effects of radial redistribution of matter (Vityazev *et al.* 1990) complicates the formulae, but this does not greatly influence the numerical values max m, max $\bar{e}$, and b.

An estimate of the number of forming planets can easily be generated from (19) for a preplanetary disk with moderate mass and distribution $\sigma(R)$ according to the standard or MK-models with pre-assigned outer and inner boundaries:

$$N = \frac{\ln(R^*/R_*)}{\ln[(1 + \max \bar{e})/(1 - \max \bar{e})]} \tag{20}$$

With low values of max $\bar{e}$ from (20) we have $N \simeq \ln(R^*/R_*)/2 \max \bar{e}$ and yield a natural explanation of the results of the numerical experiment (Isaacman and Sagan 1977): $N \propto 1/e$. For a circumsolar disk with initial mass $\lesssim 0.1 M_\odot$, the theory offers a satisfactory estimate of the number of planets which formed:

$$N(R^* \lesssim 10^2 AU, \ R_* \gtrsim 10^{-1} AU, \ \max \bar{e} \simeq e_{cr} = 1/3) \lesssim 10.$$

DYNAMIC CHARACTERISTICS OF FORMING PLANETS

Workers were long unsuccessful in developing estimates of planet eccentricities, orbital inclinations, and mean periods of axis rotation, which were formed during the planets' growth process. In other words, estimates that fit well with observational data. Perchernikova and Vityazev (1980, 1981) demonstrated that by taking into account the input of large bodies, the existing discrepancy between theory and observations could be resolved. When accreting planets approach and collide with large bodies at the earliest stages $\bar{e}$ increases. A rounding off of orbit takes place at the final stage: by the time accumulation is completed estimated values for $\bar{e}$ are close to present "mean" values (Laska 1988). The same can be said for orbital inclinations. Vityazev and Pechernikova (1981) and Vityazev *et al.* (1990) developed a theory for determining mean axial spin periods and axis

inclinations. The angular momentum vector for the axial spin of a planet K, inclined at an angle ϵ to axis z (z is perpendicular to the orbital plane), is equal to the sum of the regular component K_1, which is directed along axis z and the random component K_2, which is inclined at an angle γ to z. For K_1 in a modified Giuli-Harris approximation (for terrestrial planets) the following was generated:

$$K_1 = \frac{48}{\pi}\sqrt{\frac{2\Theta G}{5}}\left(\frac{2M_\odot}{R^3}\right)^{1/4}\left(\frac{3}{4\pi\bar{\rho}}\right)^{5/12} \cdot$$

$$m_p^{5/3}F_1(m/Q), \quad F_1(m/Q = 1) = 9.6\cdot 10^{-2}. \tag{21}$$

Dispersion K_2 is (q = 11/6):

$$DK_2 \approx 4.18\cdot 10^{-2}\bar{\rho}^{-1/3}Gm_p^{10/3}\cdot F_2(m/Q), \quad F_2(m/Q = 1) = 0.123. \tag{22}$$

The authors demonstrated that as planets accumulate what takes place is essentially direct rotation ($\bar{\epsilon}\lesssim\bar{\gamma}< 90°$). Large axis rotation inclinations and reverse rotation of individual planets are a natural outcome of the accumulation of bodies of comparable size. It is clear from (21) and (22) that the theoretical dependence of the specific axis rotation mementum ($\propto m^{2/3}$) approaches what has been observed. It is worth recalling that this theory does not allow us to determine the direction and velocity at which a planet, forming at a given distance, will rotate. It only gives us the corresponding probability (Figure 4).

COSMOCHEMICAL ASPECTS OF EVOLUTION

Vityazev and Pechernikova (1985) and Vityazev *et al.* (1989) have repeatedly discussed the problem of fusing physico-mechanical and physico-chemical approaches in planetary cosmogony. We will only mention two important findings here. Vityazev and Pechernikova (1985, 1987) proposed a method for estimating the time scale for gas removal from the terrestrial planet zone. They compared the theory of accumulation and data on ancient irradiation by solar cosmic rays of the olivine grains and chondrules of meteorite matter with an absolute age of 4.5 to 4.6 billion years.

TIME SCALE FOR THE REMOVAL OF GAS FROM THE TERRESTRIAL PLANET ZONE

It is easy to demonstrate that if gas with a density of at least 10^{-2} of the original amount remains during the formation of meteorite parent bodies in the preplanetary disk between the asteroid belt and the Sun, then
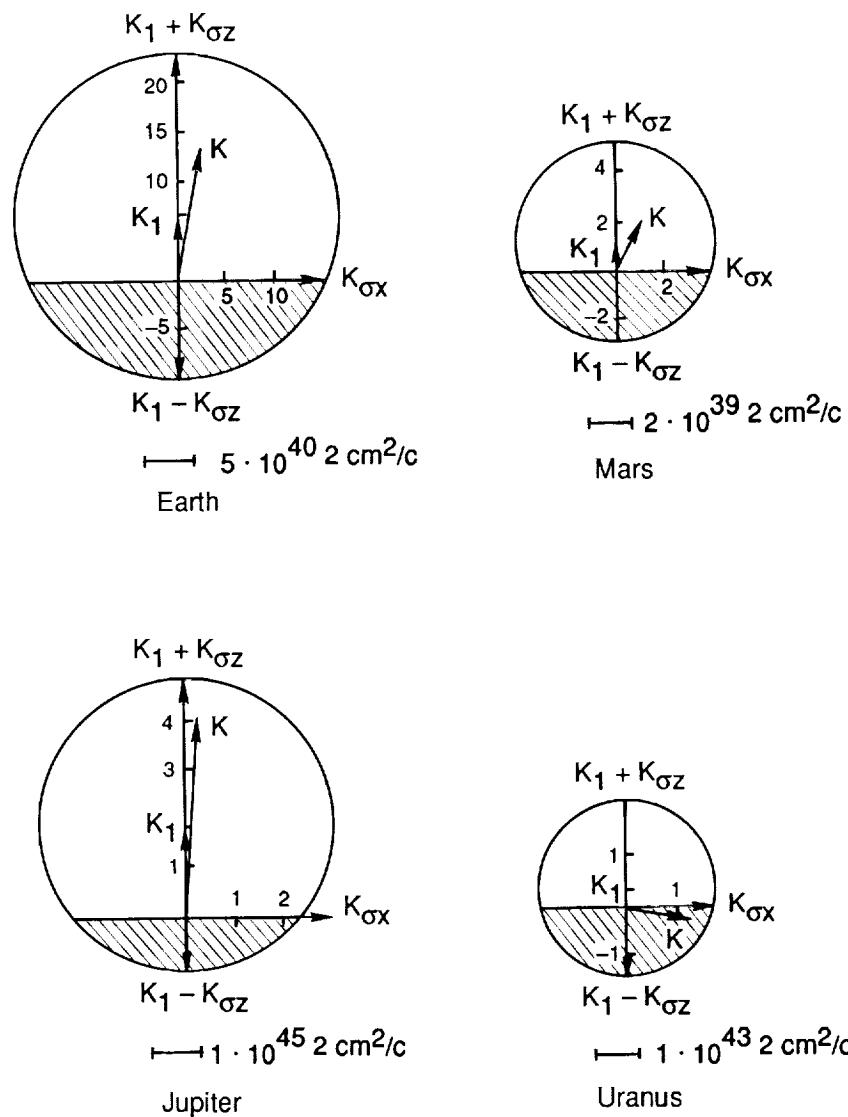
FIGURE 4 Diagrams of the components of the momentum of planetary axis rotation. Vector K of the observed rotation is shown. An initial rotational period of 10 hours was assumed for Earth. A rotational period equal to 15 hours was assumed for Uranus. The lined area corresponds to reverse rotation.

solar cosmic rays (high-energy nuclei of iron and other elements) could not have irradiated meteorite matter grains. Vityazev and Pechernikova (1985, 1987) showed that irradiation occurred when 100- to 1000-kilometer bodies appeared. The reasoning was that prior to that time nontransparency of the swarm of bodies was still sufficiently high, while less matter would have been irradiated at a later stage than the 5-10% that has been discovered experimentally. The conclusion then follows that there was virtually no gas as early as the primary stage of terrestrial planet accumulation. This conclusion is important for Earth science and comparative planetology because it is evidence in favor of the view of gas-free accumulation of the terrestrial planets at the later stages and repudiates the hypothesis of an accretion-induced atmosphere (see, for example the works of the Hayashi school).

ON THE COMPOSITION OF THE TERRESTRIAL PLANETS

According to current thinking, the Earth (and other planets) was formed from bodies of differing mass and composition. It is supposed that the composition of these bodies is, at least partially, similar to meteorites. Several constraints on the possible model composition of primordial Earth (generated by the conventional mixing procedure) can be obtained from a comparison of data on the location of meteorite groups and mafic bedrock on Earth on the diagram $\sigma^{17}0 - \sigma^{18}0$ and density data. Pechernikova and Vityazev (1989) found constraints from above on Earth's initial composition with various combinations of different meteorite groups: the portion of carbonaceous chondrite-type matter for Earth was < 10%, chondrite (H, L, LL, EH, EL) < 70% and achondrite (Euc) + iron < 80%. They proposed a method which can be used to determine multicomponent mixtures of Earth's model composition. The composition of Mars can also be determined from the hypothesis of the Martian origin of shergottites. Confirmation of the authors' hypothesis on the removal of the silicate shell from proto-Mercury (Vityazev and Pechernikova 1985; Pechernikova and Vityazev 1987) would mean that there is an approximately homogeneous composition for primary rock-forming elements in the entire zone 0.5-1.5 AU.

EARLY EVOLUTION OF THE PLANETS

Vityazev (1982) and Safronov and Vityazev (1983) showed that by the stage where 1000-kilometer bodies are formed, there commences a moderate, and subsequently, increasingly intensive process of impact processing, heating metamorphism, melting, and degassing of the matter of colliding bodies. It has been concluded that > 90% of the matter of bodies which
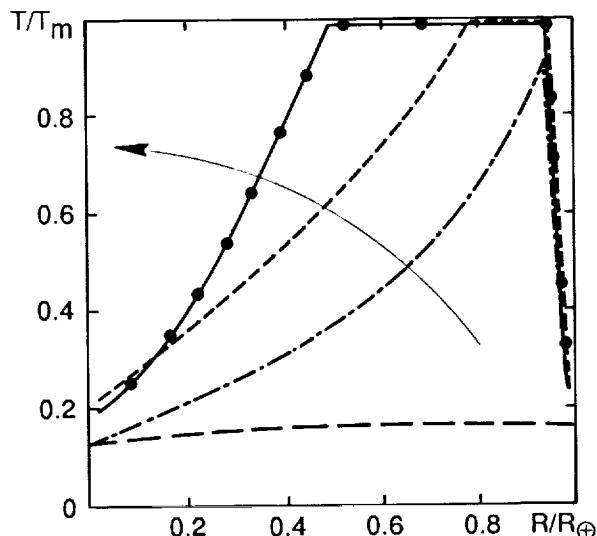
FIGURE 5 Estimates of the Earth's initial temperature: 1 is heating by small bodies (Safronov 1959); 2 is heating by large bodies (Safronov 1969); 3 is heating by large bodies (Safronov 1982); and 4 is heating by large bodies (Kaula 1980). The arrow indicates modifications in the estimates of T/T_m since the 1950s, including the authors' latest estimates.

went into forming the terrestrial planets had already passed the swarm through repeated metamorphism and melting, both at the surface and in the cores of bodies analogous to parent bodies of meteorites.

INITIAL EARTH TEMPERATURE AND ENERGY SOURCES

If we take into account the collisions of accreting planets with 1000-kilometer bodies, we conclude that there were extremely heated interiors, beginning with protoplanetary masses $\gtrsim 10^{-2}M_\oplus$. The ratio of the earliest and current estimates for primordial Earth are given in Figure 5. The latest temperature estimates indicate the possibility that differentiation began in the planet cores long before they attained their current dimensions.

The primary energy sources in the planets are known to us. In addition to energy released during the impact processes of accumulation, the most significant sources for Earth are: energy from gravitational differentiation released during stratification into the core and mantle ($\sim 1.5 \cdot 10^{38}$ ergs) and energy from radioactive decay ($\lesssim 1 \cdot 10^{38}$ ergs). It is important for specific zones to account for the energy of rotation released during tidal evolution of the Earth-Moon system ($\sim 10^{37}$ ergs) and the energy of chemical transformations $\lesssim 10^{37}$ ergs). However, these factors are not usually taken into

account in global models. The energy from radioactive decay at the initial stages plays a subordinate role. However, the power of this source may locally exceed by several times the mean value $\dot{\epsilon}r(U,Th,K) \simeq 10^{-6} erg/cm^3 s$ (in the case of early differentiation and concentration of U, Th, and K in near-surface shells). The power of the shock mechanism is significantly greater even on the average: $\dot{\epsilon}imp \sim 10^{-4}-10^{-5} erg/cm^3 s$. Intermediate values are generated for the energy of gravitational differentiation $\dot{\epsilon}_{GD} \sim 10^{-5}-10^{-6} erg/cm^3 s$ (the time scale for core formation is 0.1 to one billion years) and the source of equivalent adiabatic heating during collapse of $\sim 10^{-6} erg/cm^3 s$.

HEAT MASS TRANSFER PROCESSES

Heat-mass transfer calculations in planetary evolution models are currently made using various procedures to parameterize the entire system of viscous liquid hydrodynamic equations for a binary or even single-component medium. This issue was explored (Safronov and Vityazev 1986; Vityazev et al. 1990) in relation to primordial Earth. We will merely note here the order of values for effective temperature conductivities, which were used for thermal computations in spherically symmetrical models with a heat conductivity equations such as:

$$\frac{\partial T}{\partial t} = 1/R^2 \frac{\partial}{\partial R}(R^2 \Sigma K_j \frac{\partial T}{\partial R}) + \Sigma \dot{\epsilon}_i/\rho c_p. \tag{23}$$

The value which is normally assumed for aggregate temperature conductivity is a function of molecular mechanisms $K_m = 10^{-2} cm^2/s$. For shock mixing during crater formation, the effective mean is $K_{imp} \sim 1\text{-}10\ cm^2/s$. Similar values have effective values for thermal convection $(K_c \sim 10^{-1}\text{-}10\ cm^2/s)$ and gravitational differentiation $(K_{GD} \sim 1\text{-}10\ cm_2/s)$. It is clear from these estimates that the energy processes in primordial Earth exceeded by two orders and more contemporary values in terms of intensity.

INITIAL COMPOSITION INHOMOGENEITIES

The planets were formed from bodies with slightly differing compositions, and which accumulated at various distances. Variations in their composition and density were, on the average, on the same order as the adjacent planets:

$$|\delta c_o|/\bar{c} \approx |\delta \rho_o|/\bar{\rho} \lesssim 0.1 \tag{24}$$

These fluctuations were smoothed out during shock mixing with crater formation, but remained on the order of

$$|\delta\rho_o| \simeq |\delta c| = |\delta c_o|/\xi \simeq 10^{-3} - 10^{-4}, \tag{25}$$

where $\xi(\sim 10^2 - 10^3)$ is the ratio of mass removed from the crater to the mass of the fallen body. Using the distribution of bodies by mass (6), we can estimate the distribution of composition and density fluctuations both for a value $(\delta\rho)$ and for linear scales (1). For a fixed $\delta\rho$, spaces occupied by small-scale and large-scale fluctuations are comparable. It can be demonstrated that during planet growth, relaxation of inhomogeneities already begins (floating of light objects and sinking of heavy ones). In the order of magnitude, velocity v, relaxation time scale τ, effective temperature conductivity K and energy release rates $\dot\epsilon$ for one scale inhomogeneities occupying a part of the volume c, are obtained from the expressions:

$$|v| = |\delta\rho|g1/5n, \tau \sim R_\oplus^2/v1 \tag{26}$$

$$K \propto cv1, \quad \dot\epsilon \sim |\delta\rho|gc|v|1/R_\oplus,$$

where g is the acceleration of gravity, and η is the viscosity coefficient. Where $g \simeq 10^3 cm/s^2$, $\eta \sim 10^{20}$ poise, $\delta\rho/\rho \sim 10^{-3} - 10^{-4}$, $c \sim 0.1$, we have $|v| \gtrsim 10^{-5} - 10^{-6} cm/s$, $K \; 1 - 10^2 cm^2/s$, $\tau \sim (1-10)\cdot 10^6$ years, $\dot\epsilon \lesssim \dot\epsilon_r$ (U,Th,K). Estimates show that, owing to intensive heat transfer during relaxation of such inhomogeneities, the accreting planet establishes a positive temperature gradient (central areas are hotter than the external). This runs contrary to previous assessments (Safronov, 1959, 1969, 1982; Kaula 1980). The second important conclusion is that heavy component enrichment may occur towards the center, which is sufficient for closing off large-scale thermal convection. This is due to relaxation of the composition fluctuations during planet growth. Both of these conclusions require further verification in more detailed computations.

"THERMAL EXPLOSIONS" IN PRIMORDIAL EARTH

Peak heat releases, as these relatively large bodies fall, exceed by many orders the values for $\dot\epsilon_{imp}$ which are listed above. Entombed melt sites seek to cool, giving off heat to the enclosing medium. However, density-based differentiation, triggering a separation of the heavy (Fe-rich) component from silicates, may deliver enough energy for the melt area to expand. Foregoing the details (Vityazev *et al.* 1990), let us determine the critical dimensions of such an area. Let us write (23) in nondimensional form:

$$\frac{\partial\Theta}{\partial\tau} = \frac{\partial}{\partial\xi}(1 + Pe_o e^{\Theta/n})\frac{\partial}{\partial\xi}\Theta + \Gamma \cdot e^\Theta + \Gamma_r, \tag{27}$$

where

$$\Theta = \frac{(T - T_m)E}{RT_m^2}, \Gamma = \frac{\Delta\rho gch^2 v_o E}{4\lambda RT_m^2}, \Gamma_r = \frac{\dot{\epsilon}_r h^2 E}{4\lambda RT_m^2}, Pe_o = \left(\frac{v_o lc}{K}\right)^{1/n},$$

is the difference between heavy and light component densities, c is the heavy component portion in terms of volume, h is the characteristic size of the area (layer here), E is the energy for activation in the expression for the viscosity coefficient, λ is the heat conductivity coefficient, $v_o(\Theta = 0)$ is the Stokes' velocity, or filtration rate, whose numerical value is found from the condition $\epsilon_r = \epsilon_{GD}$. It can be demonstrated that for $\Gamma > \Gamma_{cr}$ ($\Gamma_c = 0.88$ for the layer and $\Gamma_{cr} = 3.22$ for the sphere) and $\Theta > \Theta_{cr}$ ($\Theta_{cr} = 1.2$ for the layer and $\Theta_{cr} = 1.6$ for the sphere) with $\Theta = 0$ at the area boundaries, conditions are maintained which promote a "thermal explosion." The emerging differentiation process can release enough energy to develop the process in space and accelerate it in time. For $\Delta\rho = 4.5$ g/cm^3, $\rho c = 0.2$, $\rho c_p = 10^8 erg/cm^3$ K, $RT_m^2/E = 50K$, $\ae = 10^{-2} cm^2/s$, and a Peclet number < 1, critical dimensions of the area (h_{cr}) are on the order of several hundred kilometers.

CHARACTERISTIC TIME SCALES FOR THE EARLY DIFFERENTIATION OF INTERIORS

Experimental data on meteorite material melt is too meager to make reliable judgements as to the composition of phases which are seeking to divide in the field of gravitational pull. Classical views, hypothesizing that the heavy (Fe-rich) component separates from silicates and sinks, via the filtration mechanism or as a large diapicre structures in the convecting shell, have only recently been expressed as hydrodynamic models. Complications with an estimate of the characteristic time scales for separation are, first of all, related to highly ambiguous data on numerical viscosity values for matter in the interiors. Variations in the temperature and content of fluids on the order of several percent, close to liquidus-solidus curves, alter viscosity numerical values by orders of magnitude. It is clear from this that even for Stokes' (slowed) flows, velocity $v \propto n^{-1}$ and characteristic time scales $\tau \propto n$ are uncertain. Secondly, existing laboratory data point to the coexistence of several phases (components) with sharply varied rheology. This further complicates the separation picture. Nevertheless, a certain overall mechanism which is weakly dependent on concrete viscosity values and density differences, was clearly functioning with interiors differentiation. A number of indirect signs are evidence of this; they indicate the very early and concurrent differentiation of all terrestrial planets, including the Moon.

The following are time scale estimates of the formation of the Earth's core:

1) From paleomagnetic data, the core existed 2.8 to 3.5 billion years ago;

2) Based on uranium-lead data, it was formed in the first 100 to 300 million years.

(The following do not clash with items 1 and 2 listed above).

3) Data on the formation of a protective, ionosphere or magnetosphere screen for ^{20}Ne and ^{36}Ar prior to the first 700 million years, and

4) Data on intensive degassing in the first 100 million years for I-Xe.

CONCLUSION

The solution (generated in the 1970's-80's at least in principle) to the primary problems of classical planetary cosmogony has made it possible to move towards a synthesis of the dynamic and cosmochemical approaches. The initial findings appear to be promising. However, they require confirmation. Clearly, there is no longer any doubt that intensive processes which promoted the formation of their initial shells occurred at the later stages of planet formation. At the same time, if we are to make significant progress in this area, we will need to conduct experimental studies on localized matter separation and labor-intensive numerical modeling to simulate large-scale processes of fractioning and differentiation of the matter of planetary cores.

REFERENCES

Cassen, P.M., and A. Summers. 1983. Models of the formation of the solar nebula. Icarus 53(1):26-40.

Ipatov, S.I. 1987. Solid-state accumulation of the terrestrial planets. Astron. vestn. 21(3):207-215.

Isaacman, R., and C. Sagan. 1977. Computer simulations of planetary accretion dynamics: sensitivity to initial conditions. Icarus 31: 510-533.

Kaula, W.M. 1980. The beginning of the Earth's thermal evolution. The continental crust and its mineral deposits. N.Y. pp. 25-31.

Laskar, J. 1988. Secular evolution of the Solar System over ten million years. Astron. Astrophys. 198(1/2):341-362.

Levin, B.Yu. 1964. The origin of the Earth and the planets. Nauka, Moscow.

Pechernikova, G.V. 1987. On the interim asymptotics of mass spectrum in the system of coagulating particles. Kinematics and Physics of Heavenly Bodies 3(5):85-87.

Pechernikova, G.V., and A.V. Vityazev. 1979. Masses of the largest bodies and velocity dispersion during planet accumulation. Letters to Astron. J. 5(1):54-59; Sov. Astron. Lett. 5:31-34.

Pechernikova, G.V., and A.V. Vityazev. 1980. The evolution of planetary orbit eccentricities in the process of planet formation. Astron. J. 57(4):799-811; Sov. Astron. 27:460-467.

Pechernikova, G.V., and A.V. Vityazev. 1987. Erosion of Mercury silicate shell during its accumulation. Pages 770-771. In: XVIII Lunar and Planetary Sci. Conf., Houston.

Pechernikova, G.V., and A.V. Vityazev. 1988. The evolution of dust clusters in the preplanetary disk. Astron. J. 65(1):58-72.

Pechernikova, G.V., and A.V. Vityazev. 1989. $\sigma^{17}O$-$\sigma^{18}O$ - ρ constraints on the composition of planetesimals forming the Earth. Theses XII of the National Symposium on stable isotopes in geochemistry. Moscow.

Ruzmaikina, T.V., and S.V. Maeva. 1986. Study of the process of protoplanetary disk formation. Astron. vestn. 20(3):212-227.

Safronov, V.S. 1959. On Earth's initial temperature. AS USSR Publ., Geophys. Series 1:139-143.

Safronov, V.S. 1969. The evolution of the preplanetary cloud and the formation of the Earth and the planets. Nauka, Moscow.

Safronov, V.S. 1982. The current state of the theory of the origin of the Earth. AS USSR Publ. Earth Physics 6:5-24.

Safronov, V.S., and A.V. Vityazev. 1983. The origin of the solar system. Conclusions of science and technology. Pages 5-93. In: Astronomiya, vol. 24. VINITI, Moscow.

Safronov, V.S., and A.V. Vityazev. 1986. The origin and early evolution of the terrestrial planets. Adv. Ph. Geochem. 6:1-29.

Stewart, G.R., and W.M. Kaula. 1980. A gravitational kinetic theory for planetesimals. Icarus 44:154-171.

Vityazev, A.V. 1982. Fractioning of matter during the formation and evolution of the Earth. AS USSR Publ. Earth Physics 6:52-68.

Vityazev, A.V., and G.V. Pechernikova. 1981. A solution to the problem of planet rotation within the framework of the statistical theory of accumulation. Astron. J. 58(4):869-878.

Vityazev, A.V., and G.V. Pechernikova. 1982. Models of protoplanetary disks near F - G stars. Letters to Astron. J. 6:371-377; Sov. Astron. Lett. 8:201-208.

Vityazev, A.V., and G.V. Pechernikova. 1985. Towards a synthesis of cosmochemical and dynamic approaches in planetary cosmogony. Meteoritika 44:3-20.

Vityazev, A.V., and Pechernikova G.V., 1987. When the gas was removed from the zone of terrestrial planets? Pages 1044-1045. In: XVIII Lunar and Planetary Sci. Conf., Houston.

Vityazev, A.V., G.V. Pechernikova, and V.S. Safronov. 1990. The Terrestrial Planets. Origin and Early evolution. Nauka, Moscow.

Weidenschilling, S.J. 1977. The distribution of mass in the protoplanetary system and solar nebula. Astron. Space Sci. 51:153-158.

Wetherill, G.W. 1980. Formation of the terrestrial planets. Numerical calculations. Pages 3-24. In: The Continent, Crust and Its Mineral Deposits. Geol. Assoc. of Canada.

Wetherill, G.W. 1985. Occurence of Giant Impacts during the growth of the terrestrial planets. Science 228:

Giant Planets and their Satellites:
What are the Relationships Between Their Properties and How They Formed?

DAVID J. STEVENSON
California Institute of Technology

ABSTRACT

The giant planet region in our solar system appears to be bounded inside by the limit of water condensation, suggesting that the most abundant astrophysical condensate plays an important role in giant planet formation. Indeed, Jupiter and Saturn exhibit evidence for rock and/or ice cores or central concentrations that probably accumulated first, acting as nuclei for subsequent gas accumulation. This is a "planetary" accumulation process, distinct from the stellar formation process, even though most of Jupiter has a similar composition to the primordial Sun. Uranus and Neptune are more complicated and imperfectly understood, but appear to exhibit evidence of an important role for giant impacts in their structure and evolution. Despite some interesting systematics among the four major planets and their satellites, no simple picture emerges for the temperature structure of the solar nebula from observations alone. However, it seems likely that Jupiter is the key to our planetary system and a similar planet could be expected for other systems. It is further argued that we should expect a gradual transition from solar nebula dominance to interstellar dominance in the gas phase chemistry of the source material in the outer solar system because of the inefficiency of diffusion in the solar nebula. There may be evidence for this in comets. Similar effects to this may have occurred in the disks that formed around Jupiter and Saturn during their accretions; this may show up in satellite systematics. However, each satellite system is distinctive, preventing general conclusions.

163

THE MASS DISTRIBUTION

In our solar system, over 99.5% of the known planetary mass resides in the planets beyond the asteroid belt, primarily in Jupiter, Saturn, Uranus, and Neptune. However, from the point of view of an earthbound cosmochemist, these bodies are not a "well known" (i.e. sampled) part of the solar system, but from the point of view of the astronomer seeking a general understanding of planetary system detectability and taxonomy, the giant planets should be the most important source of information.

It is convenient to divide the constituents of planets into three components: 1) "gas" (primarily hydrogen and helium; not condensable as liquid or solid under solar system conditions); 2) "ices" (volatile but condensable to varying degrees; H_2O is the most common, CO, CH_4, NH_3, and N_2 are the other main ones); and 3) "rock" (essentially everything else; primarily silicates and metallic or oxidized iron).

Jupiter defines a remarkable transition in our planetary system. Inside of Jupiter's orbit at 5 AU, the planets are small and rocky, largely devoid of both "ices" (especially water, the most abundant condensate in the universe) and "gas." By contrast, Jupiter has about 300 Earth masses of gas, and the more distant giant planets, though less well endowed, also have large reservoirs of gas. It is perhaps even more significant that Jupiter is the first place outward from the Sun at which water ice appears to become a common condensate. Although we do not know the abundance of water in Jupiter (because it forms clouds deep in the atmosphere), we see satellites such as Ganymede and Callisto which contain about as much ice as rock by mass, and we observe enhancements of other "ices" (CH_4 and NH_3) in the Jovian atmosphere.

The outer edge of the solar system is ill defined. It is possible that the cometary cloud contains a greater amount of ice and rock than do all the giant planets combined, especially if the most massive comets are substantially larger than the comets we have seen. A more conservative estimate of total cometary mass is approximately 10 Earth masses, but some increase in this estimate is justified given the recent realization that Halley is more massive than previously suspected (Sagdeev *et al.* 1988; Marochnik *et al.* 1988). The inner part of the cometary distribution, sometimes called the Kuiper belt, has now been tentatively identified as a disk rather than a spherical cloud (Duncan *et al.* 1988) and is therefore clearly associated with the planetary formation process. Planet X (a body beyond Pluto) has been frequently mentioned as a possibility, but no firm corroborative evidence currently exists.

One game that can be played is called *reconstituting the nebula*. One surveys the estimated amounts of rock and ice in each of the giant planets, then attempts to determine how much material of cosmic composition

would be required to provide that much rock and ice. Roughly speaking, this implies that each of the four major planets required ~ 0.01 $M_{\odot}$ of cosmic composition material. The cometary reservoir may have required an amount comparable to each of the planets. The similarity for each giant planet arises because they have roughly similar amounts of ice and rock (10-20 Earth masses) but diminishing amounts of gas as one proceeds outwards. The planets are also spaced in orbits that define a roughly geometric progression. In other words,

$$0.01 M_{\odot} \simeq \int_R^{2R} \sigma(R')2\pi R'dR', \tag{1}$$

independent of R, where $\sigma(R)$ is the "surface density" (mass per unit area) of the discoid nebula from which the planets form, and R is the (cylindrical) radius. This implies $\sigma(R) \sim (2 \times 10^4$ g cm$^{-2})/R^2$ where R is in astronomical units. Theoretical models for $\sigma(R)$ from accretion disk theory tend to give somewhat weaker dependences on R, implying a stronger tendency for most of the mass to be near the outer limits of the nebula. Naturally, most of the angular momentum is also concentrated in the outer extremities. The outer radius of the solar nebula is not known, but was presumably determined by the angular momentum budget of the cloud from which the Sun and planets formed.

INTERIOR MODELS

One could say a lot about how giant planets formed if one knew their internal structures. However, there are as of yet no techniques that are similar to terrestrial or solar seismology and that enable inversion for the interior densities in a detailed way. Instead, one must rely on a very small set of data, the lower-order (hydrostatic) gravitational moments, and the correspondingly small number of confident statements regarding the interiors. Even if the *quantity* of information thus obtained is low, the *quality* is high and represents a quite large investment of theoretical and computational effort, together with some important experimental data from high-pressure physics. Although the theory is not always simple, its reliability is believed to be high. The great danger exists, however, in overinterpreting the very limited data.

Good reviews on the structure of giant planets include Zharkov and Trubitsyn (1978), Stevenson (1982), and Hubbard (1984), and it is unnecessary to repeat here the techniques, data, and procedures used. In the case of Jupiter, there is no doubt that $\sim$ 90-95% of the total mass can be approximated as "cosmic" composition (meaning primordial solar composition). However, the gravitational moment J_2 (which can be thought of as a measure of the moment of inertia) indicates that there must be

some central concentration of more dense material (ice and rock). The uncertainties in hydrogen and helium equations of state are not sufficient to attribute this central density "excess" to an anomalously large compressibility of H-He mixtures or even to a helium core (since the latter can be limited in size by the observational constraints on depletion of helium in the outer regions of the planet). There is no way to tell what the "core" composition is; it could be all rock or all ice or any combination thereof. It does not even need to be a distinct core; it only needs to be a substantial enhancement of ice and/or rock in the innermost regions. The amount of such material might be as little as five Earth masses but is probably in the range of 10 – 30 Earth masses. The upper range of estimates is most reasonable if a substantial portion of this heavy material is mixed upward into the hydrogen and helium. One important point for the purposes of understanding origin is that Jupiter is enhanced in rock and ice by roughly a factor of 10 relative to cosmic composition. In other words, Jupiter formed from a cosmic reservoir containing 10^{-2} solar masses, even though its final mass is only 10^{-3} solar masses, a fact we had already noted in the previous section. The other important point about the dense material: it probably did not accumulate near the center by rainout of insoluble matter. This is in striking contrast to the Earth's core which formed because metallic iron was *both* more dense and insoluble in the mantle (silicates and oxides). The temperature in the center of Jupiter is very high ($>$ 20,000 K) and the mole fraction of the ice or rock phases, were they mixed uniformly in hydrogen, would not exceed 10^{-2}. Although solubility calculations are difficult (Stevenson 1985), there does not seem to be any likelihood that some component would be insoluble at the level of 10^{-2} mole fraction at $T \sim$ 20,000 K, since this requires an excess Gibbs energy of mixing of order kT ln 100 $\sim$ 8 eV, well in excess of any electronic estimate based on pseudopotential theory. It seems likely that Jupiter formed by first accumulating a dense core; the gas was added later. Subsequent convective "dredging" was insufficient to homogenize the planet (Stevenson 1985).

Saturn is further removed from a simple cosmic composition than Jupiter, a fact that can be deduced from the density alone since a body with the same composition as Jupiter but the same mass as Saturn would have about the same *radius* as Jupiter (Stevenson 1982). Saturn has only 83% of Jupiter's radius, implying a dense core that causes contraction of the overlying hydrogen-helium envelope. In fact, the ice and rock core of Saturn has a similar mass to that of Jupiter, but this is a larger fraction of the total mass in the case of Saturn. An additional complication in Saturn's evolution arises because of the limited solubility of helium in metallic hydrogen, predicted long ago but now verified by atmospheric abundance measurements. The presence of a helium-rich deep region is compatible with the gravity field (Gudkova *et al.* 1988) as well as being required by

mass balance considerations. As with Jupiter, the ice and rock central concentration must be primordial and form the nucleus for subsequent accretion of gas.

Despite recent accurate gravity field information (French *et al.* 1988) based on ring occultations, models for Uranus are not yet so well characterized. The problem lies not with the general features of the density structure, which are agreed upon by all modelers (Podolak *et al.* 1988), but with the interpretation of this structure, since no particular component (gas, ice, or rock) has predominance. A mixture of gas and rock can behave like ice, leading to a considerable ambiguity of interpretation. There is no doubt that the outermost $\sim 20\%$ in radius is mostly gas, and it is generally conceded that some rock is present within Uranus (though not much in a separate, central core). It is clear that the models require some mixing among the constituents: it is not possible to have a model consisting of a rock core with an ice shell and an overlying gas envelope as suggested around 1980. It is not even possible to have a model consisting of a rock core and a uniformly mixed envelope of ice and gas. The most likely model seems to involve a gradational mixing of constituents, with rock still primarily concentrated toward the center, and gas still primarily concentrated toward the outside.

Accurate models of Neptune must await the flyby in August, 1989. Based on the existing, approximate information it seems likely that the main difference between Uranus and Neptune is the extent of mixing of the constituents. Uranus has a substantial degree of central concentration (low moment of inertia), despite the inference of mixing described above. Neptune has a higher moment of inertia, suggesting far greater homogenization. At the high temperatures ($\sim 10^4$ K) and pressures (0.1 Mbar and above) of this mixing, phase separation is unlikely to occur, so the degree of homogenization may reflect the formation process (degree of impact stirring) rather than the phase diagram.

ATMOSPHERIC COMPOSITIONS AND THEIR IMPLICATIONS

Many of the minor constituents in giant planets undergo condensation (cloud formation) deep in the atmosphere and their abundances are accordingly not well known. The main exceptions are methane (which either does not condense or condenses in a region accessible to occultation and. IR studies) and deuterated hydrogen (HD). Some limited information on other species (especially NH_3) exists from radio observations, but we focus here on carbon and deuterium.

Carbon is enriched relative to cosmic by a factor of two (Jupiter), five (Saturn), and ~ 20 (Uranus). At least in the cases of Jupiter and Saturn, the

enhancement cannot be due to local condensation of the expected carbon-bearing molecules present in the primordial solar nebula (CO or CH_4) even allowing for clathrate formation. This interpretation follows from the fact that water ice did not condense closer to the Sun than about Jupiter's orbit, yet any solid incorporating CO or CH_4 requires a much lower temperature than water to condense (Lewis 1972). The enhancement of carbon must arise either through ingestion of planetesimals containing involatile carbon or "comets" (planetesimals that formed further out and were scattered into Jupiter-crossing orbits). There is increasing awareness of involatile carbon as a major carbon reservoir in the interstellar medium and it has long been recognized as a significant component of primitive meteorites. Comets also possess a substantial involatile carbon reservoir (Kissel and Kreuger 1987), but much of the cometary carbon reservoir is in C-O bonded material (part but not all as carbon monoxide; Eberhardt *et al*. 1987). If we are to judge from known carbonaceous chondrites, then the amount of such material needed to create the observed Jovian or Saturnian carbon enrichment is very large, 20 to 30 Earth masses, especially when one considers that this must be *assimilated* material (not part of the unassimilated core). Comets would be a more "efficient" source of the needed carbon, but it is also possible that the currently known carbonaceous chondrites do not reflect the most carbon-rich (but ice-poor) material in the asteroid belt and beyond. Even with comets, one needs of order 10 Earth masses of material added to Jupiter after it has largely accumulated. The implication is that estimates of ice and rock in Jupiter or Saturn, based solely on the gravity field, are likely to be lower than the true value because much of the ice and rock is assimilated (and therefore has no clean gravitational signature).

In Jupiter and Saturn, the value of $D/H \sim 2 \times 10^{-5}$ is believed to be "cosmic." However, this interpretation is still imperfectly established because of uncertainties in the "cosmic" value, and its true meaning (i.e., is it a primordial, universal value?). A cosmic value seems like a reasonable expectation, but it must be recognized that there are very strong fractionation processes in the interstellar medium which deplete the gas phase and enrich the particulate material. This enrichment is well documented for meteorites and is also probably present in comets. It is likely that the gaseous component of protoJupiter was *depleted* in deuterium, but that the assimilation of the carbon-bearing solids described above also contributed deuterium-rich materials, probably more than compensating for the gas-phase depletion. Thus, D/H in Jupiter is probably in excess of cosmic, though perhaps not by a large enough factor to be detectable in the current data. In contrast, Uranus is clearly enriched ($D/H \sim 10^{-4}$), an expected result given the far higher ratio of ice or rock to gas in that planet and the evidence of at least partial mixing discussed earlier. Neptune might be

expected to have an even larger D/H if it is more substantially mixed than Uranus.

In summary, atmospheric observations provide additional evidence of noncosmic composition and partial assimilation of "heavy" material (ice and rock) into the envelopes of giant planets.

HEAT FLOWS AND THEIR IMPLICATIONS

Jupiter, Saturn, and Neptune emit more energy than they receive from the Sun, implying significant internal energy sources. The *ultimate* source of this energy is undoubtedly gravitational, but there are several ways in which this energy can become available. In Jupiter, the heat flow is consistent with a simple cooling model in which the planet was initially much hotter and has gradually cooled throughout the age of the solar system. In this case, the gravitational energy of accretion created the primordial heat reservoir responsible for the current heat leakage. In Saturn, the heat flow is marginally consistent with the same interpretation, but the observed depletion of helium in the atmosphere requires a large gravitational energy release from the downward migration of helium droplets. This process may also contribute part of the Jovian heat flow. Even with helium rainout, it is necessary to begin the evolution with a hot planet (at least twice as hot as the present interior thermal state), but this constraint is easily satisfied by accretion models.

Uranus and Neptune have strikingly different heat flows. The Uranus internal heat output is less than 6×10^{21} erg/s and might be zero; expressed as energy output per gram, this is an even lower luminosity than the *Earth*. The Neptune heat flow is about 2×10^{22} erg/s. Although clearly much larger, it is still *less* than one would expect if Neptune were fully adiabatic and began its evolution with an internal temperature of at least twice its present value (the assumption that works so well for Jupiter). The difference between Uranus and Neptune is striking and not easily explained solely by their different distances from the Sun. It is also unlikely that these planets began "cold," that is, only slightly hotter than their present states since the energy of accretion is enough to heat the interior by $\sim 2 \times 10^4$ K. The low heat flow of Uranus may be due to stored heat of accretion; this heat is unable to escape because of compositional gradients, which inhibit thermal convection. In this way one can reconcile the low heat flow of Uranus with a high heat content and the inferred partial mixing of the interior discussed above (Podolak et al. 1990). By contrast, Neptune has a relatively high heat flow because it is more uniformly mixed. A speculative explanation for this difference in mixing efficiency is that the last giant impact on Uranus was oblique and created the large obliquity and disk from which the satellites formed. This impact was not efficient in mixing

the deep interior. By contrast, the last giant impact on Neptune was nearly head on, which is a more efficient way of heating and mixing the interior and did not lead to the formation of a compact, regular satellite system. The high heat flow of Neptune is accordingly related to its higher moment of inertia. This speculation may be testable after the Voyager encounter at Neptune.

In summary, the heat flows of giant planets support the expectation that these planets began their life hot. In some cases (e.g., Jupiter) much of this heat has since leaked out. In at least one case (Uranus) the heat has been stored and prevented from escaping by compositional gradients which inhibit convection.

SATELLITE SYSTEMS

The four giant planets exhibit a startling diversity of satellite systems. Jupiter has four large, comparable mass satellites with a systematic variation of density with distance, suggesting a "miniature solar system." Saturn has an extensive satellite system, though only one of the satellites (Titan) is comparable to a Galilean satellite. Uranus has a compact family of icy satellites, regularly spaced and in the equatorial plane. Neptune has only two known satellites, in irregular orbits. One of these is Triton, a large body that has significant reservoirs of CH_4 and possibly N_2. Satellites are common, and they probably have diverse origins (Stevenson *et al.* 1986). Some of the diversity may arise as the stochastic outcome of a common physical process (this may explain the difference between Jovian and Saturnian systems) but the Neptunian system is clearly different. One suspects that the Uranian system has a different history also, since it formed around a planet that was tipped over and never had as much gas accretion as Jupiter or Saturn. The recent enthusiasm for an impact origin of the Earth's Moon suggests that the Uranian system deserves similar attention. Impact origin seems to make less sense for Jupiter and Saturn, where the target is mostly gas, even though these planets must also have had giant impacts. The issue for Neptune is unresolved, though one wonders how a distant, nonequatorial, and *inwardly* evolving satellite such as Triton could have an impact origin. Perhaps Triton was captured.

The formation of Jovian and Saturnian satellites is commonly attributed to a disk associated with the planet's formation, and therefore crudely analogous to solar system formation. Pollack and Bodenheimer (1988) discuss in some detail the implications of this picture. Even if a disk origin is accepted, there are two distinct circumstances from which this disk arises. One scenario involves the formation of satellites from the material shed by a shrinking protoplanet. In this picture, protoJupiter once filled its Roche lobe, then shed mass and angular momentum as it cooled.

An alternative view is an accretion disk which forms and evolves before Jupiter or Saturn approaches its final mass. In this picture, the disk serves a role more similar to that of the solar nebula, though with some important dynamical differences: it is more compact (because of tidal truncation), and it is evolving more rapidly relative to the accretion time (whereas the viscous evolution time and accretion time are roughly comparable in the solar nebula). The solar system analogy must be used with care when applied to satellite systems! The choice between a disk that is shed and a true accretion disk has important implications for the chemistry (Stevenson 1990) but must be resolved by future dynamical modeling.

TEMPERATURES IN THE SOLAR NEBULA

Is there evidence in the outer solar system for the expected temperature gradient of the solar nebula? Perhaps surprisingly, the answer is no. There is a trend of *decreasing* gas content in giant planets as one goes outward, but this surely reflects formation time scales and the ability of a proto-giant planet to accrete large amounts of gas before the onset of T Tauri. Satellite compositions seem to reflect more the immediate environment of the central planet than the background temperature of the nebula. The lack of CO in Titan, and presumably Triton, may reflect the processing of solar nebula CO into CH_4 in the disk or envelope surrounding the proto-giant planet, rather than any statement about solar nebula conditions. The only statement about temperature that seems reasonably firm is the placement of water condensation (T $\sim$ 160 K) at around 5 AU at the time of condensate accumulation.

Of course, absence of evidence is not the same as evidence of absence. Nevertheless, we have to admit that we know remarkably little about the temperature variation in the solar nebula, either spatially or with time. One possible constraint could arise if there were a better knowledge and understanding of chemical trends in the outer solar system. For example, chemical processing such as catalyzed hydrogenation of CO to CH_4 and higher hydrocarbons is thermally mediated. The contamination by the products at greater radii in the nebula depends on where these reactions are quenched and how quickly or efficiently the species are dispersed by winds and turbulent diffusion. Stevenson (1990) has argued that the transport is inefficient so that the more distant regions of the nebula are dominated by interstellar speciation. Prinn (1990) has pointed out, however, that the uncertainties in momentum and species transport make it difficult to reach firm conclusions. In any event, chemical indicators are the best hope for obtaining information on outer solar system temperatures. If comets are found to have compositional trends as a function of formation position (as is suspected for asteroids) then these may provide the best clues.

GIANT PLANET FORMATION

The formation of the giant planets remains a major theoretical problem. Evidence presented above supports the idea that these planets may have formed by accumulating a core of ice and rock first, with gas accretion following—but truncated at some point, presumably because of the T Tauri mass loss or perhaps (in Jupiter's case) by tidal truncation of the accretion zone (Lin and Papaloizou 1979). The problem lies in the accumulation of the rock-ice core on a sufficiently short time scale, so that the gas is still present. Conventional accumulation models, based on Safronov's theory (1969) predict long time scales ($> 10^7$ years), even with allowances for gas-drag effects (Hayashi *et al.* 1985). Rock-ice cores may begin to accumulate gas when they are only approximately one Earth mass (Stevenson 1984) and this aids the accumulation somewhat, but does not solve the problem. Lissauer (1987) pointed out that if the surface density of solids is sufficiently high in the region of Jupiter formation then a runaway accretion may take place, forming the necessary Jupiter core in $\sim 10^5$ years. The onset of ice condensation helps increase the surface density by a factor of three, but this may not be sufficient by itself. Stevenson and Lunine (1988) suggest that a further enhancement may arise because of a diffusive transport of water molecules from the terrestrial zone into the Jupiter formation region. Several criteria must be satisfied to make this work well and they may not all be met. However, even a modest additional enhancement of the surface density in this region may make the mechanism work, at least to the extent of favoring the first (largest) giant planet at the water condensation front.

This suggests a speculative prediction for other planetary systems: giant planets should occupy the region outward from the point of water condensation. The largest of these (the extrasolar equivalent of Jupiter) may be near the condensation point. This position will vary with the mass of the central star (or with the mass of the nebula that the star once had) but is presumably a calculable quantity as a function of star mass and angular momentum budget. We await the exciting prospect of identifying Jupiters and superJupiters about nearby stars and characterizing their orbital distributions and properties.

The 1989 flyby of Neptune by Voyager reveals that Uranus and Neptune are more similar in structure than suggested above. This reduces the strength of arguments presented here for the role of giant impacts. Tremaine (preprint 1990) has suggested that the obliquity of Uranus is not related to impact.

ACKNOWLEDGEMENTS

REFERENCES

Duncan, M., T. Quinn, and S. Tremaine. 1988, Ap. J. 328:L69.

Eberhardt, P., D. Krankowsky, W. Schulte, U. Dolder, P. Lammerzahl, J.J. Berthelier, J. Woweries, U. Stubbemann, R.R. Hodges, J.H. Hoffman, and J.M. Illiano. 1987. Astron. Astrophys. 187:481.

French, R.G., J.L. Elliot, L.M. French, J.A. Kangas, K.J. Meech, M.E. Ressler, M.W. Buie, J.A. Frogel, J.B. Holberg, J.J. Fuensalida, and M. Joy. 1988. Icarus 73:349.

Gudkova, T.V., V.N. Zharkov, and V.V. Leontiev. 1988. Russian Astronomical Circular 1526:21.

Hayashi, C., K. Nakazawa, and Y. Nakagawa. 1985. Page 1100. In: Black, D. and M. Matthews (eds.). Protostars and Planets II. University of Arizona Press, Tucson.

Hubbard, W.B. 1984. Planetary Interiors. Van Nostrand Reinhold.

Kissel, J., and F.R. Kreuger. 1987. Nature 326:755.

Lewis, J.S. 1972. Earth Planet. Sci. Lett. 15:286.

Lin, D.N.C., and J. Papaloizou. 1979. Mon. Not. Roy. Astron. Soc. 186:799.

Lissauer, J.J. 1987. Icarus 69:249.

Marochnik, L.S., L.M. Mukhin, and R.Z. Sagdeev. 1988. Science 242:547.

Podolak, M., W.B. Hubbard, and D.J. Stevenson. 1990. In: Bergstrohl, J., and M. Matthews (eds.). Uranus. University of Arizona Press, Tuscon, in press.

Pollack, J.B. and P. Bodenheimer. 1988. Page 564. In: Pollack, J. and M. Matthews (eds.). Origin and Evolution of Planetary and Satellite Atmospheres. University of Arizona Press, Tucson.

Prinn, R.G. 1990. Ap. J. 348:725.

Safronov, V.S. 1969. Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets. Nauka Press, Moscow.

Sagdeev, R.Z., P.E. Elyasberg, and V.I. Moroz. 1988. Nature 331:240.

Stevenson, D.J. 1982. Ann. Rev. Earth Planet. Sci. 10:257.

Stevenson, D.J. 1984. Abs. Lunar Planet. Sci. Conf. 15:822.

Stevenson, D.J. 1985. Icarus 62:4.

Stevenson, D.J. 1990. Astrophys. J. 348:730.

Stevenson, D.J., A.W. Harris, and J.I. Lunine. 1986. Page 39. In: Burns, J. (ed.). Satellites. University of Arizona Press, Tucson.

Stevenson, D.J., and J.I. Lunine. 1988. Icarus 75:146.

Zharkov, V.N., and V.P. Trubitsyn. 1978. Physics of Planetary Interiors. Parchart, Tucson.

The Thermal Conditions of Venus

Vladimir N. Zharkov and V.S. Solomatov
Schmidt Institute of the Physics of the Earth

ABSTRACT

This paper examines models of Venus' thermal evolution. The models include the core which is capable of solidifying when the core's temperature drops below the liquidus curve, the mantle which is proposed as divided into two, independent of the convecting layers (upper and lower mantle), and the cold crust which maintains a temperature on the surface of the convective mantle close to 1200°C. The models are based on the approximation of parametrized convection, modified here to account for new investigations of convection in a medium with complex rheology.

Venus' thermal evolution, examined from the point when gravitational differentiation of the planet was completed (4.6 billion years ago), is divided into three periods: (1) adaptation of the upper mantle to the thermal regime of the lower mantle: approximately 0.5 billion years; (2) entry of the entire mantle into the asymptotic regime approximately three to four billion years; (3) asymptotic regime. The parameters of a convective planet in an asymptotic regime are not dependent on the initial conditions (the planet "forgets" its initial state) and are found analytically. The thermal flux in the current epoch is ~ 50 ergs $cm^2 s^{-1}$. We consider the connection between the thermal regime of Venus' core and its lack of magnetic field. After comparing the current thermal state of thermal models of Venus and Earth, together with the latest research on melting of the Fe-FeS system and the phase diagram of iron, we propose that Venus' lack of its own magnetic field is related to the fact that Venus' core does not solidify in the contemporary epoch. The particular situation of the iron triple point ($\gamma -$ ϵ – melt) strengthens this conclusion. We discuss the thermal regime of the

C-3

Venusian crust. We demonstrate that convection in the lower portion of the crust plays a minor role in regions with a particular crust composition, but that effusive or intrusive heat transport by melt, formed from melting of the crust's lower horizons, is the dominant mechanism for heat transport to the surface.

MODIFIED APPROXIMATION OF PARAMETRIZED CONVECTION

Models of Venus' thermal evolution, calculated in approximation of parameterized convection (APC), were examined in the works of Schubert (1979); Turcotte *et al.* (1979); Stevenson *et al.* (1983); Solomatov *et al.* (1986); and Solomatov *et al.* (1987). In parameterizing, dependencies were used that were obtained from studying convection in a liquid with constant viscosity. They were inferred to be true in cases of more complex rheology. New numerical investigations of convection in media with rheology that is more appropriate for the mantle (Christensen 1984a, b, 1985a, b) and theoretical research (Solomatov and Zharkov 1989) necessitated the construction of a modified APC (MAPC).

Let the law of viscosity be expressed as (Zharkov 1983):

$$\eta = b/\tau^{m-1} \exp\left[A_o/T \left(\rho/\rho_o\right)^L\right]\tag{1}$$

where τ denotes the second invariant of the tensor of tangential stresses, ρ is density; b and L are considered here to be the constants within the upper and lower mantles; $m \approx 3$; A_o denotes the enthalpy of activation for self diffusion (in K); and where $\rho = \rho_o$ is the reference value of density selected at the surface of each layer.

To describe convection in such a medium, following Christensen (1985 a), we will use two Raleigh figures:

$$Ra_o = \frac{\alpha g \rho \Delta T d^3}{\chi \eta_o},\tag{2}$$

$$Ra_T = \frac{\alpha g \rho \Delta T d^3}{\chi \eta_T},\tag{3}$$

where α denotes the thermal expansion coefficient; g is the acceleration of gravity; ΔT is the mean superadiabatic temperature difference in the layer; d denotes layer thickness, χ is the coefficient of thermal diffusivity; and η_o and η_T are defined by the formulae:

$$\eta_o = \frac{b}{\tau_o^{m-1}} \exp \frac{A_o}{T_o} ; \tau_o = \eta_o \frac{x}{d^2}\tag{4}$$

$$\eta_T = \frac{b}{\tau_o^{m-1}} \exp \frac{A_o}{\overline{T}} \left(\frac{\overline{\rho}}{\rho_o}\right)^L .\tag{5}$$

The dependence of the Nusselt number (determined in terms of the thickness of the thermal boundary layer δ) on Ra_o and Ra_T, where m = 3, is parameterized by the formula

$$Nu = \frac{d}{2\delta} = aRa_o^{\beta o}Ra_T^{\beta_T}. \tag{6}$$

In the case of free boundaries (lower mantle):

$$a = 0.29, \quad \beta_o = 0.37, \quad \beta_T = 0.16. \tag{7}$$

In the case of fixed boundaries (upper mantle):

$$a = 0.13, \quad \beta_o = 0.35, \quad \beta_T = 0.15. \tag{8}$$

The theoretical expression has the following appearance (Solomatov and Zharkov 1989):

$$Nu = \pi - \frac{3m+4}{2(m+2)}Ra_0^{\frac{m-1}{m+2}}Ra_T^{\frac{-1}{m+2}} = 0.23Ra_0^{0.4}Ra_T^{0.2}. \tag{9}$$

It is obtained from the balance of the capacities of viscous dissipation and buoyancy forces, and is in good agreement with (7) and (8).

The heat flow at the upper or lower boundary of the layer is equal to

$$F_i = \frac{\text{æ}\Delta T_i}{\delta}, \tag{10}$$

where ΔT_i denotes the temperature difference across the thermal boundary layer. Velocities at the boundary (u), mean tangential stresses in the layer (τ) and mean viscosity ($\overline{\eta}$) are estimated by the formulae (m=3):

$$u = \frac{\pi}{4}\frac{xd_2}{\delta}, \tag{11}$$

$$\overline{\tau} = \left(\frac{\pi u b}{d}\exp\frac{A}{T^o}\left(\frac{\overline{p}}{\rho_o}\right)^L\right)^{1/3}, \tag{12}$$

$$\overline{\eta} = \frac{b_2}{\overline{\tau}}\exp\frac{A_o}{\overline{T}}\left(\frac{\overline{p}}{\rho_o}\right)^{L'}. \tag{13}$$

The mean temperature of the layer, $\overline{T}$, and the temperature of the top of the lower thermal boundary layer, T_L, can be calculated from the adiabatic relationship through the temperature in the base of the upper thermal boundary layer, T_u:

$$\overline{T} = nT_u; \quad T_L = n_L T_u, \tag{14}$$

where n and n_L are constants.

With large Ra_T/Ra_o, the very viscous upper thermal boundary layer becomes reduced in mobility, taking an increasingly less effective part in convection, and the previous formulae are not applicable. Solomatov and Zharkov (1989) estimated that the transition to a new convective regime occurs when

$$Ra_T \gtrsim Ra_{T_{tr}} \approx 2Ra_o^2; \quad m = 3. \tag{15}$$

DESCRIPTION OF THE MODEL

There is no unequivocal answer to the question of whether convection in the Earth's mantle (or Venus') is single layered or double-layered. Previous works explored the single-layered models of convection. Possible differences in the single- and double-layered model of Venus' thermal evolution were discussed by Solomatov *et al.* (1986) and Solomatov *et al.* (1987). We propose here that convection is double-layered, and the boundary division coincides with the boundary of the second phase transition at a depth of approximately 756 kilometers (Zharkov 1983).

The thermal model of Venus (Figure 1a) contains a cold crust, whose role is to maintain the temperature in its base at approximately 1200°C (melting temperature of basalts), the convective upper mantle, the convective lower mantle, and the core. An averaged, spherically symmetric distribution, T (r,t) is completely determined by the temperatures indicated in Figure 1b.

The thermal balance equations for the upper mantle, the lower mantle, and the core are written as:

$$\frac{4}{3}\pi(R_L^3 - R_{12}^3)\rho_1 C_{p1}\frac{d\overline{T}}{dt^1} = 4\pi R_{12}^2 F_{12} - 4\pi R_L^2 F_L, \tag{16}$$

$$\frac{4}{3}\pi(R_{12}^3 - R_c^3)\rho_2 C_{p2}\frac{d\overline{T}}{dt^2} = \frac{4}{3}\pi(R_{12}^3 - R_c^3)\rho_2 Q_2 - 4\pi R_{12}^2 F_{21} + 4\pi R_c^2 F_c, \tag{17}$$

$$-\frac{4}{3}\pi R_c^3 C_{pc}\rho_c\frac{d\overline{T}}{dt^c} + Q_c\frac{dm}{dt} = 4\pi R_c^2 F_c. \tag{18}$$

Indices "1," "2," and "C" relate, respectively, to the upper mantle, lower mantle and the core. $\overline{T}$ denotes the mean layer temperature, F_L is the heat flow from the mantle under the lithosphere. The heat flow at the surface of the planet is obtained by adding to F_L, thermal flow generated by the radiogenic production of heat in the crust ($\sim$11 erg cm^{-2}s^{-1}). The radius of the lithosphere boundary (R_L) differs little from the radius of the planet (R_o) so that $R_L \approx R_o$. It is supposed that almost all of the radioactive elements of the upper mantle migrated into the crust when the
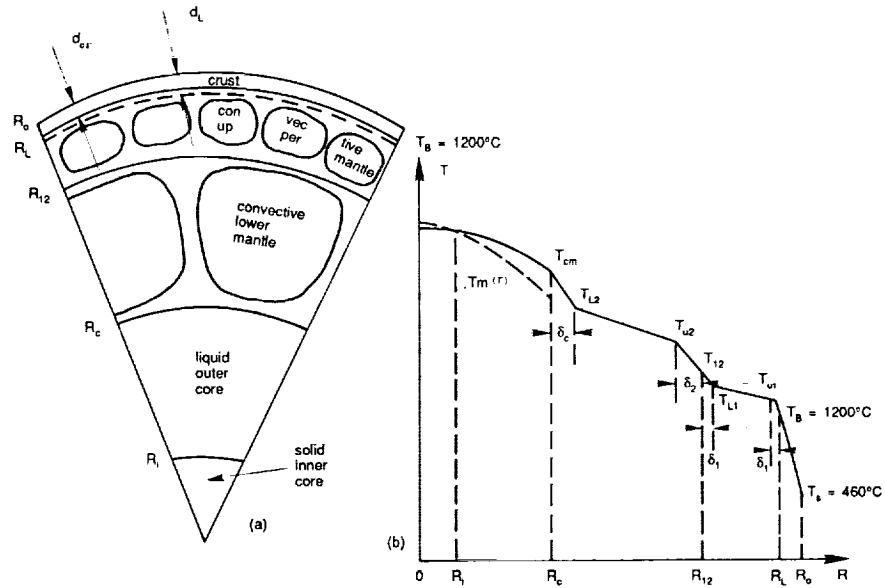
FIGURE 1 (a) Diagram of Venus's internal structure. Radii are indicated for: the planet—R_o; the base of the lithosphere—R_L; the boundary between the upper and lower mantles—R_{12}; the core—R_c; and the solid internal core —R_i. The dashed line illustrates the boundary of the lithosphere, which is an isothermic surface of T_B = $1200^\circ C$ = const; d_{cr} and d_L denote the thickness of the crust and the lithosphere. (b) Schematic, spherically symmetric temperature distribution in the cores of Venus. Reference temperatures are indicated for: the surface—T_s; lithosphere base—T_B; base of the upper thermal boundary layer of the upper mantle—T_{u1}; peak of the lower boundary layer of the upper mantle—T_{L1}; boundary between the upper and lower mantles—T_{12}; base of the upper boundary layer of the lower mantle—T_{u2}; peak of the lower boundary layer of the lower mantle—T_{L2}; and the boundary between the core and the mantle—T_{cm}. Thicknesses of the thermal boundary layers are given: δ_1 for the boundaries of the upper mantle; δ_2 and δ_c for the boundaries of the lower mantle. The dashed line indicates the core melting curve, which intersects the adiabatic temperature curve at the boundary between the outer, liquid and inner, solid core.

crust was melted. F_c denotes the heat flow from the core. The heat flow at the boundary between the upper and lower mantles with a radius of R_{12} is continuous: $F_{12} = F_{21}$.

Heat production in the lower mantle (Q_2) is defined by the sum

$$Q_2 = \sum_{i=1}^{4} Q_{oi}\exp\left[\lambda_i(t_o - t)\right] \tag{19}$$

where Q_{oi} and λ_i denote the current heat production of the radioactive

isotopes K, U, and Th for one kilogram of undifferentiated silicate reservoir of the mantle and their decay constants. The concentrations of K, U, and Th are selected in accordance with O'Nions *et al.* (1979): U = 20 mg/t, K/U = 10^4, Th/U = 4.

The term, Q_c dm/dt, in (18) describes heat release occurring when the core solidifies after the core adiabat drops below the core liquidus curve. It is supposed that the core consists of the mixture, Fe-FeS, and as solidification begins from the center of the planet, sulfur remains in the liquid layer, reducing the solidification temperature. The value Q_c is composed of the heat of the phase transition and gravitational energy.

According to the estimates of Loper (1978); Stevenson *et al.* (1983), and Solomatov and Zharkov (1989), Q_c = (1-2) · 10^{10}erg g^{-1}.

Mean temperature of the adiabatic core:

$$\overline{T}_c \approx n_c T_{CM} \tag{20}$$

where $n_c \approx 1.2$, the mean core density is $\overline{p} = 10.5$ g cm^{-3}, $C_{PC} = 4.7 \cdot 10^6$ erg g^{-1} K^{-1} (Zharkov and Trubitsyn 1980; Zharkov 1983).

The formulae for the melting $T_m(\rho)$ and adiabatic $T_{ad}(\rho)$ curves are written as follows:

$$T_m(\rho) = T_o(1 - \propto x)\left(\frac{\rho}{\rho cm}\right)^{2.24}, \tag{21}$$

$$T_{ad}(\rho) = T_{cm}\left(\frac{\rho}{\rho cm}\right)^{1.45}, \tag{22}$$

$$\frac{\rho}{\rho cm} - 1.224 - 0.009405\frac{r}{R_c} - 0.1586\left(\frac{r}{R_c}\right)^2 - 0.05672\left(\frac{r}{R_c}\right)^3. \tag{23}$$

Here T_o denotes the temperature at which pure iron melts at the boundary core of the radius R_c, where $\rho = \rho_{CM} = 9.59$ g cm^{-3}; $\alpha \approx 2$; x is the mass portion of sulfur in the liquid core, depending upon the radius of the solid core R_i, and the overall amount of sulfur in the core, x_o:

$$x(R_i) = \frac{x_o R_c^3}{R_c^3 - R_i^3} \tag{24}$$

The intersection of (21) and (22) defines the radius of the solidified core R_i (Figure 1b).

Convection in the upper mantle is parameterized by the MAPC with the parameters (7) (for fixed boundaries) in the lower mantle; and by the same formulae with the parameters (8) (for free boundaries). However, the difference between (7) and (8) is not very substantial. The parameters for the upper mantle are: $b_1 = 4.3 \cdot 10^{15}$dyne^{3}cm^{-6}s, $A_{o1} = 6.9 \cdot 10^4$K,

$\alpha_1 = 3 \cdot 10^{-5}\mathrm{K}^{-1}, \overline{p}_1 = 3.7\ \mathrm{g\ cm}^{-3}, \chi_1 = 10^{-2}\mathrm{cm}^{-2}\mathrm{s}^{-1}, \mathrm{æ}_1 = 4.5 \cdot 10^5\mathrm{erg}$ $\mathrm{cm}^{-1}\mathrm{s} \cdot 10^7\mathrm{erg\ g}$

and for the lower mantle:

$b_2 = 1.1 \cdot 10^{17}\mathrm{dyne}^3\mathrm{cm}^{-6}\mathrm{s},\ A_{o2} = 1.3 \cdot 10^5\mathrm{K},\ \alpha_2 = 1.5 \cdot 10^{-5}\mathrm{K}^{-1}, \overline{p}_2$ $= 4.9\ \mathrm{g\ cm}^{-3}, \chi_2 = 3 \cdot 10^{-2}\mathrm{cm}^{-2}\mathrm{s}^{-1}, \mathrm{æ}_2 = 1.8 \cdot 10^6\mathrm{erg\ cm}^{-1}\mathrm{s}^{-1}\mathrm{K}^{-1},\ g =$ $900\mathrm{cm\ s}^{-2}, C_{p2} = 1.2 \cdot 10^7\mathrm{erg\ g}^{-1}, n_2 = 1.13, n_{L2} = 1.26.$

NUMERICAL RESULTS AND ASYMPOTOTIC SOLUTION OF THE MAPC EQUATIONS

The evolution of the planet began approximately 4.6 billion years ago. The initial state of the planet was a variable parameter. The initial value of T_{u2} is the most significant, since T_{u1}, due to the low thermal inertia of the upper mantle, rapidly adapts to the thermal regime of the lower mantle ($t < 0.5$ billion years). The core has little effect on the evolution of the planet in general and the majority of models do not take into account its influence. T_{u2} was selected as equal to 2500, 3000, 3500 K. The upper value is limited by the melting temperature of the mantle, since a melted mantle is rapidly freed from excess heat.

The upper mantle adapts to the thermal regime of the lower mantle for the first approximately 0.5 billion years (Figure 2a, b). Then, after approximately three to four billion years the entire mantle enters an asymptotic regime which is not dependent upon the initial conditions. The evolution picture in general is similar to the one described by Solomatov et al. (1986); and Solomatov et al. (1987). The planet is close to an asymptotic state in the present epoch. Contemporary parameters of the models are:

$T_{u1} = (1700\text{-}1720)$ K,	$u_1 = (2.1\text{-}2.4)\mathrm{cm\ yr}^{-1}$
$T_{12} = (2500\text{-}2530)$ K,	$u_2 = (0.8\text{-}1.0)\mathrm{cm\ yr}^{-1}$
$T_{u2} = (2840\text{-}2870)$ K,	$\overline{r}_1 = (5\text{-}6)$ bars,
$F_L = (35\text{-}40)\mathrm{erg\ cm}^{-2}\ \mathrm{s}^{-1}$,	$\overline{r}_2 = (110\text{-}120)$ bars,
$\sigma_1 = (28\text{-}30)\mathrm{km}$,	$\overline{\eta}_1 = (1\text{-}2)10^{21}$ poise,
$\sigma_2 = (130\text{-}140)\mathrm{km}$,	$\overline{\eta}_2 = (3\text{-}10)10^{22}$ poise.

According to the criterion (15) MAPC are applicable throughout the entire evolution, just as with the quasistationary criterion (Solomatov et al. 1987).

The asymptotic expression for F_L in the first approximation is formulated as (Solomatov et al. 1987):

$$F_L = F_Q \left(1 + \frac{t_{in}}{t_r}\right),\tag{25}$$

where F_Q denotes the thermal flow created by the radioactivity of the lower mantle, and t_r is the characteristic time of decay:
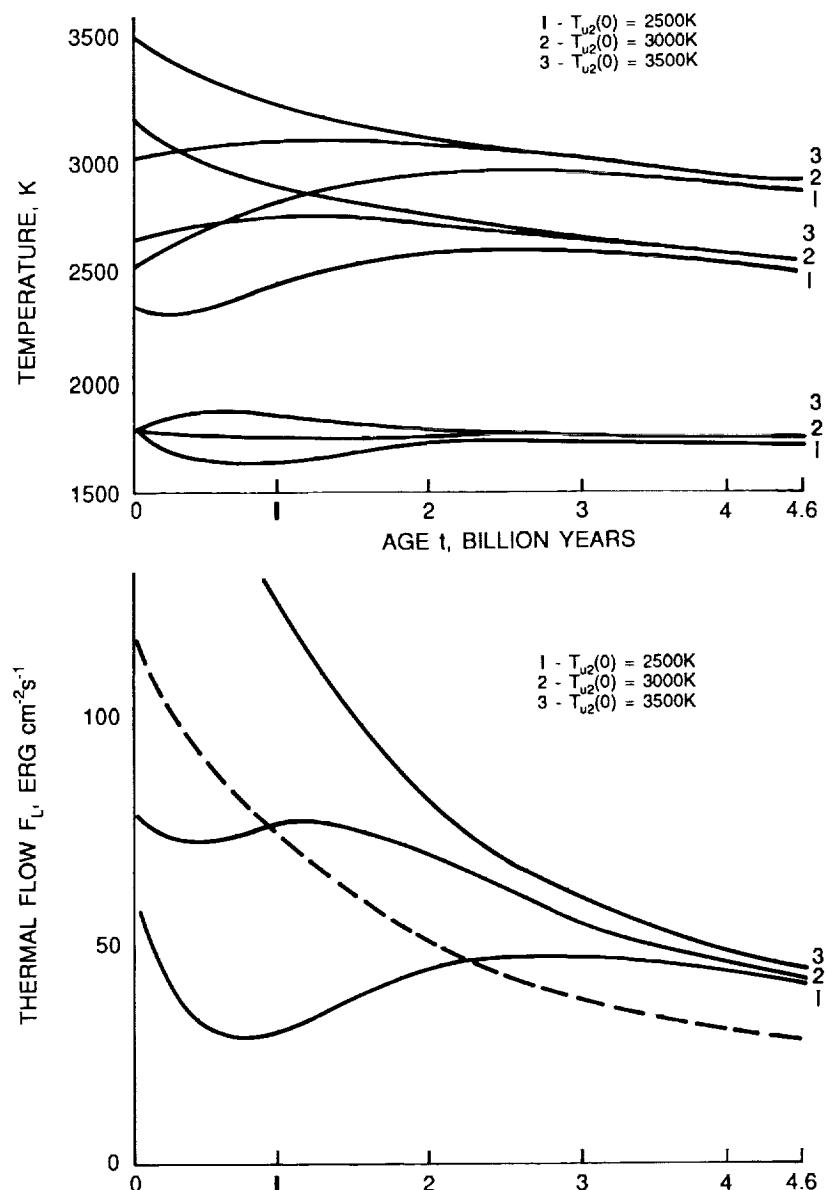
FIGURE 2 (a) Evolution of the base temperatures, T_{u1}, T_{12} and T_{u2}, with differing initial conditions. The core is not taken into account and is considered to be $T_{cm} = T_{L2}$ (Figure 1). (b) Evolution of the thermal flow under the lithosphere, F_L, with differing initial conditions. The thermal flow to the surface is obtained by adding ~ 11 erg cm^{-1}s^{-1} to F_L, generated by the radioactive elements of the crust. The dashed line indicates the thermal flow generated by radioactive elements of the lower mantle.

$$t_r = -\left(\frac{dlnQ_2}{dt}\right)^{-1}, \tag{26}$$

equal $\sim 5.67 \cdot 10^9$ years at the current time.

The time scale of thermal inertia of the mantle is equal to

$$t_{in} = t_{in1} + t_{in2}(1 + \delta), \tag{27}$$

where t_{in1} and t_{in2} are the time scales for the inertia of each of the layers individually:

$$t_{in1} = \frac{M_1 n_1 C_{p1} \Delta T_1}{M_2 Q_2}\left(1 + \beta_1 + \frac{\beta_1 A_1(T_{u1} - T_B)}{T_{u1}^2}\right)^{-1}, \tag{28}$$

$$t_{in2} = \frac{n_2 C_{p2} \Delta T_2}{Q_2}\left(1 + \beta_2 + \frac{\beta_2 A_2(T_{u2} - T_{12})}{T_{u2}^2}\right)^{-1}, \tag{29}$$

$$\xi = \frac{\Delta T_1}{\Delta T_2}\frac{1 + n_{L1}\bar{s}}{\bar{s}}\frac{1 + \beta_o + \beta_2 - 1/3(\beta_o - 2\beta_2)A_{o2}(T_{u2} - T_{12})T_{u2}^{-2}}{1 + \beta_1 + \beta_1 A_1(T_{u1} - T_B)T_{u1}^{-2}}, \tag{30}$$

M_1 and M_2 denote the layer masses, $\bar{s} = R_{12}{}^2 / R_L{}^2$. The values of (27) through (30) are calculated in a zero approximation: $F_L = F_Q$, and $A = A_o/n \ (\rho/\rho_o)^L$.

The time scale for thermal inertia, as compared with models based on the conventional APC (Solomatov *et al.* 1986, 1987), has increased from $\sim 2.5 \cdot 10^9$ years to $\sim 3.5 \cdot 10^9$ years. The mantle and core temperatures obtained are somewhat lower (by 300 K near the core). In the models with the core, $T_{cm} \approx 3720$K and $F_c \approx 15$ erg cm^{-2}s^{-1} in the contemporary epoch. Since the adiabatic value is $F_c \approx 30$ erg cm^{-2}s^{-1}, there is no convection if solidification is absent, and a magnetic field cannot be generated (Stevenson *et al.* 1983).

MAGNETISM AND THE THERMAL REGIME OF THE CORES OF EARTH AND VENUS

We will estimate the temperatures in the Earth using the MAPC. MAPC is not applicable to the Earth's upper mantle, since convection in the Earth involves the surface layer, and rheology is, in general, more complex. However, MAPC is applicable to the Earth's lower mantle. Let us assume a value for the temperature at the boundary between Earth's upper and lower mantles of

$$T_{12} = (2300 - 2500)K, \tag{31}$$

(Zharkov 1983), and the thermal flow at this boundary is

$$F_{21} \approx \frac{R_o^2}{R_{12}^2} \left(F - F_{cr} + C_p M_1 \frac{d\overline{T}_1}{dt} \right) \approx 80 erg cm^{-2} s^{-1},$$

where $F - F_{cr} \approx 70$ erg $cm^{-2}s^{-1}$ is the medium thermal flow from the Earth, after deducting heat release in the crust (according to the estimates of Sclater *et al.* 1980), $d\overline{T}_1/dt \approx -100K/\text{billion}$ years. (Basaltic Volcanism Study Project 1981).

We will assume the following values for the rheologic parameters:

$$b = 6.6 \cdot 10^{16} dyne^3 cm^{-6} s, A = 1.5 \cdot 10^5 K, and A_o = 1.3 \cdot 10^5 K.$$

The remaining parameters are listed in Zharkov (1983).

As a result, we have:

$$T_{u2} = (2800\text{-}2900) \text{ K},$$
$$T_{L2} = (3600\text{-}3800) \text{ K},$$
$$T_{cm} = (3800\text{-}4000) \text{ K},$$
$$\delta_2 = 100 \text{ km}.$$

The value of $T_{cm} - T_{L2}$ depends upon F_c, which we will assume to be equal to the adiabatic value of $F_{ad} \approx 30$ erg $cm^{-2}s^{-1}$.

Therefore, the temperature at the boundary of the Earth's core, T_{CME} is greater than for Venus (T_{CMV}) by a value of

$$T_{CME} - T_{CMV} = (100 - 300)K. \tag{33}$$

In order for Venus' core not to solidify, it is necessary that the adiabat of the Venusian core not drop below the solidification curve during cooling. Otherwise, solidification of the core will cause the core to mix by chemical or thermal convection, and it will trigger the generation of a magnetic field (Stevenson *et al.* 1983). We will estimate the difference between T_{CME} and the temperature, T_{crV} (which is critical for the beginning of solidification), at the boundary of Venus' core, with a single pure iron melting curve, T_m (P) and a single equation of the state for iron $\rho(P)$.

T_{CME} is found from the intersection of the adiabat of the Earth's core (22) and the liquidus curve (23), with a sulfur content in the core of x_E at the boundary of the Earth's inner core. T_{crV} is obtained from the intersection of the adiabat of Venus' core (22) and the liquidus curve (23) with a sulfur content of x_V. We then obtain (Solomatov and Zharkov 1989):

$$T_{CME} - T_{crV} = (+300) \div (-300), K, \tag{34}$$

where $x_E - x_V = 0 \div 0.07$. Therefore, if $x_V \gtrsim x_E - (0 \div 0.02) = 0.07 \div 0.12$; (with $x_E = 0.09 \div 0.12$; Aherns 1979), the core of Venus is not solidifying at the present time, and a magnetic field is not being generated.

Complete solidification of the core would have led to the absence of a liquid layer in the core and would have made it impossible for the magnetic field to be generated. However, for this, the temperature near the boundary of Venus' core should have dropped below the eutectic value, which, according to the estimates of Anderson *et al.* (1987) is ~ 3000 K and according to Usselman's estimate (1975) is ~ 2000 K. Such low temperatures of the core seem to be of little probability.

Pressure in the iron triple point, $\gamma - \epsilon - 1$, (liquid) approaches the pressure in the center of Venus. This is significant for interpretation the absence of the planet's own magnetic field. According to the estimates of Anderson (1986), $P_{tp} \approx 2.8$ Mbar (Figure 3), but is also impossible to rule out the larger values of P_{tp}. In the center of Venus, $P_{cv} \approx 2.9$ Mbar; at the boundary of the solid inner Earth's core, $P_{IE} = 3.3.$Mbar; and in the Earth's center, $P_{CE} = 3.6$ Mbar (Zharkov 1983). If $P_{cv} \lesssim P_{tp}$, the conclusion that there is no solidification of Venus' core is further supported, since, in this case, the core's adiabat (critical for the beginning of solidification) drops several hundred degrees lower. This stems from the fact that reduction in the temperature of solidification of the mixture, $\gamma-$Fe $-$ FeS is greater than for the mixture $\epsilon-$ Fe $-$ FeS by $\Delta S_\epsilon / \Delta S_\gamma$ times, where $\Delta S_\epsilon / \Delta S_\gamma$ denotes the ratio of enthropy jumps (during melting) equal to ~ 2, according to Anderson's estimates (1986).

The latest experimental data on the melting of iron allow us to estimate the melting temperature in the cores of Earth and Venus. Figure 3 shows melting curves obtained by various researchers, and the $\gamma - \epsilon$ boundary, computed by Anderson (1986). With $x = 0.09$-0.12 and $\alpha = 1 \div 2$ (formula 21), the full spread of temperatures at the boundary of the Earth's core, leading to the intersection of the liquidus and adiabat curves, is equal to

$$T_{CME} = 3500 \div 4700K, \tag{35}$$

- with data from Brown and McQueen (1986) and

$$T_{CME} = 4300 \div 5400K, \tag{36}$$

- with data from Williams *et al.* (1987).

The effect of pressure on viscosity of the lower mantle reduces the effective Nusselt number and increases the temperature of T_{CME} and T_{CMV} by ~ 300 K (Solomatov and Zharkov 1989). We obtain the estimate

$$T_{CME} = 3800 \div 4300K, \tag{37}$$

which is the best fit with (35).
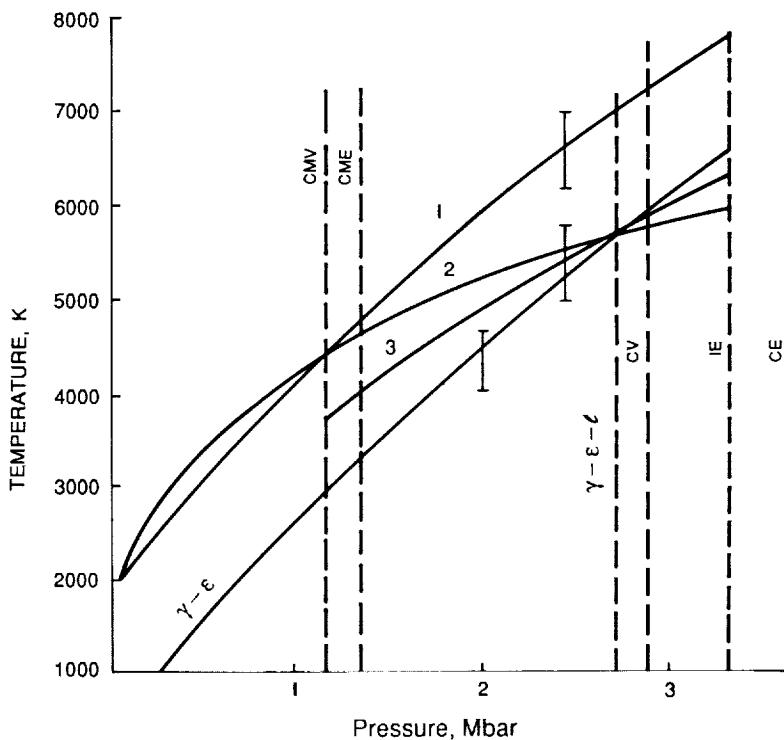
FIGURE 3 Pure iron melting curves according to various data and the boundary of the phase transition γ—Fe—ϵ—Fe. The figures indicate melting curves: 1 is from the study by Williams *et al.* (1987); 2 is from the study by Anderson (1986) and experimental data from Brown and McQueen (1986); and 3 is based on the Lindeman formula with the Grunehaisen parameter of G = 1.45 = const, using experimental data from Brown and McQueen (1986). The transition boundary, γ—ϵ, was constructed in the study by Anderson (1986) and together with the melting curve gives us the location of the triple point. The vertical segments illustrate errors in determining temperature for curves 1 and 2. The boundary labels are as follows: CM is the boundary between the core and the mantle; C denotes the center of the planet; I is the boundary of the solid inner core; and the final letter indicates Earth (E) or Venus (V). γ - ϵ - I denotes the location of the triple point.

THE THERMAL REGIME OF THE VENUSIAN CRUST

What are the mechanisms by which heat is removed from Venus's interior to the surface? Global plate tectonics, like Earth's, are absent on the planet (Zharkov 1983; Solomon and Head 1982), although in individual, smaller regions, it is possible that there are features of plate tectonics (Head and Crumpler 1987). A certain portion of heat may be removed by the mechanism of hot spots (Solomon and Head 1982; Morgan and Phillips

1983). Conductive transport of heat through the crust clearly plays a large role. However, with a large crust thickness ($\gtrsim$ 40 km), as indicated by data from a number of works (Zharkov 1983; Anderson 1980; Solomatov *et al.* 1987), the heat flow of $F \approx$ (40-50) erg cm^{-2}s^{-1} triggers the melting of the lower crust layers and removal of approximately one half of heat flow by the melted matter. The latter flows to the surface as lava or congeals in the crust as intrusions.

We shall discuss the possibility of solid-state convection in the crust as an alternative mechanism of heat removal through the Venusian crust.

The crust is constituted of an upper, resilient layer with a thickness of d_e and a viscous one with a thickness of $d_v = d_{cr} - d_e$.

The crust thickness of d_{cr} is constrained in our model by the phase transition of gabbro-eclogite, since eclogite, with a density higher than that of the underlying mantle, will sink into the mantle (Anderson 1980; Sobolev and Babeiko 1988). The depth of this boundary is dependent upon the composition of basalts of Venus' crust. We assume that $d_{cr} =$ 70 km (Yoder 1976; Zharkov 1983; Sobolev and Babeiko 1988). It is considered that radioactive elements are concentrated in the upper portion of the crust, and the primary heating occurs via the heat flow from the mantle of $F_L \approx$ 40 erg cm^{-2}s^{-1}. The boundary (d_e) of the resilient crust is defined as the surface of the division between the region effectively participating in convection and the nonmobile upper layer. Crust rheology is described by law (1) with the parameters from Kirby and Kronenberg (1987). Four modeled rocks are considered: quartz diorite, anorthosite, diabase and albite. Parameter m in (1) is close to 3 for them, so that the formulae MAPC with parameters (8) for fixed boundaries is fully suitable for estimating. In addition, criterion (15) is used to define the boundary, d_e. At this boundary, which is regarded as the upper boundary of the convective portion of the crust, the temperature is equal to

$$T_o = 733 + 20 d_e(km), K. \tag{38}$$

The following physical parameters are assumed (Zharkov *et al.* 1969):

$$\rho = 2.8 g \ cm^{-3}, \alpha = 2 \cdot 10^{-5} K^{-1}, \ \text{æ} = 2 \cdot 10^5 erg \ cm^{-1} s^{-1} K^{-1}.$$

Computations have demonstrated that the thickness of the resilient crust is 20-30 kilometers. The mean temperature of the convective layer of the crust has been calculated at $\sim$ 1600K, 1700K, 1900K and 2000K, respectively, for quartz diorite, anorthosite, diabase, and albite, and exceeds the melting temperature for basalts by hundreds of degrees. This means that convection does not protect the crust from melting, and heat is removed by the melted matter.

We can estimate that portion of the heat which is removed by convection. For this, let us assume the temperature in the base of the crust to be T_B = 1500K (approximately the melting temperature). We obtain the Nusselt number from the formulae of MAPC, and find that it is ~1.7; 1; 0.6; and 0.5, respectively, for quartz diorite, anorthosite, diabase, and albite. For η = const, and for complex rheology, as in (1), (Christensen 1984a and 1985a), convection begins at Nu ~ 1.5-2, as calculated according to the formulae of APC (MAPC). Thus, where $d_{cr} \approx$ 70km, perhaps only quartz diorite is convective, removing 25-30 erg cm^{-2}s^{-1}, while 10-20 erg cm^{-2}s^{-1} is removed by the melt. The rate of convective currents is three to five millimeters per year. It can be demonstrated that

$$Nu \sim (d_{cr} - d_e)^{0.9} \tag{39}$$

With d_{cr} ~ 120 km, convection also begins in anorthosite, while in quartz diorite it occurs virtually without melting.

These estimates show that convection in the crust may play some role in individual regions, depending on the thickness and composition of the crust. It is not excluded that in individual regions, convection in the crust makes its way to the surface. The bulk of the heat is, apparently, removed by melted matter. The rate of circulation of material from the crust is, with this kind of volcanism, 50-100 km^{3}yr^{-1}. This is three to five times greater than crust generation in the terrestrial spreading zones. Another process by which basalt material circulates is where new portions of melted basalt reach the crust from the upper mantle, and basalt returns back to the mantle in the eclogite phase. This process may trigger the accumulation of eclogite in the gravitationally stable region between the upper and lower mantles. It may lead to the chemical differentiation of the mantle. This process has been noted for the Earth by Ringwood and Irifune (1988).

Figure 4 illustrates various processes involved in heat and mass transport.

The geological structures observed on Venus may be related to these processes. Flat regions may be tied to effusive, basalt volcanism. Linear structures in the mountainous regions may be related to horizontal deformations which are an appearance of convections in the crust or mantle. Ring structures may stem from melted intrusion or hot plume lifted towards the surface from the bottom of the upper mantle.

Convection both in the mantle and in the crust is, apparently, nonstationary (as on Earth). This nonstationary nature stems from instabilities occurring in the convective system. The characteristic time scale for such fluctuations is t ~ d/u, where d ~ 10^8-10^9 cm is the characteristic dimension, and u ~ 1 cm per year is the characteristic velocity. Therefore, characteristic "lifespans" for various occurrences of instability are t ~ 10^8-10^9 years.
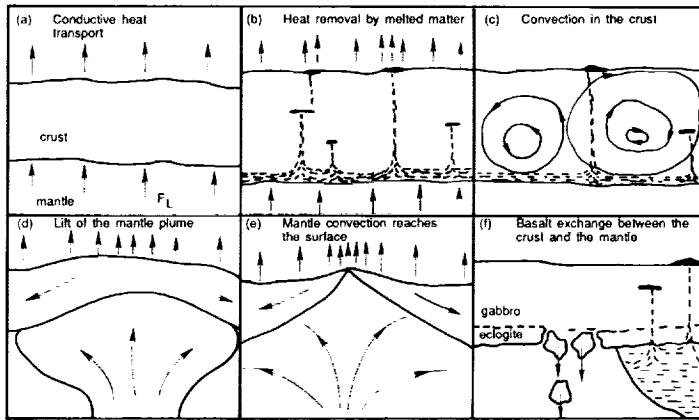
FIGURE 4 Process of heat transport through the Venusian crust: (a) transfer of heat
carried from the mantle by the mechanism of conductive thermal conductivity; (b) heat
removal by the melted matter which is formed as the basalt crust melts. The melted matter
may flow to the surface or form intrusions; (c) convection in the crust which does not
reach the surface or reaching the surface; (d) lifting of hot plume from the bottom of the
upper mantle to the crust of Venus, triggering enhanced heat flow, a flow of crust material,
and the sinking of the crust; (e) involvement of the crust in mantle convection with the
formation of spreading zones; (f) basalt exchange between the crust and the mantle. Basalt
is formed when the upper mantle partially melts and is returned back as eclogite.

Regional features of tectonic structures, thermal flows, volcanic activity and
so on, can exist for this length of time.

CONCLUSION

1. Modification of the approximation of parameterized convection
for the case of nonNewtonian mantle rheology led to no marked increase
in the time scale of thermal inertia of the mantle from two to three or three
to four billion years in comparison with the usual parameterization. The
thermal flow at the surface of Venus of ~ 50 erg cm^{-2}s^{-1}, is the product
of radiogenic heat release from the mantle (50%), heat release in the
crust (20%), and cooling of the planet (30%). These figures for Earth are
approximately 40, 20, and 40% for the double-layered convection models.
Temperatures in the upper portion of the mantle are approximately 1700 K,
which is 50-100 K greater than on Earth. Given the existing uncertainties
in the concerning parameters, $T_{cm} \sim 3700\text{-}4000$ K at the core boundary
and may, possibly, be greater. This temperature is 100-300 K less than for
the Earth.

2. The magnetic field on Venus is absent. This is most likely due to
the lack of core solidification and, respectively, the lack of energy needed

to maintain convection in the liquid core. For this, it is enough for the sulfur content in Venus' core to be even a little bit less than on the Earth (but not more than 20-30% less). This conclusion is stronger if the triple point in the phase diagram for iron, $\gamma - \epsilon - 1$, lies at pressures that are greater than in Venus' center, but approximately less than in Earth's center.

3. At the values obtained for heat flows from Venus' mantle, the crust melts, and the melted matter ($\lesssim 100$ km^3 per year) removes about one half of the entire heat. The remainder is removed conductively. In individual regions, depending on the crust thickness and composition, a portion of the heat may be removed by convection, reducing crust temperature and the portion of heat removed by the melted matter. Flow velocities comprise several millimeters per year. Due to the insufficiently high temperature of the surface, convection in the crust separates from the surface as a highly viscous, nonmobile layer with a thickness of 20-30 kilometers. In individual regions, crust convection may emerge to the surface. Basalt circulation also occurs by another way: the basalt is melted out of the upper mantle and returned back in the form of eclogite masses, which sink into the lighter mantle rock. It is possible that this process triggers the accumulation of eclogite at the boundary between the upper and lower mantles, resulting in the chemical separation of the mantle. These processes may explain the formation of various geological structures on Venus.

4. The nonstationary nature of convection in Venus' mantle and crust determine regional features of tectonic, thermal and volcanic appearances on the surface of the planet which have a characteristic duration of $\sim 10^8$-10^9 years.

REFERENCES

Ahrens, T.J. 1979. J. Geophys. Res. 846:985-990.
Anderson, D.L. 1980. J. Geophys. Res. Lett. 7:101-102.
Anderson, O.L. 1986. Geophys. J. R. Astron. Soc. 84:561-579.
Basaltic Volcanism Study Project. 1981. Basaltic volcanism on the Terrestrial Planets. Pergamon, New York.
Brown, J.M., and R.G. McQueen. 1986. J. Geophys. Res. 91:7485-7494.
Christensen, U.R. 1984a. Geophys. J. R. Astron. Soc. 77:343-384.
Christensen, U.R. 1984b. Phys. Earth Planet. Inter. 35:264-282.
Christensen, U.R. 1985a. Phys. Earth Planet. Inter. 37:183-205.
Christensen, U.R. 1985b. J. Geophys. Res. 90:2995-3007.
Head, J.W., and L.S. Crumpler. 1987. Science 238:1380-1385.
Kirby, S.H., and A.K. Kronenberg. 1987. Rev. Geophys. 25:1219-1244, 1680-1681.
Loper, D.L. 1978. Geophys. J. R. Astron. Soc. 54:389-404.
Morgan, P., and R.J. Phillips. 1983. J. Geophys. Res. 88: 8305-8317.
O'Nions, R.K., N.M. Evensen, and P.J. Hamilton. 1979. J. Geophys. Res. 24:6091-6101
Parmentier, E.M., D.L. Turcotte, and K.E. Torrance. 1976. J. Geophys. Res. 81:1839- 1846.
Richter, F.M., H.C. Nataf, and S.F. Daly. 1983. Fluid Mech. 129:173-192.
Ringwood, A.E., and T. Irifune. 1988. Nature 331:131-136.
Schubert, G. 1979. Ann. Rev. Earth Planet. Sci. 7:289-342.
Sclater, J.C., C. Jaupart, and D. Galson. 1980. Rev. Geophys. Space Phys. 18:269-311.

Sobolev, S.V., and A.Yu. Babeiko. 1988. Pages 103-104. In: Theses from the 8th Soviet-American working meeting on planetology, 22-28 August, 1988.

Solomatov, V.S., V.V. Leont'ev, and V.N. Zharkov. 1986. Astron. vestn. 20:287-305.

Solomatov, V.S., V.V. Leont'ev, and V.N. Zharkov. 1987. Gerlands Beitr. Geophysik. 96:73-96.

Solomatov, V.S., and V.N. Zharkov. 1989. Icarus, in press.

Solomon, S.C., and J.W. Head. 1982. J. Geophys. Res. 87: 9236-9246.

Stevenson, D.J., T. Spohn, and G. Schubert. 1983. Icarus 54: 466-489.

Turcotte, D.L., F.A. Cook, and R.J. Willeman. 1979. Proc. Lunar Planet. 10:2375-2392.

Usselman, T.M. 1975. Amer. J. Sci. 275:291-303, 275-291.

Williams, Q., R. Jeanloz, J. Bass, B. Svendsen, and T.Y. Ahrens. 1987. Science 236:181-182.

Yoder, Jr. 1976. Generation of basaltic magma. National Academy of Sciences. Washington, DC.

Zharkov, V.N. 1983. The Moon and Planets 29:139-175.

Zharkov, V.N. 1983. The Internal Structure of the Earth and the Planets. Nauka, Moscow.

Zharkov, V.N., V.L. Pan'kov, A.A. Kalachnikov, and A.I. Osnach. 1969. Introduction to the Physics of the Moon. Nauka, Moscow.

Zharkov, V.N., and V.P. Trubitsyn. 1980. The Physics of Planetary Interiors. Nauka, Moscow.

Degassing

JAMES C.G. WALKER
The University of Michigan

ABSTRACT

Measurements of the concentrations of rare gases and trace elements in oceanic basalts have provided new information concerning the structure of the Earth's mantle and its evolution. This review is based principally on papers by Allegre, Staudacher, Sarda, O'Nions, Oxburgh, and Jacobsen. Approximately 35% of the mantle lost more than 99% of its rare gas content in the first 100 million years of solar system history. A comparable volume of the mantle has also been depleted in radioactive and other large ion lithophile elements, the depleted elements being concentrated in continental crust. But depletion was a much slower process than degassing. The average age of continental crust is 1.8 billion years, but the average age of the rare gas atmosphere is 4.4 billion years. There has been very little mixing of material between the degassed and depleted portion (presumably the upper mantle) and the undegassed and relatively undepleted portion (presumably the lower mantle).

Gas fluxes from the mantle indicate that degassing today is inefficient, affecting only the top few hundred meters of oceanic crust. It is not likely that sea floor spreading processes like those now operating could have degassed the entire upper mantle within a 100 million years, even given large initial heat fluxes. At the same time, it is not likely that sea floor spreading processes could have dissipated the initial heat of a nearly molten Earth. Lava flooding could have removed initial heat efficiently and at the same time degassed the upper mantle rapidly.

Rare gases do not make an atmosphere, of course. There is new information concerning the release of carbon dioxide from the mantle. As

pointed out most forcefully by Marty and Jambon, the exogenic system (atmosphere, ocean, and sedimentary rocks) is deficient in carbon by a factor of 100 relative to rare gases when present amounts are compared with present fluxes from the mantle. It appears that carbon dioxide did not participate in the initial rapid degassing that released rare gases from the upper mantle. Instead, carbon has been modestly concentrated into the continental crust like other incompatible, but not atmophile, elements. Less than 10% of upper mantle carbon has been transferred to the crust, and the total mantle amount may be 40 times the amount in the exogenic system.

INTRODUCTION

Important new information has become available in recent years concerning the release of gases from the interior of the Earth. The most fruitful source of information has been the measurement of rare gas concentrations in sea floor basalts. The results set important constraints that need to be incorporated into any comprehensive understanding of the early history of the planets. In my review here, I will describe some of the highlights of these results and give an indication of how they are derived. I cannot provide a complete description of all of the evidence that is used to reach the conclusions presented.

RESERVOIRS

Measurements on sea floor basalts have provided clear indications of two major reservoirs within the mantle. The larger reservoir, constituting about 65% of the mantle, is undegassed and relatively undepleted in incompatible elements. The remaining 35% of the mantle was degassed very early in Earth history (within 100 million years of the beginning), and more than 99% of the initial gas content of this reservoir was released. Throughout the whole of Earth history there has been very little mixing between these reservoirs (O'Nions 1987; Anderson 1989).

These conclusions are based on measurements of the concentrations in sea floor basalts of the radioactive parent elements shown in Figure 1, along with their radiogenic daughter isotopes and non-radiogenic cousin isotopes also shown in the figure (Allegre *et al.* 1983). The important feature of these isotope systems is that the ratio of daughter/cousin increases through time as a result of the radioactive decay of the parent, and that there are no other processes that will cause the ratio of daughter/cousin to change because they are chemically and physically almost identical.

Figure 2 shows how the ratio of daughter/cousin, called ALPHA, increases at a rate that depends on the ratio of parent/cousin, called MU.

Parent	Daughter	Cousin
K40	Ar40	Ar36
U,Th	He4	He3
I129	Xe129	Xe130

■ ALPHA (=DAUGHTER/COUSIN) increased by decay

■ MU (=PARENT/COUSIN) determines rate of increase

FIGURE 1 Isotope taxonomy.

The solid line in the top panel of the figure shows the evolution of the amount of radiogenic ^{40}Ar resulting from the decay of radioactive ^{40}K. The bottom panel shows the evolution of the ratio of ^{40}Ar/^{36}Ar, ALPHA. The effect of a degassing episode fairly early in Earth history is indicated by the left hand arrow labeled Degas 50%. The degassing episode reduces the concentration of ^{40}Ar by a factor of two, as shown in the top panel. Because ^{36}Ar concentration is also reduced by a factor of two there is initially no change in ALPHA. The rate of increase of ALPHA with time is larger after the degassing episode, however, because there is less ^{36}Ar in the denominator of the ALPHA ratio. This evolution is shown by the dashed line in the figure. The effect of a second degassing event at -1 billion years is also shown in the figure. The impact of the second degassing event on the evolution of ALPHA is smaller because, later in Earth history, there is less radioactive ^{40}K left to decay. Thus, early degassing leads to large increases in ALPHA; late degassing has a smaller effect. The event in the middle of Figure 2 shows the effect of a depletion by 50% in the concentration of radioactive ^{40}K. Depletion reduces the rate of increase of ALPHA in the manner indicated by the dashed line. In this way it is possible to deduce the history of MU from measurements of ALPHA.

The basic data concerning mantle degassing appear in Figure 3. They are ALPHA values measured for He, Ar, and Xe in mid-ocean ridge basalts and in ocean island basalts. The mid-ocean ridge basalts appear to sample the upper mantle, whereas the ocean island basalts are assumed
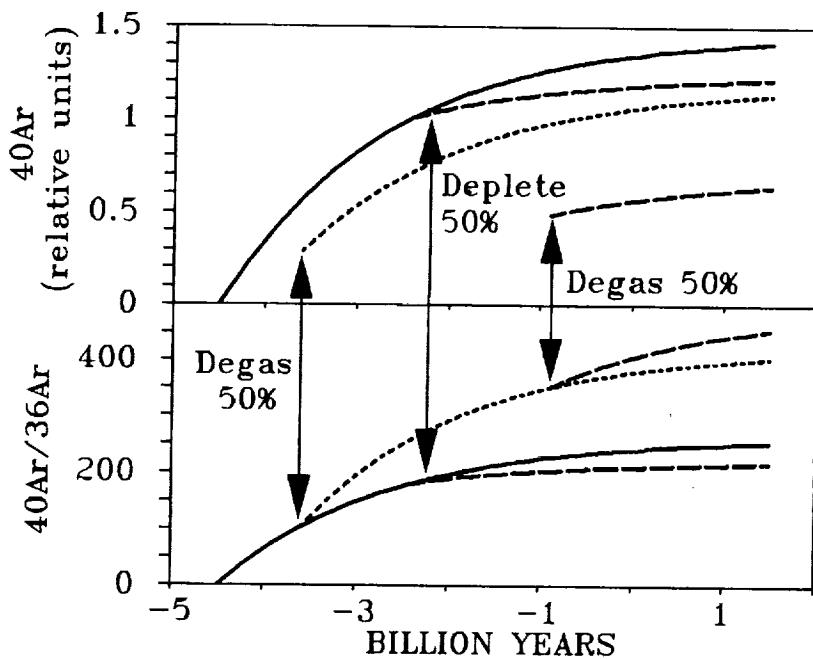
FIGURE 2   Measure ALPHA to deduce history of MU.

to sample plumes of material rising from the lower mantle. There is a range of compositions of ocean island basalts representing various degrees of mixing between lower mantle material and upper mantle material. As representative of the least contaminated material I show results for Loihi sea mount in Hawaii. The point is that ALPHA is larger in MORB than in Loihi material, which indicates that MORB material is more degassed. The enhancement in ALPHA has been large for He and Ar. From data such as these it is now possible to derive important results concerning mantle reservoirs.

First, the bulk Earth concentration of K gives the ^{40}Ar concentration in undegassed mantle material. The ALPHA value observed in Loihi basalts gives the ^{36}Ar concentration in undegassed mantle material. The mass of ^{36}Ar in the atmosphere then gives the mass of the mantle that has been degassed. From a comprehensive study of rare gas isotope systematics Allegre *et al.* (1987) deduce that 46% of the mantle has been degassed. To increase the ALPHA value of Ar from the Loihi value to the MORB value it is necessary that no more than 390/25000 = 1.6% of the initial ^{36}Ar complement be retained in degassed mantle material. This value would apply in the case of early degassing from undepleted material. Delayed degassing or prior depletion of ^{40}K would reduce the permitted degree of

ALPHA

	He	Ar	Xe
MORB	86,000	25,000	6.95
Loihi	25,000	390	6.48

- MORB samples degassed reservoir (upper mantle)

- Loihi samples undegassed (lower mantle)
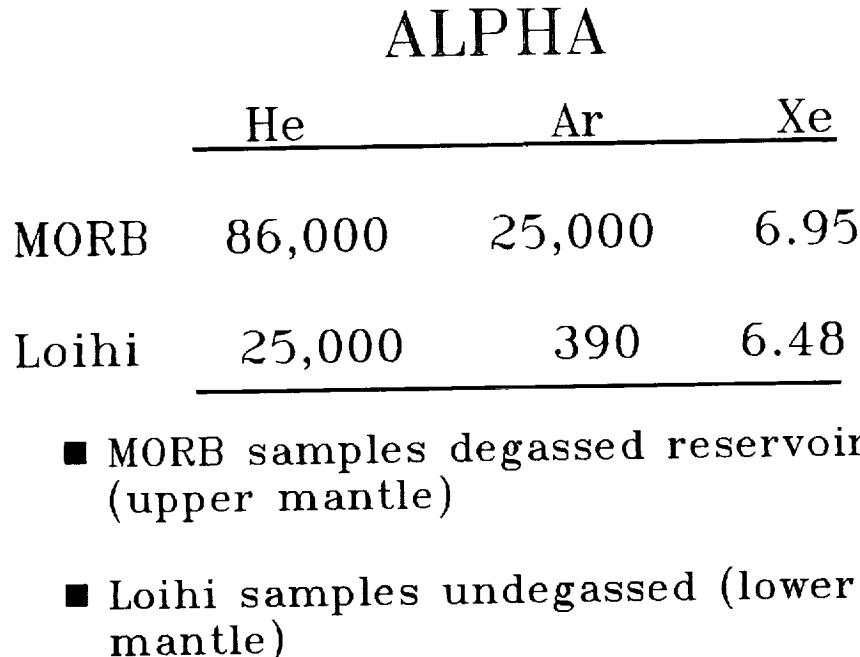
FIGURE 3   Data that constrain degassing (Allegre *et al.* 1987).

retention. The conclusion is that degassing has been very thorough indeed. At the same time, because the difference between ALPHA values in Loihi and MORB is so great, it is possible to conclude that just 2% contamination of MORB material by Loihi material would reduce the ALPHA value of the degassed mantle by a factor of two. There is therefore evidence for strong isolation of the mantle reservoirs from one another.

The increase in the ALPHA value for Xe between Loihi and MORB, although modest, demonstrates that degassing took place very early in Earth history. For ALPHA to have changed, degassing must have occurred before all of the parent ^{129}I had decayed away. But the half life of ^{129}I is only 17 million years. Therefore, the division of the mantle into two reservoirs, the very thorough degassing of one of these reservoirs, and the nearly total isolation of the two reservoirs all took place very early in Earth history. At this time it is not clear to me how to reconcile these surprising conclusions with our current understanding of the growth of the Earth by planetesimal impact, in which planetesimals were vaporized and degassed, at least during the later stages of accretion. Neither is it clear how to reconcile with these data the current thinking concerning the formation of the Moon by a giant impact event occurring near the end of Earth

accretion. It appears likely that such an impact would have completely remixed and homogenized the mantle. On the other hand, it is not clear that such an impact event would have led to complete degassing of the mantle or to complete removal from the Earth of any atmosphere released during the course of previous accretion. Also unclear is what physical process causes the separation of the mantle into two distinct reservoirs. In my further analysis I shall assume that degassing of the upper mantle was a consequence of mantle convection, possibly driven by accretional energy, but that most of the impacts, and in particular the giant Moon-forming impact had already occurred before the processes that brought about the presently observable state had begun.

In this interpretation then, degassing should be related to continental growth and the depletion of the upper mantle in incompatible elements. Studies of continental growth and depletion are based on precisely the same kind of isotopic arguments as the studies of degassing already described. The only difference in depletion is that the daughter and cousin isotopes are concentrated in the continents instead of in the atmosphere. Analyses of Sm-Nd, Lu-Hf, and Rb-Sr isotopes in sea floor basalts, summarized in Figure 4, indicate that 30% of the mantle has been depleted to form the continents (Jacobsen 1988). The average age of the continents is 1.8 Ga. Allegre *et al.* (1983, 1988), in a similar analysis, conclude that 35% of the mantle has been depleted while 47% of it has been degassed (Sarda *et al.* 1985). The average age of the rare gas atmosphere deduced in their analysis is 4.4 Ga. My tentative conclusion is that the degassed and depleted reservoirs are probably the same, but that degassing occurred much earlier than depletion.

FLUXES

Fluxes of gases from the mantle to the atmosphere can be deduced from the measured flux of ^{3}He and the concentration ratios in sea floor basalts. These fluxes lead to the very interesting conclusion that heat is released much more readily from the mantle than are the rare gasses (O'Nions and Oxburgh 1983; Oxburgh and O'Nions 1987). Further it can be argued that degassing today is inefficient. Processes now operating could not have degassed the upper mantle rapidly and thoroughly. A comparison of the fluxes of heat, helium, and argon is presented in Figure 5. The sources are mainly concentrated in the lower mantle because the upper mantle is depleted in radioactive incompatible elements. The heat flux through the surface of the Earth exceeds the sum of upper and lower mantle sources because the interior of the Earth is cooling down. This fact is reflected in the Urey ratio of source/flux. For heat this ratio has a value of about 0.6 (Pollack 1980). For ^{4}He the Urey ratio is 6.8, indicating that
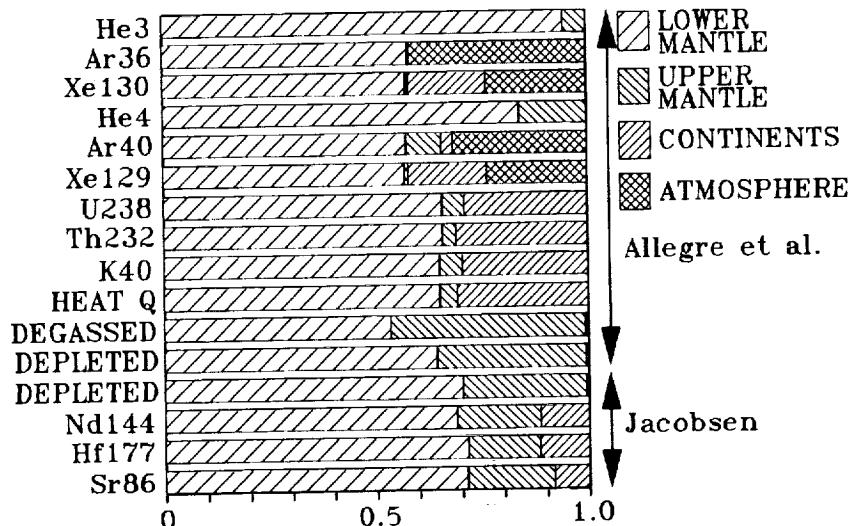
FIGURE 4 Comparison of deductions concerning degassing and depletion (Allegre *et al.* 1983, 1987; Jacobsen 1988). The bars indicate what fraction of the terrestrial complement of each isotope is in the indicated reservoir. HEAT Q refers to heat source.

most radiogenic helium is retained within the Earth and that the flux from mantle to atmosphere is much less than the production within the Earth. However, the flux does exceed upper mantle production. Helium must be flowing from the lower mantle to the upper mantle at a significant rate. For ^{40}Ar, on the other hand, the flux is less than the upper mantle source. There is no evidence of flow from lower mantle to upper mantle; the Urey ratio is 23, and ^{40}Ar is accumulating even in the depleted upper mantle. These observations provide strong support for the notion of a two-layer convective structure in the mantle.

It is entirely reasonable to suppose that heat is more mobile than helium which is in turn more mobile than argon. The argon flux from the mantle is 6.2 × 10^6 mole/y. The ^{40}Ar concentration in the upper mantle is 3 × 10^{-10} mole/g. Therefore, the rate at which upper mantle material is degassed, calculated from the ratio of these two numbers, is 2 × 10^{16} g/y. Since the mass of the upper mantle (35% of the total mantle) is 1.4 × 10^{27} g, it would take 70 Ga to degas the upper mantle at this rate. But the xenon isotope data indicate that the upper mantle was degassed in less than .1 Ga. Therefore, the present rate of degassing is too slow to explain the observations by a factor of 1000.

Furthermore, degassing today is inefficient, in the sense illustrated in Figure 6. Ocean crust is formed by the partial melting of upper-mantle

```
                          Heat        He4          Ar40
    UREY  RATIO          0.6         6.8           23
    SOURCE/FLUX


    FLUX:  3.6E15  ▲    9.6E7  ▲       .62E7  ▲
  _____|_____|_____
  Upper  mantle     |           |               |
                    |           |               |
    SOURCE:      .09E15        4E7            1E7
                  watt                mole/y
                    ▲           ▲
                        5.5E7   |
  _____|_____|_____
  Lower  mantle     |           |
    SOURCE:      1.4E15        63E7           13E7
```
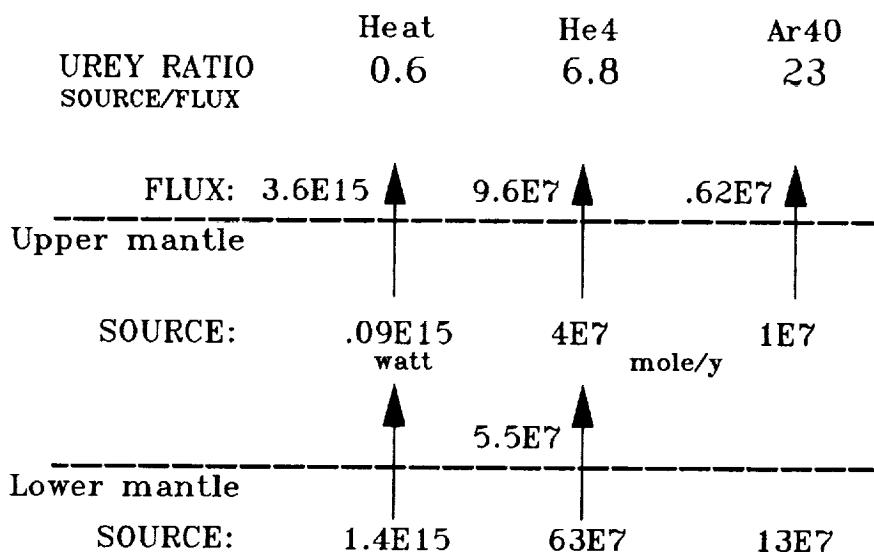
FIGURE 5   Fluxes of heat and gases and ratios of sources to fluxes.

material. The degree of partial melting can be deduced from the concentrations of the completely incompatible element potassium. Potassium concentration is enhanced in ocean crust by a factor of 10, more or less, so we have approximately a 10% partial melt of 60 kilometers of upper mantle material to produce six kilometers of ocean crust. About the same increase by a factor of 10 can be expected in the concentration of ^{40}Ar, also presumably a completely incompatible element. New ocean crust is generated at the rate of 3 km^2 per year. To produce the ^{40}Ar degassing flux of 6.2 × 10^6 mole per year it would be necessary to extract ^{40}Ar from just the top 250 meters of ocean crust. This extraction presumably occurs by interaction between sea water and the ocean crust. The ^{40}Ar does not diffuse directly out of the crust or bubble out of the magma. It must be extracted by leaching at relatively shallow depths in the crust. During the lifetime of the sea floor before subduction, heat will be extracted from a lithospheric layer approximately 60 kilometers thick, but Ar will be extracted only from 250 meters of ocean crust. This thickness of crust is equivalent, before partial melting, to 2.5 kilometers of upper mantle, so the release of Ar is about 25 times less efficient than the release of heat. The flux data indicate that radiogenic rare gases are accumulating in the mantle. The degassing process now operating is inefficient and slow. It seems that the process that originally degassed the upper mantle completely and rapidly must have been markedly different from the process now operating.

Ar40 degassing flux
= 6.2E6 mole/y
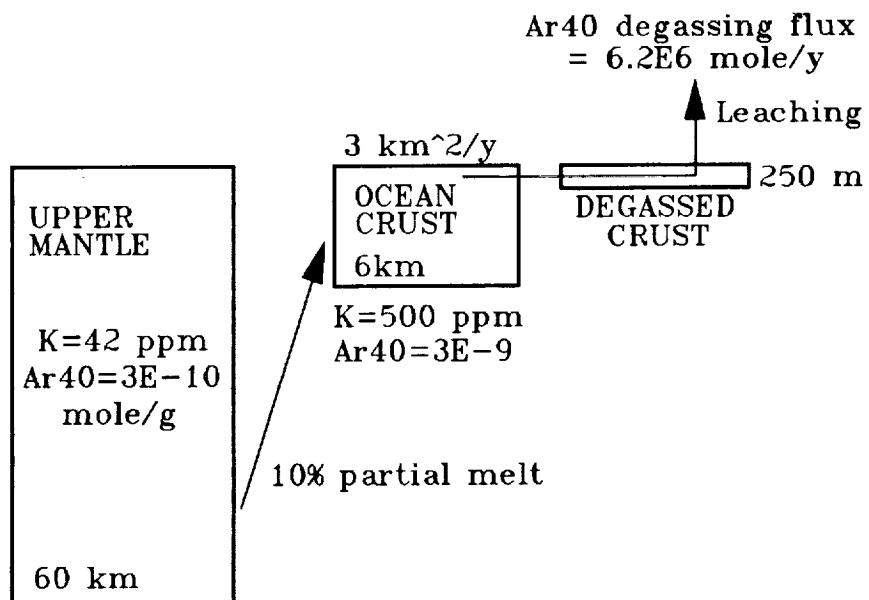
FIGURE 6 Degassing is inefficient compared with the extraction of heat.

CARBON DIOXIDE

To what extent can the rare gas results be applied to more important constituents of the atmosphere? It turns out that there is significant information concerning carbon dioxide (Des Marais 1985; Marty and Jambon 1987). The data and results are summarized in Figure 7. The flux of carbon dioxide from mantle to atmosphere today is 2×10^{12} mole per year. The flux of ^{36}Ar is 250 mole per year, so the ratio of the fluxes is 8×10^9. On the other hand, the ratio of the amounts in atmosphere, ocean, and continental crust is 1.8×10^6. The ratio of fluxes is very much larger than the ratio of amounts now present in the surface layers of the Earth. Carbon is missing from the surface layers relative to argon.

This conclusion can be seen also in the accumulation times calculated by dividing the flux into the amount. Carbon would accumulate at present rates in 5×10^9 years, but it would take 2.2×10^{13} years for the ^{36}Ar now in the atmosphere to accumulate at the present flux. The conclusion is that while ^{36}Ar was massively degassed earlier in Earth history, carbon did not participate in this early degassing. If carbon was rapidly released from the mantle early in Earth history it was just as rapidly returned to the mantle.

Carrying this analysis further it can be concluded that carbon is a lithophile and not an atmophile element. From the flux ratio of carbon to

Accumulation times: 5E9   2.2E13 y
        AMOUNT/FLUX

            Amounts: 1E22   5.6E15 mole
                       Ratio C/Ar = 1.8E6

CRUST AND
ATMOSPHERE   C ↑   ↑ Ar 36

───

MANTLE

            Fluxes: 2E12   250 mole/y
                      Flux ratio C/Ar = 8E9
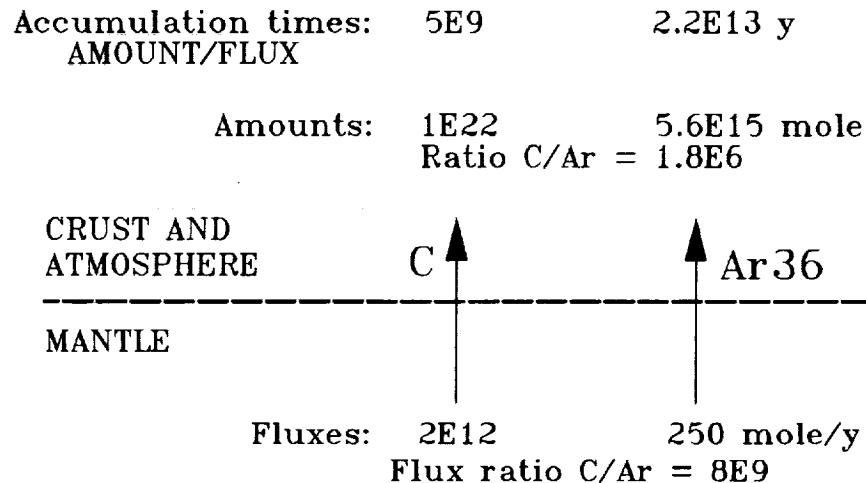
FIGURE 7 Compared to argon, carbon is deficient in the crust and atmosphere.

[36]Ar and the concentration of [36]Ar in the upper mantle we can calculate the concentration of carbon in the upper mantle. The value is 1×10^{-4} mole/g. This calculation assumes that carbon is not more mobile than Ar, surely a reasonable assumption. If carbon is less incompatible than Ar the required upper mantle concentration of carbon would be larger. From the concentration and the mass of the upper mantle I calculate that there are 1.4×10^{23} moles of carbon in the upper mantle. The amount in the crust and atmosphere and ocean is only 1×10^{22} mole (Wilkinson and Walker 1989). Therefore less than 10% of upper-mantle carbon has been degassed. By way of contrast, more than 99% of upper mantle [36]Ar has been degassed. Continuing the analysis and assuming that the lower- mantle concentration is given by the upper-mantle concentration augmented by crustal carbon mixed back in, the total amount of carbon in the mantle is 4.2×10^{23} mole, which is 42 times the amount in crust, ocean, and atmosphere. It seems that the fate of most carbon dioxide released from the mantle is to be incorporated into oceanic crust in weathering reactions and to be carried back into the mantle on subduction. Only a small fraction of the carbon is captured in the exogenic system as cratonic carbonate rocks. The average carbon concentration in continental crust is 5×10^4 mole/g. The concentration in the upper mantle is 1×10^4 mole/g. Therefore, the crust is only moderately enriched in carbon dioxide relative to the upper mantle and, unlike the rare gases, carbon is a modestly incompatible element.

CONCLUSION

The rare gas data indicate that there was early, thorough degassing of the upper mantle, but that there remain large amounts of primordial rare gases in the undegassed, lower mantle reservoir. The time scales and rates of degassing and depletion are very different. Depletion and continental growth occurred much later in Earth history than degassing. Degassing today, by weathering of the sea floor, is a slow and inefficient process and could hardly have provided the rapid and total early degassing that apparently occurred. Carbon dioxide did not degas like the rare gasses and is only modestly incompatible in the upper mantle. With the example of carbon dioxide in mind, we must be cautious about deducing degassing histories of other important atmospheric gases like nitrogen or water from the rare gas data. In the absence of relevant observations it is not immediately clear whether other atmospheric gases have behaved more like argon or more like carbon dioxide. By analogy with the Earth, it does seem likely that large amounts of both rare gases and carbon dioxide may be retained within the interiors of Mars and Venus. This possibility must be kept in mind in the study of the origin of planetary atmospheres. I do not feel that we yet have a satisfactory description even in qualitative terms of the origin of the Earth and the atmosphere. The challenge is to reconcile the ideas of planetary growth by accretion, impact degassing during the course of accretion, the origin of the Moon by a giant impact, and the data described in this paper concerning the degassing history of the mantle.

ACKNOWLEDGEMENTS

REFERENCES

Allegre, C.J., S.R. Hart, and J.-F. Minster. 1983. Chemical structure and evolution of the mantle and continents determined by inversion of Nd and Sr isotopic data, II. Numerical experiments and discussion. Earth Planetary Sci. Letters 66:191-213.

Allegre, C.J., T. Staudacher, P. Sarda, and M. Kurz. 1983. Constraints on evolution of Earth's mantle from rare gas systematics. Nature 303:762-766.

Allegre, C.J., T. Staudacher, and P. Sarda. 1987. Rare gas systematics: formation of the atmosphere, evolution and structure of the Earth's mantle. Earth Planetary Sci. Letters 81:127-150.

Anderson, D.L. 1989. Composition of the Earth. Science 243: 367-370.

Des Marais, D.J. 1985. Carbon exchange between the mantle and the crust and its effect upon the atmosphere: today compared to Archean time. Pages 602-611. In: Sundquist, E.T., and W.S. Broecker (eds.). Natural Variations in Carbon Dioxide and the Carbon Cycle. American Geophysical Union, Washington, D. C.

Jacobsen, S.B. 1988. Isotopic and chemical constraints on mantle-crust evolution. Geochim. Cosmochim. Acta 52: 1341-1350.

Marty, B., and A. Jambon. 1987. $C/^3He$ in volatile fluxes from the solid Earth: implications for carbon geodynamics. Earth Planetary Sci. Letters 83:16-26.

Oxburgh, E.R., and R.K. O'Nions. 1987. Helium loss, tectonics, and the terrestrial helium budget. Science 237:1583-1588.

O'Nions, R.K., and E.R. Oxburgh. 1983. Relationships between chemical and convective layering in the Earth. J. Geological Soc. London 144:259-274.

O'Nions, R.K., and E.R. Oxburgh. 1983. Heat and helium in the Earth. Nature 306:429-431.

Pollack, H.N. 1980. The heat flow from the Earth: a review. Pages 183-192. In: Davies, P.A., and S.K. Runcorn (eds.). Mechanisms of Continental Drift and Plate Tectonics. Academic Press, New York.

Sarda, P., T. Staudacher, and C.J. Allegre. 1985. $^{40}Ar/^{36}Ar$ in MORB glasses: constraints on atmosphere and mantle evolution. Earth Planetary Sci. Letters 72:357-375.

Wilkinson, B.H., and J.C.G. Walker. 1989. Phanerozoic cycling of sedimentary carbonate. American J. Sci. 289:525-548.

The Role of Impacting Processes in the Chemical Evolution of the Atmosphere of Primordial Earth

Lev M. Mukhin and M.V. Gerasimov
Institute of Space Research

ABSTRACT

The stability of the chemical composition of the planets' atmospheres on any time scale is determined by the ratio of source and sink strengths for various atmospheric constituents. Beginning with the Earth's formation and continuing throughout its history, these ratios have undergone significant alterations. Such changes are determined by various physical processes that are tied to the evolution of the Earth as a planet. A complete theory of the origin and evolution of the Earth's atmosphere is still far from complete at this time. This is due, in particular, to a certain randomness in the selection of a number of important physical parameters of the preplanetary cloud. They include, for example, the time scale for the dissipation of the gaseous nebula component; the planet accretion scale; and the chemical composition and distribution by size of planetesimals. It will therefore be useful to consider some physical constraints on the process by which the Earth's atmosphere was formed.

Currently existing paleontological data offer clear evidence of the presence of life on Earth 3.5 billion years ago (Schopf and Packer 1987). Furthermore, we can make the fundamental conclusion from the analysis of carbon isotope ratios $^{13}C/^{12}C$ that an almost contemporary biogeochemical carbon cycle (Schidlowski 1988) existed on Earth 3.8 billion years ago. Moreover, there is some reason to believe that this last dating could be pushed further back to four billion years (Schidlowski, personal communication). There is no doubt that both the prebiological evolution processes and the global biogeochemical carbon cycle can only occur in a sufficiently

dense atmosphere and hydrosphere that have already formed or are forming. Taking into account data from isotope systematics (Faure 1986), with which we can estimate the Earth's age at 4.55 to 4.57 billion years old, we conclude that the time scale for prebiological evolution and the emergence of life on Earth was sufficiently brief: possibly less than 0.5 billion years. This fact means, however, that the Earth's protoatmosphere must have been formed prior to the appearance of the biogeochemical cycle, that is, over a period of time significantly less than 0.5 billion years. Additional indirect evidence of the early emergence of the Earth's atmosphere can be found in analyzing the isotope relationships of the noble gases Ar and Xe (Ozima and Kudo 1972; Ozima 1975; Kuroda and Crouch 1962; Kuroda and Manuel 1962; Phinney *et al.* 1978).

Therefore, existing and observed data provide evidence of the very early formation of a dense atmosphere on Earth. We shall consider possible scenarios for the formation of the Earth's early atmosphere and its initial chemical composition.

EARTH'S INITIAL ATMOSPHERE

A model for Earth's early atmosphere, formed from solar composition gas captured gravitationally during the final stages of accretion, was explored in Hayashi's studies (1981, 1985). Hydrogen and helium are the primary components of this atmosphere. According to the estimates of several authors, the total mass of the initial atmosphere could have reached 10^{26}-10^{31} g (Hayashi *et al.* 1985; Cameron and Pine 1973; Lewis and Prinn 1989). However, the inference by these models that the entrapment process occurred isothermally may lead to significant error: they overlook the heating up of the gas during accretion (Lewis and Prinn 1984). We will note that there are some additional difficulties in a model of an isothermal, initial atmosphere. They stem from the diffusive concentration of heavy gases in the initial atmosphere (Walker 1982). The pressure of the initial atmosphere for Earth is only 0.1 bar in the more realistic models of the adiabatic, gravitational capture of gas from a protoplanetary nebula (Lewis and Prinn 1984).

Clearly, an initial atmosphere could only have formed if the processes of gas dissipation from the protoplanetary nebula had not ceased by the time Earth's accumulation had concluded. Gas dissipation from a protoplanetary nebula is determined by a solar wind from a young T Tauri Sun and EUV (Sekiya *et al.* 1980; Zahnle and Walker 1982; Elmegreen 1978; Horedt 1978;). Canuto *et al.* (1983) have estimated the time scale for dissipation of the gaseous component of a protoplanetary nebula, using observed data on T Tauri star luminosity. Their estimates show that this time scale is not more than 10^7 years. It may only be several million years, beginning with

the onset of the convective phase in the history the Sun's development. This time scale is appreciably shorter than the approximately 10^8 years estimated for Earth's accumulation (Safronov 1969; Wetherill 1980). Hence, current theories of stellar evolution, coupled with observed astronomical data, raise the possibility that only a very weak initial atmosphere existed at the very inception of planets' accumulation process. This is a very serious argument against the formation of an initial atmosphere on Earth as proposed by Hayashi.

Nevertheless, one should note that the accuracy of estimates of the time scales for a significant portion of cosmogonic processes at the early stages of protosolar nebula evolution is not reliable enough to rule out, with absolute certainty, the possibility that an appreciable initial atmosphere existed on Earth. According to this scenario, the initial atmosphere encounters considerable additional difficulties stemming from its chemical composition. Adopting, again, the adiabatic model of the gravitational capture of gas from a protoplanetary nebula (Lewis and Prinn 1984), we can see that with a 0.1 bar value of the pressure of the primordial atmosphere, the nitrogen levels in it would be 10^5 times lower than present levels. At the same time, neon levels would exceed current neon atmospheric levels by approximately 40 times. Walker's more detailed computations, where he accounts for the radiative-convective structure of the primordial atmosphere (1982), also show an inconsequential pressure for this kind of atmosphere at the surface (approximately 0.2 bars). This means that arguments raised by Lewis and Prinn (1984) against a primordial atmosphere remain valid in this instance. Hence, various scenarios for a dense initial atmosphere appear to be highly improbable for the above reasons. Therefore, the very rapid formation of a dense Earth atmosphere was apparently a function of different physical processes.

CONTINUOUS DEGASSING

One of the suggested mechanisms for the rapid formation of the atmosphere is the intense, continuous degassing of Earth in the conventional sense of this term, including magmatic differentiation and volcanism (Walker 1977; Fanale 1971). There is still no answer to the question of whether continuous degassing of the Earth could have been intensive enough to allow for the formation of the atmosphere and the hydrosphere over a very short time span ($< 10^8$ years). Such a possibility exists where there is a strong, overall heating of the planet during its accretion. However, this entails the inclusion of a number of additional inferences regarding a very brief accretion scale: $< 5 \times 10^5$ years (Hanks and Anderson 1969).

Walker's recent analysis of the process of continuous degassing (Walker, this volume) shows that the present rate of degassing is clearly insufficient

to be responsible for forming the atmosphere in a very brief time scale. At the same time, Walker's proposed numerical estimate of the accumulation rate of hot lava on primordial Earth, which could have provided for the necessary intensity of degassing (1.3×10^{19} g/yr), appears to be unjustifiably high. Such intensive volcanism infers that $\sim 1.3 \times 10^{27}$ g of magma would pour out over 10^8 years of accumulation on the surface, or a quarter of the entire present mass of the Earth.

Under the currently adopted planetary accumulation models with an approximate 10^8 year time scale, such a hypothesis would only be justified in the case of a gigantic impact (Kipp and Melosh 1986; Wetherill 1986). There is no question that a considerable portion of the planet would have melted as a result of a gigantic impact which would have released a huge quantity of gases. This amount would have been sufficient to have formed an atmosphere. Yet, there is no detailed, physical-chemical model of this process at the present time. Furthermore, it cannot be considered that the very fact of a giant impact in the Earth's history has been firmly established. Clearly, if such a catastrophic event did actually take place in the earliest history of our planet, its consequences would have been so great that they would have been reflected in the geochemical "records." However, the scenario explored by Walker (this volume) of two reservoirs of volatiles, one of which is virtually entirely degassed (the upper mantle), and the other which has conserved its store of volatiles (lower mantle), is difficult to reconcile with the idea of a gigantic impact. Truly, such a clear separation of the two mantle reservoirs, whose presence has been determined with sufficient certainty from observed data (Allegre *et al.* 1987) is difficult to expect in the case of a gigantic impact. An attempt should be made to analyze the possible geochemical consequences of a gigantic impact and reconcile them with existing data. For now, a gigantic impact remains a reality only in computational models of the evolution of preplanetary swarms of bodies. However, this episode is only a specific case of the impact-induced degassing of matter. Yet impact-induced degassing was, apparently, a determining physical process which led to the very rapid formation of protoEarth's atmosphere.

IMPACT PROCESSES AND THE EARTH'S PROTOATMOSPHERE

The role of impact processes in forming Earth's protoatmosphere was discussed more than 20 years ago in Florensky's study (1965). During the years that followed, this idea was developed in a number of studies (Arrhenius *et al.* 1974; Benlow and Meadows 1977; Gerasimov and Mukhin 1979; Lang and Ahrens 1982; Gerasimov *et al.* 1985). Impact-induced degassing is rooted in the idea of Earth accumulation from large solids. This idea does not require the inclusion of any serious additional proposals. It is

this circumstance that brings us to the conclusion that impact degassing was a real mechanism in the formation of Earth's atmosphere and hydrosphere. Departing from the gigantic impact issue, we can see that the maximum dimension of those bodies which regularly fell on the planet, and from which Earth was formed, could have attained hundreds of kilometers (Safronov 1969). These dimensions fit with estimates of the diameters of impacting bodies which formed the largest craters on the terrestrial planets and their satellites (O'Keef and Ahrens 1977). The velocity of the collision of a random body of a preplanetary swarm with the embryo of primordial Earth is no less than the escape velocity. Therefore, the range of velocities of planetesimal collision with a growing planet embryo varies from meters per second at the initial stages of its expansion, to $\geq$ 12 kilometers per second at the final stages of accumulation. In reality, collision velocities could have been significantly higher in the case of a collision between Earth and bodies escaping from the asteroid belt, or with comets. In this case, when the mass of an expanding Earth reached approximately 10% of its current mass, the escape velocity became equal to roughly five kilometers per second; the melting of silicate matter began during a collision between planetesimals and an embryonic planet (Ahrens and O'Keef 1972). Beginning with the point where the mass of the embryo was equal to approximately 0.5 of the mass of present-day Earth, impact processes were paralleled by the partial vaporization of silicates (Ahrens and O'Keef 1972). The more high-speed planetesimals, reaching a speed of more than 16-20 kilometers per second, were completely vaporized. Several works (Ahrens and O'Keef 1972; Gault and Heitowit 1963; McQueen *et al.* 1973; Gerasimov 1979; O'Keef and Ahrens 1977) analyze in detail the physics of how colliding matter is heated as a shock wave passes through that matter and estimate the efficiency of impact-induced degassing. They demonstrate that impact degassing is most efficient when melting and partial vaporization of silicate matter begin. It is worth noting that the release of the primary volatile components begins long before the point when the planet's accumulation process reaches collision velocities corresponding to the melting point of matter. Pioneering works (Lang and Ahrens 1982, 1983; Katra *et al.* 1983) establish the beginning of volatile loss at extremely low collision velocities of approximately one kilometer per second. Using the Merchison meteorite as an example, the authors established that about 90% of the initial amount of volatiles is already lost at a collision velocity of 1.67 kilometers per second (Tuburczy *et al.* 1986). This loss is due to breakdown of the meteorite's water-, carbon-, and sulphur-containing minerals at impact. These experiments are proof of the beginning of volatile release at a very early stage of the Earth's accumulation. Its mass was less than 0.01 of its final mass. Water and carbon dioxide are the main constituents involved in the processes of shock-induced dehydration and decarbonatization of minerals.

As the mass of the embryonic planet increases, escape velocity rises. Consequently, there is also a rise in the velocity at which planetesimals fall. There occurs a corresponding increase in the heating of matter of the planetesimal and the surface of the planet in the central shock zone. The nature of degassing processes undergoes qualitative changes. The shock process becomes a considerably high-temperature one. Chemical reactions occurring in the vaporized cloud become increasingly significant. High-temperature chemical processes begin to play a predominant role during the final stages of Earth's accumulation, instead of the relatively simple processes involved in the breakdown of water-containing compounds and carbonates. Colliding matter undergoes total meltdown in the central zone of impact at collision velocities of five to eight kilometers per second: volatile components are released from the melted matter and interact with it. The chemical composition of the released gases must correspond to the volcanic gases for magma of corresponding composition and temperature. CO_2, H_2O, and SO_2 will clearly be the primary components in such a gas mixture. The gases CO, H_2, H_2S, CH_4, and others may be competitors to these components, depending on the extent to which planetesimal matter is reduced (Holland 1964). Where collision velocities exceed eight to nine kilometers per second, vaporization at impact of a portion of planetesimal matter becomes significant.

Vaporization of silicate matter is supported by thermodissociation of planetesimal mineral constituents. As a result of this process, a considerable quantity of molecular and monatomic oxygen is released to the cloud of vaporized matter. Melt-vapor equilibrium determines the conditions for vaporization. The characteristic vaporization temperature for silicates is approximately 3000-5000 K. Vapor pressure is approximately 1-100 bars (Bobrovskiy et al. 1974). In these conditions, thermodynamic equilibrium in a gaseous phase is reached over a time scale $t_{chem} < 10^{-5}$ seconds (Gerasimov et al. 1985). Therefore, when large-scale impact episodes occur, where the characteristic time scale for the expansion of a cloud of vaporized matter tcool is counted in seconds, thermodynamic equilibrium ($t_{cool} \gg t_{chem}$) is clearly reached at the initial stage of the expansion of such a cloud. Consequently, gases are formed in the cloud from the entire range of volatile elements present, regardless of what form they displayed in the initial matter. In such conditions, H_2O and CO_2 will be formed inside the dense, hot cloud if hydrogen, carbon, and oxygen are present in it. This will occur regardless of the initial presence in the planetesimals of carbonates and hydrated minerals. At the same time, where carbonates and hydrated minerals are present in planetesimals, a portion of the hydrogen and carbon in these minerals will be used to form other hydrogen- and carbon-containing components. Less H_2O and CO_2 will be produced than with the simple thermal breakdown of minerals. As

the vapor-gas cloud expands, its density and temperature drop, and t_{chem} increases. The expansion process commences when $t_{cool} \sim t_{chem}$. This is the point where the reaction products undergo chemical hardening, since the time scale for chemical reactions becomes greater than the characteristic time scale for cloud expansion as the cloud expands further. Gases with a chemical composition corresponding to the moment of hardening are mixed with atmospheric gases, interacting with both these gases and bedrock from the uppermost layers of the planet (regolith). If the atmosphere is dense, mixing of gases in an expanding cloud with atmospheric gases may occur earlier, when overall pressure of the expanding cloud $pv2/2 + P$ ($p =$ density, $v =$ mass velocity, $P =$ pressure in the vapor-gas cloud) becomes approximately the pressure in the atmosphere. The portion of the vaporized silicate matter α, which is condensed in the cloud by the time constituents harden (occurring at a given temperature T^*), is defined by the simple ratio (Anisimov *et al.* 1970):

$$\alpha \sim 1 - T*/T_o,$$

where T_o denotes temperature in the vaporized cloud as it begins to expand. Estimates demonstrate that for impact of a planetesimal with a dimension of approximately 100 kilometers, the value is $T^* \sim 2000$ K. Therefore, for a vaporization temperature of $T_o \sim 5000$ K, the value is $\alpha \sim 60\%$. This means that approximately one half of molecular and monatomic oxygen, comprising $\sim 30\%$ of the cloud's pressure, (Gerasimov *et al.* 1985) will be released into the atmosphere with each impact. Clearly, the impact-vaporization mechanism is a powerful source of free oxygen in Earth's earliest atmosphere.

The chemistry of a high-temperature, gas-vapor cloud is too complex to be able to judge it solely in terms of the thermodynamical equilibrium in its gaseous phase. Condensation of silicate particles and catalytic activity occur during expansion and cooling of the vaporized cloud. These processes can significantly alter the equilibrious gas chemical composition. This circumstance imposes certain requirements on the search for both theoretical and experimental approaches to the study of chemical processes in a cloud of impact-vaporized matter.

Studies (Gerasimov *et al.* 1984; Gerasimov and Mukhin 1984; Gerasimov *et al.* 1987) used laser radiation to examine the chemical composition of gases which form during high-temperature, pulse-induced vaporization of the Earth's rockbed and meteorites. They demonstrated that molecular oxygen is actually the most abundant constituent in a cloud of vaporized matter (Gerasimov *et al.* 1987), and that it determines the chemical processes occurring in the cloud. Regardless of how reduced the initial matter is, the primary volatile elements H, C, and S are released as oxides: H_2O,

CO, CO_2 ($CO/CO_2 \geq 1$), and SO_2. In addition to the oxides, a certain amount of reduced components and organic molecules is formed in the cloud: H_2, H_2S, CS_2, COS, HCN, saturated, unsaturated, and aromatic hydrocarbons from CH_4 to C_6 and CH_2CHO. Molecular nitrogen is released. The vaporization-induced gaseous mixtures for samples belonging to both crust and mantle rock, as well as for conventional and carbonaceous chondrite, are qualitatively homogenous. This is seen in the preponderant formation of oxides, and in both the comparable (within one order of magnitude) ratios of the gases CO/CO_2 and the correlation between the various hydrocarbons. The gas mixtures formed at high temperatures are in nonequilibrium for normal conditions. Therefore, their chemical composition will be easily transformed under the impact of various energy sources.

One should note that the passage of a planetesimal through the atmosphere exerts a substantial influence on the formation of the atmosphere's chemical composition. Studies (Fegly *et al.* 1986; Prinn and Fegley 1987; Fegley and Prinn 1989) have analyzed this issue in the greatest detail in recent years. In the physics of the process, a body of large dimension (approximately 10 kilometers in diameter) passes through primordial Earth's atmosphere at a speed of approximately 20 kilometers per second; shock waves send a large amount of energy into the atmosphere. If we put the planetesimal density at ~ 3 g/cm^3 (which matches the chondrite composition), and the angle of entry into the atmosphere at 45°, as first approximation, we would have 2.2×10^{27} ergs. This figure is 0.07% of the energy of an asteroid. The energy passes directly into the atmosphere as the body flies through, and even more energy ($\sim 3.2 \times 10^{29}$ ergs) is "pumped" into the atmosphere by a supersonic discharge of matter from the impact crater (Fegley and Prinn 1989). The shock wave front (formed in the atmosphere during this process) compresses and heats the atmospheric gas to several thousand degrees Kelvin. It is clear that various thermochemical and plasmochemical reactions are occurring in this region. Due to recombination processes, new compounds are formed as cooling occurs. This substantially alters the initial chemical composition of the atmosphere.

We can estimate the chemical composition of the gas mixtures at high temperatures, using the standard methods of thermodynamical equilibrium (given the presupposition that the time scale for the breakdown of a given molecule is less than the time required to cool the elementary gas exchange. The most detailed computations of these processes were made in the studies of Fegley *et al.* (1986) and Fegley and Prinn (1989). It seems obvious that these findings must be critically dependent on the initial chemical composition of an unperturbed atmosphere. The atomic ratio C/O is an important parameter here: it determines the "oxidized" and "reduced" state of the atmosphere. Fegley and Prinn (1989) demonstrated that if

C/O > 1 (reduced atmosphere), precursors of biomolecules, such as HCN and H_2CO, are formed as the asteroid passes through this atmosphere. Nitrogen oxides appear in the oxidized atmosphere instead of prussic acid, supporting the formation of nitric acid if rain falls. Fegley and Prinn (1989) considered several possible chemical compositions for unperturbed atmospheres, and they calculated the commensurate alterations in the composition owing to the passage of large bodies through the atmosphere.

I would like to make the following comment regarding their work. The authors used computation methods employed for purely gaseous reactions. Heterogeneous catalysis on mineral particles (present in the atmosphere during the impact-induced discharge of matter) must play a significant role in the actual natural process. Heterophase reactions must considerably affect the evolution of the atmosphere's chemical composition during impact reprocessing. However, it is quite difficult to account for these reactions at this time. An account of the fluxes of such important components as prussic acid and formaldehyde is an unquestionable achievement in the work of Fegley and Prinn. Their efforts made it possible to estimate the stationary concentrations of these components in a modeled early atmosphere. We shall note that the numerical values of the strength of the source of cyanic hydrogen formation in atmospheric reprocessing and in experiments on laser modeling of the processes of shock-induced vaporization are comparable (Mukhin *et al.* 1989). Hydrocarbon and aldehyde output in the latter case is significantly higher.

THE EVOLUTION OF AN IMPACT-GENERATED ATMOSPHERE

Gas fluxes from the atmosphere present one of the most difficult issues related to the early evolution of the Earth's atmosphere. There are no accurate estimates at this time of the reverse fluxes of released gases to the surface rock of the young planet. It is therefore impossible to reliably estimate the stationary concentrations of these gases in the protoatmosphere.

The strength of a shock source of atmospheric gases is so great that the atmospheric accumulation of gases released during the fall of planetesimals at the early stage of accretion would create a massive atmosphere from water vapors. A runaway greenhouse effect would develop. Such a scenario was developed in several studies (Abe and Matsui 1985; Matsui and Abe 1986; Zahnle *et al.* 1988). The massive atmosphere "locks in" heat released by the impact. This in turn triggers the heating of the atmosphere and melting of the upper layers of the planet's silicate envelope. Zahnle *et al.* (1988) have estimated that the massive water atmosphere should condense and form an ocean as planetesimals cease falling. The problem however,

requires that all atmospheric components be taken into account, particularly CO_2 and SO_2. These gases could also support a greenhouse effect, making it irreversible under the effects of solar radiation alone. The authors considered the processes of water dissolution in magma and thermal volatization as examples of fluxes for water vapors. Lang and Ahrens (1982) looked at the hydration reactions of phorsterite and enstatite transported by planetesimals in order to determine H_2O flux rates. It should be noted that other flux mechanisms may be significant for primordial Earth.

An actualistic approach to the problem of fluxes is not entirely justified. To assess flux rates it is necessary to take into account physical-chemical conditions which correspond to the period of accumulation. One of the significant factors for the outflow of atmospheric gases to the regolith is the formation of a large quantity of condensed silicate particles from planetesimal matter vaporized during impact. With collision velocities of approximately 12 kilometers per second, approximately 30% of the matter in the central zone of impact is vaporized (Gerasimov 1979). Particle condensation takes place as the vaporized cloud expands and cools. These particles of small dimensions are discharged into the atmosphere, settle there, and fall to the surface of the accreting planet. Jakosky and Ahrens (1979) estimated rates of flux from the atmosphere for water vapor to a meter layer of regolith consisting of approximately 50 micron particles. Comparison of the rates at which water is released into the atmosphere during impact-induced degassing and absorbed by regolith has yielded a value approximately 0.008 bars for equilibrium pressure of water vapor in the atmosphere. This means that virtually all of the released water flows into the regolith. A similar view was put forward by Sleep and Langan (1981).

A structural-chemical analysis of condensed particles forming in model laser experiments (Gerasimov *et al.* 1988) has demonstrated that the condensed matter is nonequilibrious to a significantly great degree. It concurrently has phases with a high degree of silicon-oxygen tetrahedron (quartz type) polymerization and ostrov-type phases. It also includes silicon phases with an intermediary degree of Si^{2+} oxidation and metal silicon Si^0. The condensed matter primarily contains oxidized iron FeO, as well as metal iron Fe^0. These facts are possible evidence of high chemical activity of the condensed particles which form planets' regolith during accumulation. A number of experiments illustrate that the condensed particles in contact with water during heating may easily form laminated silicates that absorb appreciable quantities of water (Nuth *et al.* 1986; Nelson *et al.* 1987). Chemically active atmospheric gas components may also react with such regolith particles and have flux rates that are significantly higher than on present-day Earth with its oxidized crust.

The rapid (not more than a few seconds) and virtually complete

absorption by condensed particles of oxygen released during vaporization is an important experimental finding (Gerasimov 1987). This is evidence that, despite the impact-induced release of large amounts of oxygen into the atmosphere, molecular oxygen was absent in the primordial atmosphere: it was almost instantaneously absorbed by the regolith. It is possible that the same fate (albeit with greater time scales) is also characteristic for other chemically active gases. In the future, absorbed gases will be repeatedly released into the atmosphere owing to shock-induced processing of planet surface matter by falling planetesimals. Chemically inert gases, such as the noble gases and molecular nitrogen, may have accumulated in the protoatmosphere. Therefore, the degassing fate of noble gases and chemically active gases may have been entirely different (Gerasimov *et al.* 1985).

If we extrapolate the conditions on earliest Earth to the following period of "continuous" degassing, with a weaker source of atmospheric gases, atmospheric density would have been extremely small. This is due to the conservation of high flux rates. The appearance of a volatile-enriched (particularly oxygen) protocrust is one of the most important conditions for atmospheric stabilization. The appearance of an acidified protocrust ensures low flux rates for the majority of atmospheric gases and the accumulation of an ocean. In turn, the formation of an ocean also governs the composition and density of the atmosphere. The appearance of a protocrust is currently being considered in models of magmatic differentiation (Taylor and McLennan 1985). The earlier appearance of a protocrust in this model runs up against the same difficulties as the catastrophic formation of an atmosphere in the continuous degassing model. One possible avenue for the appearance of a protocrust is shock-induced differentiation (Mukhin *et al.* 1979; Gerasimov *et al.* 1985), as a result of which the atmosphere and the protocrust are formed within the same process of accumulation. However, the question of shock differentiation has yet to be theoretically analyzed in-depth and requires additional investigation.

The shock-degassing source was operative virtually throughout the planets' entire accumulation process. The formation of a sufficiently dense atmosphere at the earlier stages of accumulation of the Earth and the terrestrial planets is problematic owing to the possible rapid flux of atmospheric components to the regolith. Other factors also had an effect, such as shock-induced "cratering" of the atmosphere (Walker 1986), intensive T Tauri-like solar wind, EUV, and thermal volatilization. The probability of the formation of a dense atmosphere increases with the accumulation of the planet's mass, since by the time accretion is completed there is not likely to be any action of intensive solar wind and EUV. The loss of gases from a planet with a large gravitating mass is also difficult. Uncertainties in estimating the density of an impact-generated atmosphere at the final stage

of accretion are largely related to uncertainties in estimating the rates at which gases flow into the regolith.

Therefore, impact-induced degassing, despite the possible parallel input of an accretive source and a continuous degassing source was clearly the most probable and important source of atmospheric gases during the earliest epoch of the Earth's evolution: the period of its accumulation. Achievements made in recent years in investigating the impact source allow us to assess its strength and the chemical composition of the gases that were released. Nevertheless, the question of the evolution of the composition and density of the impact-generated atmosphere continues to remain an open one, primarily due to ambiguities in estimating gas flux rates from the atmosphere during accumulation. Therefore, it is our view that fluxes are one of the most important issues pertaining to the origin of Earth's earliest atmosphere.

REFERENCES

Abe, Y., and T. Matsui. 1985. The formation of an impact-generated H^2O atmosphere and its implications for the early thermal history of the Earth. Proc. 15th Lunar and Planet. Sci. Conf., part 2. J. Geophys. Res. Supp. 90:C545-C559.

Ahrens, T.J., and J.D. O'Keef. 1972. Shock melting and vaporization of lunar rocks and minerals. The Moon 4:214-249.

Allegre, C.J., T. Staudacher, and P. Sarda. 1987. Rare gas systematics: formation of the atmosphere, evolution and structure of the Earth's mantle. Earth and Planet. Sci. Letters 81:127-150.

Anisimov, S.I., Ya.A. Imas, G.S. Romanov, and Yu.V. Khodyko. 1970. The Effect of High-Power Radiation on Metals. Nauka, Moscow.

Arrhenius, G., B.R. De, and H. Alfven. 1974. Origin of the ocean. In: Goldberg, E.D. (ed.). The Sea. Vol. 5. Wiley, New York.

Benlow, A., and A.J. Meadows. 1977. The formation of the atmospheres of the terrestrial planets by impact. Astrophys. Space Sci. 46:293-300.

Bobrovskiy, S.V., V.M. Gogolev, B.V. Zamyshlyaev, and V.P. Lozhkina. 1974. Break-away velocity in a solid medium under the effect of a strong shock wave. The Physics of Combustion and Explosion 6:891-89.

Cameron, A.G.W., and M.R. Pine. 1973. Numerical models of the primitive solar nebular. Icarus 18:377-406.

Canuto, V.M., J.S. Levine, T.R. Augustsson, C.L. Imhoff, and M.S. Giampapa. 1983. The young Sun and the atmosphere and photo-chemistry of the early Earth. Nature 305:281-286.

Elmegreen, B.G. 1978. On the interaction between a strong stellar wind and a surrounding disk nebular. Moon and Planets 19:261-277.

Fanale, F.P. 1971. A case for catastrophic early degassing of the Earth. Chemical Geology 8:79-105.

Faure, G. 1986. Principals of Isotopes Geology. Sec. ed. Wiley, New York.

Fegley, B., Jr., and R.G. Prinn. 1989. Chemical reprocessing of the Earth's present and primordial atmosphere by large impacts. In: Visconti, G. (ed.). Interactions of the Solid Planet with the Atmosphere and Climate, in press.

Fegley, B., Jr., R.G. Prinn, H. Hartman, and G.H. Watkins. 1986. Chemical effects of large impacts on the Earth's primitive atmosphere. Nature 319(6051):305-308.

Florenskiy, K.L. 1965. On the initial stage of the differentiation of Earth's Matter. Geochemiya 8:909-917.

Gault, D.E., and E.D. Heitowit. 1963. The partition of energy for hypervelocity impact craters formed in rock. Proc. 6th Hypervelocity Impact Symp. 2:419-456.

Gerasimov, M.V. 1987. On the release of oxygen from the intensively shocked meteorites and terrestrial rocks (abstract). Pages 320-321. In: Lunar and Planetary Sci. 18. Lunar and Planetary Institute, Houston.

Gerasimov, M.V. 1979. On the mechanisms of impact degassing of planetesimals. Letters to AJ. 5(5):251-256.

Gerasimov, M.V., Yu.P. Dikov, L.M. Mukhin, and V.I. Rekharsky. 1988. Structural-chemical peculiarities of the state of silicon in the processes of shock metamorphysm (abstract). Pages 383-384. In: Lunar and Planetary Sci. 19. Lunar and Planetary Institute, Houston.

Gerasimov, M.V., and L.M. Mukhin. 1984. Studies of the chemical composition of gaseous phase released from laser pulse evaporated rocks and meteorites materials (abstract). Pages 298-299. In: Lunar and Planetary Sci. 15. Lunar and Planetary Institute, Houston.

Gerasimov, M.V., and L.M. Mukhin. 1979. On the mechanism forming the atmosphere of the Earth and the terrestrial planets at the stage of their accretion. Letters to AJ. 5(8):411-414.

Gerasimov, M.V., L.M. Mukhin, Yu.P. Dikov and V.I. Rekharskiy. 1985. The mechanisms of Earth's early differentiation. Vestnik of the USSR Academy of Sciences 9:10-25.

Gerasimov, M.V., L.M. Mukhin, and M.D. Nusinov. 1984. Study of the chemical composition of the gaseous phase formed under the effect of pulse laser radiation on certain rocks and minerals. Papers of the USSR Academy of Sciences 275(3):646-650.

Gerasimov, M.V., B.L. Satovsky, and L.M. Mukhin. 1987. Mass- spectrometrical analyses of gases originated during impulsive evaporation of meteorites and terrestrial rocks (abstract). Pages 322-323. In: Lunar and Planetary Sci. 18. Lunar and Planetary Institute, Houston.

Hanks, T.C., and D.L. Anderson. 1969. The early thermal history of the Earth. Phys. Earth Planet. Interiors 2:19-29.

Hayashi, C. 1981. Formation of planets. Pages 113-128. In: Sugimoto, D., D.Q. Lamb, and D.N. Schramm (eds.). Fundamental Problems in the Theory of Stellar Evolution. IAU Symposium No. 93. D. Reidel, Dordrecht.

Hayashi, C., K. Nakazawa, and Y. Nakagawa. 1985. Formation of the solar system. Pages 1100-1153. In: Black, D.C., and M.S. Matthews (eds.). Protostars and Planets. II. University of Arizona Press, Tucson.

Holland, H.D. 1964. On the chemical evolution of the terrestrial and cytherian atmospheres. Pages 86-101. In: Brancazio, P.J., and A.G.W. Cameron (eds.). The Origin and Evolution of Atmospheres and Oceans-101. Wiley, New York.

Horedt, G.P. 1978. Blow-off of the protoplanetary cloud by a TTauri-like solar wind. Astron. Astrophys. 64:173-178.

Jakosky, B.M., and T.J. Ahrens. 1979. The history of an atmosphere of impact origin. Proc. Lunar. Planet. Sci. Conf. 10:2727-2739.

Kipp, M.E., and H. Melosh. 1986. Origin of Moon: A preliminary numerical study of colliding planets (abstract). Pages 420-421. In: Lunar and Planetary Sci. XVII. Lunar and Planetary Institute, Houston.

Kotra, R.K., T.H. See, E.K. Gibson, F. Horz, M.H. Cintala, and R.H. Schmidt. 1983. Carbon dioxide loss in experimentally shocked calcite and limestone (abstract). Pages 401-402. In: Lunar and Planetary Sci. 14th. Lunar and Planetary Institute, Houston.

Kuroda, P.K., and W.A. Crouch, Jr. 1962. On the chronology of the formation of the solar system. 2. Iodine in terrestrial rocks and the xenon 129/136 formation in the Earth. J. Geophys. Res. 67:4863-4866.

Kuroda, P.K., and O.K. Manuel. 1962. On the chronology of the formation of the solar system. 1. Radiogenic xenon 129 in the Earth's atmosphere. J. Geophys. Res. 67:4859-4862.

Lang, M.A., and T.J. Ahrens. 1983. Shock-induced CO_2 production from carbonates and a proto-CO_2-atmosphere of the Earth (abstract). Pages 419-420. In: Lunar and Planetary Sci. 14th. Lunar and Planetary Institute, Houston.

Lang, M.A., and T.J. Ahrens. 1982. The evolution of an impact generated atmosphere. Icarus 51:96-120.

Lang, M.A., and T.J. Ahrens. 1982. Shock-induced dehydration of serpentine: First quantitative results and implications for a primary planetary atmosphere (abstract). Pages 419-420. In: Lunar and Planetary Sci. 13th. Lunar and Planetary Institute, Houston.

Lewis, J.S., and R.G. Prinn. 1984. Planets and their atmospheres: origin and evolution. In: Donn, W.L. (ed.). International Geophysics Series. Vol. 33. Academic Press, Inc., Orlando.

Matsui, T., and Y. Abe. 1986. Evolution of an impact-induced atmosphere and magma ocean on the accreting Earth. Nature 319:303-305.

McQueen, R., S. Marsh, J. Taylor, G. Fritz, and W. Carter. 1973. A formula for solid state based on the results of investigation into shock waves. Pages 299-427. In: Nikolaevskiy, B.N. (ed.). High Velocity Impact phenomena. Mir, Moscow.

Mukhin, L.M., M.V. Gerasimov, and Yu.P. Dikov. 1979. Model of the Earth's accretive differentiation (abstract). Pages 876-877. In: Lunar and Planetary Sci. 10. Lunar and Planetary Institute, Houston.

Mukhin, L.M., M.V. Gerasimov, and E.N. Safonova. 1989.Hypervelocity impacts of planetesimals as a source of organic molecules and of their precursors on the early Earth (abstract). In: Lunar and Planetary Sci. 20. Lunar and Planetary Institute, Houston.

Nelson, R., J.A. Nuth, and B. Donn. 1987. A kinetic study of the hydrous alteration of amorphous MgSiO smokes: Implications for cometary particles and chondrite matrix. Proc. 17th Lunar and Planet. Sci. Conf., part 2. J. Geophys. Res., Supp. 92:E657-E662.

Nuth, J.A., B. Donn, R. Deseife, A. Donn, and R. Nelson. 1986. Hydrous alteration of amorphous smokes: First results. Proc. 16th Lunar and Planet. Sci. Conf., part 2, J. Geophys. Res., Supp. 91:D533-D537.

O'Keef, J.D., and T.J. Ahrens. 1977. Impact induced energy partitioning, melting, and vaporiation on terrestrial planets. Proc. Lunar Sci. Conf. 8:3357-3374.

O'Keef, J.D., and T.J. Ahrens. 1977. Shock effects during the collision of large meteorites with the Moon. The mechanics of funnel formation during impact and explosion. Pages 62-79. In: Ishlinskiy, A.Yu., and G.G. Cherniy (eds.). Mechanika. What's New in Science from Abroad, No. 12. Mir, Moscow.

Ozima, M. 1975. Ar isotopes and earth-atmosphere evolution models. Geochim. et cosmochim. Acta 39:1127-1134.

Ozima, M., and K. Kudo. 1972. Excess argon in submarine basalts and the Earth-atmosphere evolution model. Nature Phys. Sci. 239:23.

Phinney, D., J. Tennyson, and V. Frick. 1978. Xenon in CO2 well gas revisited. J. Geophys. Res. 83:2313-2319.

Prinn, R.G., and B. Fegley, Jr. 1987. Bolide impacts, acid rain, and biospheric traumas at the Cretaceous-Tertiary boundary. Earth Planet Sci. Letters 83:1-15.

Safronov, V.S. 1969. Evolution of the Preplanetary Cloud and the Formation of the Earth and the Planets. Nauka, Moscow.

Schidlowski, M. 1988. A 3,800-million-year isotopic record oflife from carbon in sedimentary rocks. Nature 333(6171): 313-318.

Schidlowski, M. Private communication.

Schopf, J.W., and B.M. Packer. 1987. Early arhean (3.3 billion to 3.5 billion-year-old) microfossils from Warrawoona Group, Australia. Science 237:70-73.

Sekiya, M.K., Nakazawa, and C. Hayashi. 1980. Dissipation of the primordial terrestrial atmosphere due to irradiation of the solar EUV. Prog. of Theor. Phys. 64:1968-1985.

Sleep, N.H., and R.T. Langan. 1981. Thermal evolution of the Earth: Some recent developments. Adv. Geophys. 23:1-23.

Taylor, S.R., and S.M. McLennan. 1985. The continental crust: Its composition and evolution. Blackwell Scientific Publ, Oxford.

Tyburczy, J.A., B. Frisch, and T.J. Ahrens. 1986. Shock-induced volatile loss from a carbonaceous chondrite: Implications for planetary accretion. Earth and Planet. Sci. Letters 80:201-207.

Walker, J.C.G. 1990. Degassing. This volume.

Walker, J.C.G. 1982. The earliest atmosphere of the Earth. Precambrian Research 17:147-171.

Walker, J.C.G. 1986. Impact erosion of planetary atmospheres. Icarus 68:87-98.

Walker, J.C.G. 1977. Evolution of the atmosphere. Macmillan, New York.

Wetherill, G.W. 1986. Accumulation of the terrestrial planets and implications concerning lunar origin. Pages 519-550. In: Hartmann, W.K., R.J. Phillips, and G.J. Taylor (eds.). Origin of the Moon. Lunar and Planetary Institute, Houston.

Wetherill, G.W. 1980. Formation of the terrestrial planets. Ann. Rev. Astron. Astrophys. 18:77-113.

Zahnle, K.J., J.F. Kasting, and J.B. Pollack. 1988. Evolution of a steam atmosphere during Earth's accretion. Icarus 74:62-97.

Zahnle, K.J., and J.C.G. Walker. 1982. Evolution of solar ultraviolet luminosity. Rev. Geophys. Space Phys. 20:280-292.

Lithospheric and Atmospheric Interaction on the Planet Venus

VLADISLAV P. VOLKOV
V.I. Vernadskiy Institute of Geochemistry and Analytical Chemistry

ABSTRACT

A host of interesting problems related to the probability of a global process of chemical interaction of the Venusian atmosphere with that planet's surface material has emerged in the wake of flights by the Soviet space probes, "Venera-4, -5, -6, and -7" (1967-70). It was disclosed during these flights that the temperature of Venus' surface attains 750 K, pressure is approximately 90 atm., and CO_2 constitutes 97% of the atmosphere. We shall explore several of these issues which were discussed in the pioneering works of Mueller (1963, 1969) and Lewis (1968, 1970):

- Is Venus' troposphere in a state of chemical equilibrium?
- Can we assume that the chemical composition of the troposphere is buffered by the minerals of surface rock?
- What are the scales and mechanisms involved as exogenic processes take place?
- To what degree is the composition of cloud particles tied to the process of lithospheric-atmospheric interaction?

We have succeeded in resolving a number of these problems over the past 20 years. At the same time, critical issues such as the chemical constituents of the near-surface layer of Venus' atmosphere, cloud particle chemistry, and the mineralogy of iron and sulfur in surface rock obviously cannot be definitely resolved until further landing craft will have been sent to the surface of Venus.

Several research projects have been conducted in the USSR and the United States, which used physical-chemical and thermodynamic methods

218

for computing multi-component systems. These projects have helped us to understand the particularity of the natural process occurring on the surface of Venus (Lewis and Kreimendahl 1980; Barsukov *et al.* 1980; Volkov *et al.* 1986; and Zolotov 1985).

Factual material from the studies of the atmosphere and surface of Venus, gathered with the "Venera" series spacecraft and during the "Pioneer Venus" mission, can be used to compare our view of the distribution, chemical composition, and physical properties of products of lithospheric-atmospheric interaction on Venus.

MANIFESTATION OF EXOGENIC PROCESSES USING PHOTOGEOLOGICAL DATA

The following conclusions on the nature and scope of exogenic processes were made after the probes "Venera-15" and "Venera-16" finished mapping Venus:

- The present surface relief of Venus was formed as a result of the combined processes of crater formation, volcanism, and tectonic activity;

- The rate of renewal of Venus' relief is estimated to take a million years for the first several centimeters (in the last three billion years), as compared to hundreds of meters on Earth and the first several meters on Mars (Nikolaeva *et al.* 1986);

- There is no evidence of exogenic processes on a global scale, such as lunar regolith;

- There are no traces of fluvial or aeolian processes having occurred on a scale that matches the resolution of the radar images (one kilometer).

At the same time, microscale exogenic processes have been quite clearly manifested. TV-panoramas from "Venera-9" and "Venera-10" recorded three types of processes: the formation of cracks; degradation with the emergence of desert aeolian weathering ridges; and corrosion, akin to porous aeolian or chemical weathering. The "Venera-13" and "Venera-14" images show laminated formations which have been interpreted (Florenskiy *et al.* 1982) as aeolian-sedimentation rock. Their formation can be described as a cycle: weathering—transport—deposition— lithiphication—weathering. . .

Experiments to estimate such physical properties of surface rock on Venus as porosity, and carrying capacity such as "Venera-13" and "Venera-14" (Kemurdgian *et al.* 1983) confirmed the existence of loose, porous bedrock. Loose, porous bedrock with an estimated thickness of 10 centimeters exists at the landing site of the Soviet "Venera-13" and "Venera-14" probes. The question of their geological nature is still unanswered: Are they products of chemical weathering or aeolian activity?

There is no direct evidence of the existence of aeolian forms as yet. No global aeolian, martian-type structures were revealed in the area mapped by the radars of the "Venera-15" and"Venera-16" probes. Nor do any of the four TV-panoramas show aeolian forms.

Experimental simulation (Greely *et al.* 1984) demonstrated that in an atmosphere of CO_2, with a pressure of approximately 100 atm. and wind speed of up to 3 m/sec^{-1}, signs of rippling occur when saltation of particles of up to 75 μm in diameter takes place. Theoretical estimates of the threshold rates of the separation of particles of varying dimensions produced similar results. Dust fraction, transported as suspension, will have a diameter of < 30 μm.

Scrupulous investigation of the TV-panoramas, incorporating data from measurements of the optical properties of the near-surface atmosphere, have demonstrated that the formation of dust clouds from the aerodynamic landing of Soviet probes is a reality. It is considered that the nature of particle behavior during wind activity on Venus is similar to the sorting of material at the bottom of the ocean at a depth of about 1000 meters.

Let us sum up the information on exogenic processes that was generated by research on the morphology and properties of the surface of Venus:

• The rate of exogenic processing of the Venusian surface relief is extremely low; the morphology of the ancient (0.5 to one billion years) relief has been excellently preserved;

• physical weathering (the equivalent of terrestrial, geological processes) has not been found: there is no aqueous water, living matter, or climatic contrasts;

• Regolith-like forms of relief are not developed;

• Aeolian activity on present-day Venus does not lead to the formation of global forms which can be differentiated on radar maps;

• The television images show traces of chemical weathering in the form of rock corrosion and degradation.

The findings from X-ray-fluorescent analyses on the Soviet "Venera-13" and "Venera-14" and "Vega-2" probes and K, U, and Th determinations on the Soviet "Venera-8, -9, and -10," and "Vega-1, and -2" probes (Surkov 1985) have given us information regarding the chemical nature of the surface rock. It is merely important for this paper to note that all of this rock belongs to the basalt group and contains almost 10 times more sulfur than their terrestrial equivalents.

THE CHEMICAL COMPOSITION AND A CHEMICAL MODEL OF THE TROPOSPHERE OF VENUS

It became clear, following the flight by the Soviet "Venera-4" probe in 1967, that CO_2 accounts for 97% of the Venusian troposphere, N_2 is approximately 3%, and the remaining constituents account for approximately 0.1% (by volume). Unfortunately, we lack instrumental data on chemical composition at elevations below 20 kilometers. This creates considerable difficulty as we attempt to understand the chemical processes at the boundary between the atmosphere and the surface.

The troposphere of Venus can be seen as a homogenous, well mixed, gaseous envelope for the major constituents (CO_2 and N_2) and the inert gases. It is clear that complex relationships exist between the physical (turbulent mixing, and horizontal and vertical planetary circulation of gas masses) and chemical (condensation and vaporization of cloud particles, gas-phase reactions, and gas-mineral types of interaction) processes in the atmosphere which lead to the existence of vertical and horizontal gradients of microconstituent concentrations (H_2O, SO_2 and CO; see Figure 1).

Venus' high surface temperature can be regarded as a factor which enhances the chemical interaction of the atmosphere with surface rock and, as a consequence, yields a dependency of the atmosphere's composition on heterogenic chemical reactions at the atmosphere-surface boundary.

Mueller (1964) proposed 25 years ago that three zones may exist in the vertical profile of Venus' atmosphere, depending on the predominance of varying types of chemical processes:

- The zone of thermochemical reactions in which the composition of the atmosphere is buffered by surface rock minerals;
- The zone of "frozen" chemical equilibrium, where the composition of gases corresponds to their equilibrium ratios in the near-surface layer of the troposphere;
- The zone of photochemical reactions in the upper atmosphere.

According to this model, chemical reactions at the planet's surface take place amid a constant influx of reactive matter from the crust reservoir, as geological and tectonic activity also occur. Using the principle of global chemical quasiequilibrium in the atmosphere-crust system, we can apply thermodynamic computations to estimate the equilibrium concentrations of atmospheric gases that are not accessible to direct measurements.

Lewis (1970) obtained more complete data on calculations of the chemical composition of the near-surface atmosphere; he took into account the results of the atmospheric analyses performed by the Soviet "Venera-4, -5 and -6" probes. Unlike Mueller (1964), he only considered chemical equilibrium at the atmosphere-surface boundary (Table 1). Both of the
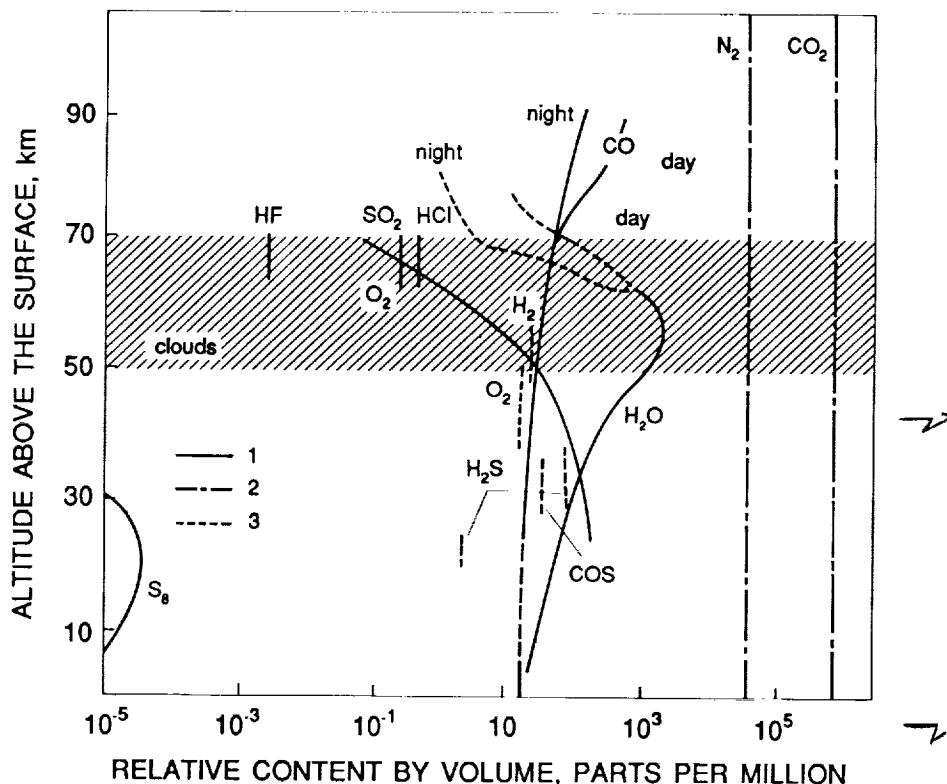
FIGURE 1   A schematic vertical cross-section of the troposphere of Venus. It shows the distribution of macro- and microconstituents based on data from measurements performed by the "Venera" series and the "Pioneer Venus" probes. 1: microconstituents; 2: macroconstituents; 3: data requiring refinement.

above models used the existence of chemical equilibrium throughout the troposphere, to its upper cloud boundary.

The literature has been discussing Urey's (1951) hypothesis for quite some time. He proposed "Wollastonite" equilibrium as a mechanism for buffering P_{CO_2} in the global, equilibrium atmosphere-crust system (Mueller 1963; Vinogradov and Volkov 1971; Lewis and Kreimendahl 1980):

$$CaCO_3 + SiO_2 \Leftrightarrow CaSiO_3 + CO_2$$

calcite quartz Wollastonite
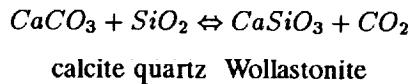
The thermodynamic calculations performed in these studies demonstrated that the mineral association of calcite-quartz-Vollastonite on the surface of Venus can buffer P_{CO_2} ($\sim$ 90 bar) at a temperature of 742 K. This is virtually

TABLE 1 Chemical Models of Venus' Troposphere

Authors	Quasi-Zone of Chemical Equilibrium	Zone of "Frozen" Chemical Equilibrium	Initial Data
R. Mueller 1963, 1969	Troposphere + lithosphere	Troposphere to an altitude of 80 km*	Spectroscopic measurements of CO, HCl, HF, and H_2O
J. Lewis 1970	Troposphere + surface rock	" "	Spectroscopic measurements of CO, HCl, HF; H_2O ("Venera-5 & -6" data)
Florenskiy et al. 1976	" "	Troposphere to lower cloud boundary	Chemical analysis of the atmosphere on Soviet "Venera -4 & -10" probes
Khodakovskiy et al. 1979	" "	Near-surface troposphere	Chemical analysis of the atmosphere on Soviet "Venera -11 & -12" probes & "Pioneer Venus"
Krasnopol'skiy and Parshev 1979	" "	Troposphere to an altitude of 60 km**	" "

*Upper boundary of the cloud layer
**According to Krasnopol'skiy and Parshev (1979): to the "zone of photochemical reactions"

commensurate with the surface conditions. However, interpretation of the multisystem computations has shown that carbonates are unstable. Yet, the high concentration of SO_2 in the troposphere is one of the determining factors of this process (Zolotov 1985; Volkov et al. 1986). Consequently, "Wollastonite" equilibrium can scarcely be seen as the basis for a chemical model of Venus' atmosphere.

Florenskii et al. (1976) developed the idea in 1976 that there may be chemical equilibrium in the subcloud portion of the troposphere. The lower atmosphere was divided into three zones:

- The stratosphere with an upper layer of clouds, which is the zone of photochemical processes;
- The main cloud layer zone, where photochemical (above) and thermochemical (below) processes compete;
- The portion of the troposphere below the cloud base, which is the zone where thermochemical equilibria are predominant.

This model brought us to a closer understanding of the Venusian troposphere as a complex, predominantly nonequilibrious system, even though numerical estimates of microconstituent concentrations (primarily SO_2) departed greatly from the actual values (Table 2).

TABLE 2   Chemical Composition of the Venusian Near-Surface Troposphere from Computational Data (Relative Levels of Microconstituents by Volume).

Gas	1	2	3	4	5	6
CO	$2 \cdot 10^{-4}$	$5 \cdot 10^{-5}$	$1.7 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$	$7.2 \cdot 10^{-6}$	$1.7 \cdot 10^{-5}$
H_2O	$5 \cdot 10^{-4}$	$3.2 \cdot 10^{-4}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-4}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-5}$
SO_2	$3 \cdot 10^{-7}$	$8 \cdot 10^{-6}$	$1.3 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$
H_2S	$5 \cdot 10^{-6}$	$1.2 \cdot 10^{-6}$	$5.2 \cdot 10^{-8}$	$3 \cdot 10^{-7}$	$8 \cdot 10^{-9}$	$8 \cdot 10^{-5}$
COS	$5 \cdot 10^{-5}$	$3.2 \cdot 10^{-5}$	$2.3 \cdot 10^{-5}$	$2 \cdot 10^{-5}$	$3 \cdot 10^{-6}$	$4 \cdot 10^{-5}$
S_2	$2 \cdot 10^{-8}$	$4 \cdot 10^{-8}$	$1.8 \cdot 10^{-7}$	10^{-7}	$1.3 \cdot 10^{-8}$	$2 \cdot 10^{-8} \text{-} 8 \cdot 10^{-7}$
H_2	$7 \cdot 10^{-7}$	10^{-7}	$2.4 \cdot 10^{-9}$	$2 \cdot 10^{-8}$	10^{-9}	$2.5 \cdot 10^{-5}$
O_2	$8 \cdot 10^{-26}$	10^{-24}	10^{-23}	--	10^{-23}	$1.8 \cdot 10^{-5}$

Notes: The underlined figures are initial data of measurements on space probes or ground-based facilities; 1: Mueller (1969); 2: Lewis (1970); 3: Khodakovskiy et al. (1979); 4: Krasnopol'skiy and Parshev (1979); 5: Zolotov (1985).

Column 6 tabulates data of measurements made on the "Venera" and "Pioneer Venus" probe series; no measurements were performed below an altitude of 20 kilometers (Figure 1).

Following measurements of the chemical composition of the troposphere by the Soviet "Venera-11 and -12" probes and the "Pioneer Venus" probe, the computational and experimental values of microconstituent concentrations were compared. Khodakovskii et al. (1979) and Krasnopol'skii and Parshev (1979) concurrently and independently proposed models (Table 2). These models were the first to compare gas-phase reaction rates with troposphere mixing rates. These consequences were generated:

• The troposphere is generally in nonequilibrium, with the exception of the near-surface layer with a thickness of the first kilometer, where the highest temperatures are dominant. However, the processes of heterogeneous catalysis at the atmosphere-surface boundary may favor the establishment of chemical equilibrium in relation to certain constituents;

• The chemical composition of the microconstituents in the vertical cross-section below the cloud base region of the troposphere does not vary: it corresponds to the "frozen" equilibrium at the atmosphere-surface boundary (T = 735 K; P = 90 atm).

The principal of "frozen" equilibrium was applied in order to theoretically estimate the chemical composition of cloud particles; this enables us to better understand the sulfur and chlorine cycles in the atmosphere-crust system (Volkov 1983; Volkov et al. 1986).

The lack of instrumental determinations of microconstituents in the

troposphere at altitudes below 20 kilometers prevents us from solving the critical problem of the ratio of gaseous sulfur: $H_2S + COS > SO_2$ (Lewis 1970) or $H_2S + COS < SO_2$ (Khodakovskii *et al.* 1979; see Table 2). Furthermore, gas chromatographic determination of oxygen by the Soviet "Venera-13, and -14" probes cannot be reconciled with the concurrent presence in these same samples of 80 ppm H_2S and 40 ppm COS (see Volkov and Khodakovskii 1984 for greater detail). The only original attempt to experimentally estimate the oxidation-reduction regime on Venus' surface, using a "Kontrast" detector on the Soviet "Venera-13, and -14" probes (Florenskii *et al.* 1983) pointed to the presence in the near-surface layer of the troposphere of a reducing agent (CO). However, it does not give us a clear-cut solution to the oxygen problem.

The results from estimations of the chemical composition of the troposphere and the nature of the processes occurring in its near-surface layer can be summarized in three conclusions:

(1) Chemical equilibrium in the troposphere of Venus has generally not been reached.

(2) The vertical gradients of SO_2, H_2O and CO concentrations are a function of the competition between physical and chemical processes in the troposphere.

(3) The near-surface troposphere can be seen as a layer in a state of "frozen" chemical equilibrium.

Unfortunately, we have yet to resolve the question of the oxidation-reduction regime on Venus' surface, as well as the problem of the existence of free oxygen in the troposphere.

THE MINERAL COMPOSITION OF SURFACE ROCK ON VENUS

Many investigations have attempted to estimate the possible mineral associations on the surface of Venus using chemical thermodynamic methods.

Mueller published the first such study as a component of the aforementioned chemical model of the atmosphere (Mueller 1963) and obtained the following results:

- Temperature and pressure on Venus' surface are consistent with silicate-carbonate equilibrium, and carbon is bound in the rock in $CaCO_3$ form;
  - Oxygen partial pressure is buffered by Fe-containing minerals;
  - Graphite and the native metals are not stable;
  - Nitrogen is not bound in the condensed phases;

● A number of chlorine- and fluorine-containing minerals are stable at the surface.

Lewis (1970) calculated 64 mineral equilibria in order to estimate P and T on the surface before the probes performed these measurements. One out of three proposed options for the P and T values was in satisfactory agreement with the actual values obtained a year later. Lewis yielded the following, additional forecast estimates:

● Surface rock contains H_2O molecules bound in the form of tremolite;

● sulfur is bound in the cloud layer in the form of mercury sulfides. Carbonyle-sulfide is the dominant form in which sulfur is found in the troposphere. This prediction proved only partially true: sulfur is actually the main component of cloud particles, but the latter consist primarily of H_2SO_4.

A series of studies to calculate mineral composition was conducted in 1979-83 at the V.I. Vernadskiy Institute using the computation of the phase ratios in multicomponent, gaseous systems, modeling the atmosphere/surface-rock system. The computations were based on troposphere chemical analysis data from the Soviet "Venera" series of probes, "Pioneer Venus," thermodynamic constants of about 150 phases, the chemical components of terrestrial magmatic rock, and the results of x-ray-fluorescent analysis of rock at three probe landing sites ("Venera-13," "Venera-14," and "Vega-2"). Compilation of this material can be found in Volkov *et al.* (1986). It should be stressed that three important predictive conclusions were made before the first data on the chemical composition of Venus' bedrock were obtained:

● Sulfur may be bound as sulfates ($CaSO_4$) and/or sulfides (FeS_2), and its concentration greatly exceeds known sulfur levels in terrestrial equivalents;

● Water-containing minerals are unstable;

● Carbonates are unstable;

● Magnetite Fe_3O_4 must be a widespread constituent both as primary and as altered bedrock.

These conclusions were generally confirmed, albeit with some refinement, by comparing them with X-ray-fluorescent analyses at "Venera-13," "Venera-14" and "Vega-2" landing sites, and by further, more detailed theoretical investigations (Zolotov 1985, 1989).

Sulfur at the surface probe landing sites, if we judge from the data of additional, postflight calibration investigations (Surkov *et al.* 1985), is in an anhydrite form ($CaSO_4$). Sulfur content may serve as a measure of convergence to the state of chemical equilibrium relative to SO_2 in the

atmosphere-crust system (Lewis and Prinn 1984; Volkov *et al.* 1986). It may be possible that rock with a maximum level of sulfur (1.9 mas. %, "Vega-2") were in contact with the atmosphere longer than the bedrock at the landing sites of the Soviet "Venera-13" and "Venera-14" probes.

In his 1985 study, Zolotov conducted thermodynamic assessments of carbonate stability depending on the concentration of SO_2, since a reaction such as:

$$CaCO_3 + 1.5SO_2 \Leftrightarrow CaSO_4 + CO_2 + 0.25S_2$$

takes place in Venus surface conditions free of kinetic constraints. As it turned out, the presence of SO_2 in quantities exceeding 1 ppm excludes the existence of calcite and dolomite. However, magnesite ($MgCO_3$), as a product of the alteration of pure forsterite, $MgSiO_4$ (Fo_{100}), may be stable at altitudes of 1.5 to eight kilometers.

Zolotov demonstrated in this same study (1985) that hematite (Fe_2O_3) may even be stable at an altitude of more than 1.5 kilometers (Figure 2), in addition to magnetite (Fe_3O_4) (the product of water vapor-driven oxidation of Fe-containing silicates, CO_2 and SO_2). Hematite stability is apparently confirmed by the results obtained from interpreting the surface color on the TV images from "Venera-13 and -14" (Shkuratov *et al.* 1987).

Nevertheless, in their 1980 study Lewis and Kreimendahl retain the conclusion regarding calcite ($CaCO_3$) and wüstite (FeO) stability, while allowing for the prevalence of H_2S and COS over SO_2 in conditions of total chemical equilibrium at the surface-atmosphere boundary. They come to the same logical conclusion that in this case, the surface rock of Venus' crust is characterized by an extremely low degree of oxidation (Fe+3/Fe+2 at one to two orders lower than the terrestrial value). Strictly speaking, the ultimate solution to the problem of the oxidation of Venus' crust has not been found due to the lack of instrumental data.

In 1975, Walker (1975) drew attention to the possible dependence of the mineral constituents of Venus' surface on the hypsometric level. The pressure ($\approx$ 65 atm.) and temperature ($\approx$ 100 K) gradients are actually so great that they may alter the composition of the phases of rock during their exogenous cycle, that is, under the influence of aeolian transport. If we take into account the fact that our knowledge of Venus' mineralogy does not go beyond the framework of theoretical forecasting, the factor of "hypsometric control" must still be considered hypothetical.

Let us summarize the theoretical investigations of the chemical interaction of Venus' rock with its atmosphere.

● Alteration of the composition of Venus' basalts during interaction with the atmosphere is highly probable;
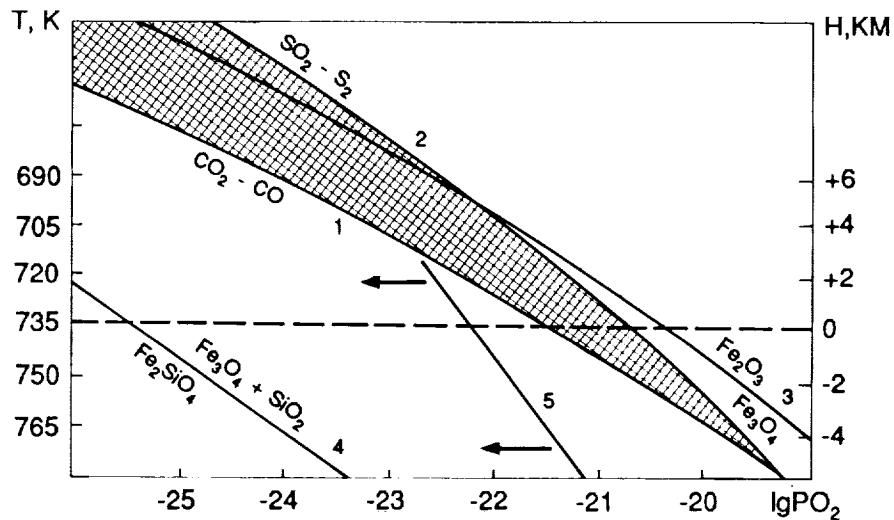
FIGURE 2   Estimates of the oxidation-reduction regime in the troposphere and on the surface of Venus from data produced by measurements (1,2,5) and computations (3,4). (Zolotov 1985). 1. $CO_2 = CO + \frac{1}{2}O_2$ ($C_{CO_2} = 96.5\%$; $C_{CO} = 20$ ppm). 2. $SO_2 = \frac{1}{2}S_2 + O_2$ ($C_{SO2} = 130 \div 185$ ppm; $C_{S2} = 20$ ppb). 3. $3Fe_2O_3 = 2Fe_3O_4 + \frac{1}{2}O_2$ (buffer HM). 4. $2Fe_3O_4 + 3SiO_2 = 3Fe_2SiO_4 + O_2$. 5. "Kontrast" detector (Florenskiy *et al.* 1983).

- Apparently, the primary outcome stemming from this interaction will be the sink of sulfur in the crust as anhydrite ($CaSO_4$) and/or iron sulfides (FeS and FeS_2);

- The existence of carbonates (besides $MgCO_3$), free carbon and nitrogen compounds on the surface of Venus is thermodynamically prohibited;

- The lack of complete factual data prevents our making a clear-cut conclusion as to the stability of water-bearing minerals and the degree of oxidation of the Venusian crust.

THE CYCLES OF VOLATILE COMPONENTS

Interpretation of data on Venus' atmospheric chemistry, and in particular, consideration of the photochemical processes in the stratosphere (Krasnopolskii 1982; Yung and De More 1982) demonstrated that nitrogen and carbon cycles are completed in the atmosphere. The H_2O cycle poses more problems, since we are not yet clear on the vertical profile of H_2O concentrations in the near-surface atmosphere.

Clearly, sulfur is the only volatile element on Venus which, in the
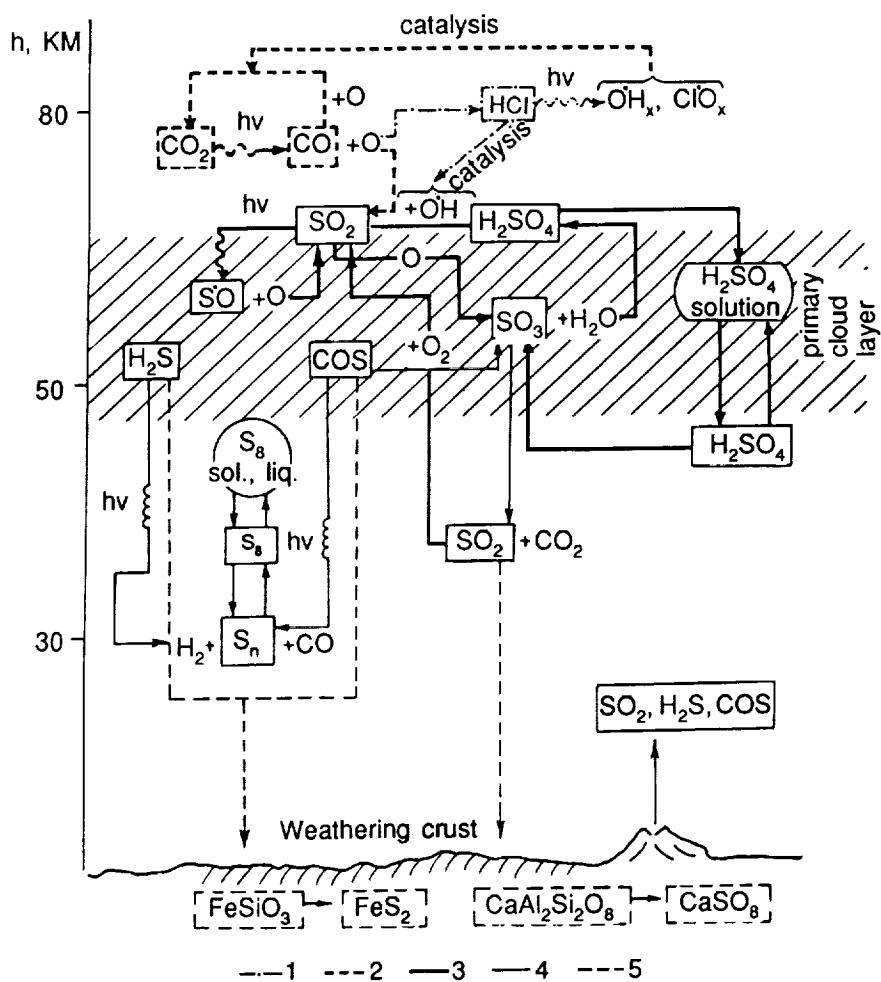
FIGURE 3 Diagram of the cycles of CO_2, sulfur and chlorine in the Venusian atmosphere 1: chlorine cycle; 2: CO_2 cycle; 3: rapid sulfur cycle; 4: slow sulfur cycle; 5: sulfur flux into the crust.

contemporary geological epoch, participates in the cyclical mass exchange between the atmosphere and the crust. Sulfur's behavior as a constituent of the cloud layer essentially determines its structure and dynamics. Three cycles (Figure 3) have been discerned, depending on the rates at which these processes unfold (Lewis and Prinn 1984).

The rapid cycle takes place in the stratosphere and the clouds and sets the stage for the photochemical emergence and thermal destruction of sulfuric acid aerosols. The residence time for the SO_2 molecule is estimated

TABLE 3 Mineral Composition of Venusian Surface Rock Based on Theoretical Assessments
(Secondary Minerals)

1	2	3
Carbon in $CaCO_3$ $C_{(graphite)}$ is unstable	Carbonates Ca and Mg	Carbonates are unstable $MgCO_3$ (?)
H_2O in amphiboles and micas	H_2O in amphiboles	Amphiboles (?)
Fe_2O_3, Fe_3O_4	FeO, Fe_3O_4 where $X_{COS} + X_{H2S} > X_{SO2}$	Fe_3O_4, Fe_2O_3 (?) where $X_{COS} + X_{H2S} < X_{SO2}$
Sulfur in sulfides of Fe and anhydrite $(CaSO_4)$		Predominantly $CaSO_4$, sulfides of Fe are stable

Nitrogen-bearing minerals are unstable
Chlorine- and fluoride-bearing minerals are stable (fluorite/apatite?)

1. Mueller 1963, 1969
2. Lewis 1970; Lewis and Kreimendahl 1980
3. Khodakovskiy et al. 1978; Volkov 1983; Zolotov 1989.

to be from several hours to several years. Two alternative scenarios are proposed in Table 4. It is difficult to select between the two because of the lack of experimental data on the rates of certain photochemical reactions.

The slow atmospheric cycle is most likely a function of photochemical and thermodynamic reactions in the lower atmosphere which lead to the existence of reduced forms: H_2S and COS and elementary sulfur. Apparently, the stratosphere is the region of H_2S and COS flux: they either photodissociate there or are oxidized by molecular oxygen to SO_3. The time span of sulfur molecules in the cycle is estimated to be several dozen years (Lewis and Prinn 1984).

The mass exchange between Venus' crust and its atmosphere is carried out in a "geological" sulfur cycle. The source of sulfur is crust matter which produces sulfur-bearing gases through both volcanism and the interaction of minerals with atmospheric gases, such as FeS_2 with CO_2, H_2O, and CO.

These gases repeatedly participate in photo- and thermodynamic processes in the atmosphere. The rapid atmospheric cycle brings about the long-term existence of a cloud cover made up of condensed H_2SO_4 particles. The competition of photo- and thermochemical reactions in the slow cycle apparently support the existence of SO_2 as the dominant form of sulfur in the atmosphere. An excess of SO_2 compared with its equilibrium concentration in the atmosphere-crust system create an SO_2 flux in the form of sulfates in surface rock.

Two factors determine the scales and rates of flux:

TABLE 4 Sulfur Cycles on Venus

Cycle		
Timeframe	I. Fast cycle (stratosphere and cloud layer)	
	(Winick and Stewart 1980)	(Krasnopol'skiy 1982)
	$CO_2 + hv \rightarrow CO + O$	$SO_2 + hv \rightarrow SO + O$
< 10 yrs.	$SO_2 + O + M \rightarrow SO_3 + M$	$SO + O + M \rightarrow SO_2 + M$
	(OH, HO_2 are catalysers)	$SO_2 + O + M \rightarrow SO_3 + M$
	$SO_3 + H_2O \rightarrow H_2SO_4$ (Sol)	$SO_3 + H_2O \rightarrow H_2SO_4$ (Sol)

	II. Slow cycle (lower atmosphere and cloud layer)	
	$SO_3 + 4CO \rightarrow COS + 3SO_2$	$SO_3 + H_2 + 3CO \rightarrow H_2S + 3CO_2$
$\gtrsim$ 10 yrs.	$COS + hv \rightarrow CO + S$	$H_2S + hv \rightarrow \overset{.}{H}S + \overset{.}{H}$
	$COS + 1.5O_2 \rightarrow SO_3 + CO$	$H_2S + 1.5O_2 \rightarrow SO_3 + H_2$

	III. Geological Cycle	
	$CaSiO_3$, $CaAl_2Si_2O_8 + SO_2 \rightarrow CaSO_4$	
> 10^6 yrs.	$FeSiO_3$, $Fe_3O_4 + COS$ (H_2S) $\rightarrow$ FeS(FeS$_2$)	

The time frame for a cycle to run its course depends on:
1) Mineral $\leftrightarrows$ gas reaction rates on the planet's surface
2) Length of time during which mineral particles are in contact with the atmosphere
   such as surface relief renewal rates

• The rate of heterogeneous mineral = gas reactions on the planet's surface;

• The residence span in which mineral particle are in contact with the atmosphere, for example, the surface relief renewal rate.

The completing of the "geological" cycle probably occurs as the altered surface rock (rich $CaSO_4$) is re-melted in the deep regions of the crust. Attenuated volcanic and tectonic activity on Venus ultimately reduces the thickness of the cloud layer because sulfur is fixed in the crust and depleted in the atmospheric reservoir.

GENERAL CONCLUSIONS

We can cite at least four firmly established facts that determine the existence of the chemical interaction of Venus' atmosphere with its surface rock. These are:

• Loosely porous rock on the planet's surface is developed; massive rock display traces of corrosion and degradation;

• There is no global regolith; aeolian transport on a limited scale is supported by weak winds in the near-surface atmosphere;

• The troposphere contains reactive gases (microconstituents): SO_2, H_2O, CO, and others;- Venus' basalts contain one to 1.5 more orders of sulfur than their terrestrial equivalents.

We can make the following conclusions based on our interpretation of the entire set of observational data:

(1) The processes of lithospheric-atmospheric interaction substantially alter primary basalts and subject them to chemical weathering. The scale of this process cannot be estimated;

(2) The troposphere is generally not in a state of chemical equilibrium with the surface rock, and the chemical composition of the near-surface layer may correspond to a "frozen" equilibrium which is buffered by the minerals.

(3) Sulfur is in a state of cyclical mass exchange between the atmosphere and the crust.

(4) Nitrogen and oxygen in the crust's rock do not form stable phases. Their cycles become completed in the atmosphere.

REFERENCES

Barsukov, V.L., V.P. Volkov, and I.L. Khodakovsky. 1982. The crust of Venus: theoretical models of chemical and mineral composition. J. Geophys. Res. 87(suppl. Part 1):A3-A9.

Florenskiy, K.P., A.T. Bazilevskiy, and A.S. Selivanov. 1982. Panoramas of the landing sites of "Venera-13" and "Venera-14" (a preliminary analysis). Astron. Vestnik. 16(3):131-138.

Florenskiy, K.P., V.P. Volkov., and O.V. Nikolaeva. 1976. Towards a geochemical model of the troposphere of Venus. Geokhimiya 8: 1135-1150.

Florenskiy, K.P., O.V. Nikolaeva, V.P. Volkov, et al. 1983. On the oxidation-reduction conditions on the surface of Venus, based on data from the "Kontrast" geochemical detector deployed on the "Venera-13" and "Venera-14" probes. Kosmich. issled. 21(3):351-354.

Greely, R., J. Iverson, R. Leach, et al. Windblown sand on Venus: preliminary results of laboratory simulations. Icarus 57(1):112-124.

Kemurdjian, A.L., P.N. Brodskiy, V.V. Gromov, et al. 1983. Preliminary results in determining the physical and mechanical properties of bedrock by the Soviet automated "Venera-13" and "Venera-14" probes. Kosmich. issled. 21(3):323-330.

Khodakovskiy, I.L., V.P. Volkov, Yu.I. Sidorov, et al. 1978. The mineralogical composition of bedrock, and the processes of hydration and oxidation of the outer envelope of the planet Venus (a preliminary forecast). Geokhimiya 12:1821-1835.

Khodakovskiy, I.L., V.P. Volkov, Yu.I. Sidorov, et al. 1979. A geochemical model of the troposphere and crust of the planet Venus, based on new data. Geokhimiya 12:1748-1758.

Krasnopol'skiy, V.A. 1982. Photochemistry of the atmospheres of Mars and Venus. Nauka, Moscow.

Krasnopol'skiy, V.A., and V.A. Parshev. 1979. On the chemical composition of the Venusian troposphere and cloud layer, based on measurements made by "Venera-11 and -12" and "Pioneer Venus." Kosmich. issled. 17(5):763-771.

Lewis, J.S. 1968. An estimate of the surface conditions of Venus. Icarus 8(3):434-457.

Lewis, J.S. 1970. Venus: atmospheric and lithospheric composition. Earth Planet. Sci. Lett. 10(1):73-80.

Lewis, J.S., and F.A. Kreimendahl. 1980. Oxidation state of the atmosphere and crust of Venus from Pioneer-Venus results. Icarus 42(3):330-337.

Lewis, J.S., and R.G. Prinn. 1984. Planets and their atmospheres. Origin and evolution. Academic Press, Orlandi.

Mueller, R.F. 1963. Chemistry and petrology of Venus: preliminary deductions. Science 141(3585):1046-1047.

Mueller, R.F. Planetary probe: origin of the atmosphere of Venus. Science 163(3873):1322-1324.

Nikolaeva, O.V., L.B. Ronka, and A.T. Bazilevskiy. 1986. Circular formations on the plains of Venus as evidence of its geological history. Geokhimiya 5:579-589.

Surkov, Yu.A. 1985. Cosmic investigations of the planets and their satellites. Nauka, Moscow.

Surkov, Yu.A., L.P. Moskaleva, O.P. Shcheglov, et al. 1985. The method, equipment and results from determinations of the elementary composition of the Venusian bedrock made by the "Vega-2" space probe. Astron. Vestnik. 19(4):275-288.

Shkuratov, Yu.G., M.A. Kreslavskiy, and O.V. Nikolaeva. 1987. A diagram of the albedo-color of a parcel on the surface of Venus and its interpretation. Astron. vestnik. 21(2):152-154.

Urey, H.C. 1951. The origin and development of the Earth and other terrestrial planets. Geochim. et Cosmochim. Acta. 1(2):209-277.

Vinogradov, A.P., and V.P. Volkov. 1971. On Vollastonite equilibrium as a mechanism determining the composition of Venus' atmosphere. Geochimiya 7:755-759.

Volkov, V.P. 1983. The Chemistry of the atmosphere and surface of Venus. Nauka, Moscow.

Volkov, V.P., and I.L. Khodakovskiy. 1984. Physical-Chemical modeling of the mineral composition of the surface rock of Venus. Pages 32-37. In: Geokhimiya i kosmochimiya. Papers at the 27th session of the Int'l. Geol. Congr. Section C. 11. Nauka, Moscow.

Volkov, V.P., I.L. Khodakovsky, and M.Yu. Zolotov. 1986. Lithospheric-atmospheric interaction on Venus. Chemistry and physics of terrestrial planets. Adv. Phys. Geochim. 6:136-190.

Walker, J.C. 1975. Evolution of the atmosphere of Venus. J. Atm. Sci. 32(6):1248-1256.

Winick, J.R., and A.I.F. Stewart. 1980. Photochemistry of SO2 in Venus' upper cloud layers. J. Geophys. Res. 85(A 13):7849-7860.

Yung, Y.L., and W.B. de More. 1982. Photochemistry of the stratosphere of Venus: implications for atmospheric evolution. Icarus 51(2):199- 247.

Zolotov, M.Yu. 1989. Chemical weathering on Venus and Mars: Similarities and Differences. Pages 71-79. In: Cosmochemistry and comparitive planetology. Proceedings of Soviet scientists for the 28th session of the Int'l. Geol. Congr. Nauka, Moscow.

Zolotov, M.Yu. 1985. Sulfur-containing gases in the Venus' atmosphere and stability of carbonates. Pages 942-944. In: Abstr. XVI Lunar and Planet. Sci. Conf. Houston, part 2.

Runaway Greenhouse Atmospheres:
Applications to Earth and Venus

JAMES F. KASTING
The Pennsylvania State University

ABSTRACT

Runaway greenhouse atmospheres are discussed from a theoretical standpoint and with respect to various practical situations in which they might occur. The critical solar flux required to trigger a runaway greenhouse is at least 1.4 times the solar flux at Earth's orbit (S_o). Rapid water loss may occur, however, at as little as 1.1 S_o, from a type of atmosphere termed a "moist greenhouse." The moist greenhouse model provides the best explanation for loss of water from Venus, if Venus did indeed start out with a large amount of water. The present enrichment in the D/H ratio on Venus provides no unambiguous answer as to whether or not it did. A runaway greenhouse (or "steam") atmosphere may have been present on the Earth during much of the accretion process. Evidence from neon isotopes supports this hypothesis and provides some indication for how long a steam atmosphere may have lasted. Finally, the theory of runaway and moist greenhouse atmospheres can be used to estimate the position of the inner edge of the continuously habitable zone around the Sun. Current models place this limit at about 0.95 AU, in agreement with earlier predictions.

INTRODUCTION

The topic of runaway greenhouse atmospheres has received renewed attention over the past several years for three different reasons. The first concerns the history of water on Venus. Although there is still no concensus as to whether Venus had much water to begin with (Grinspoon 1987;

234

Grinspoon and Lewis 1988), some recent theories of accretion (Wetherill 1985) predict extensive radial mixing of planetesimals within the inner solar system. If this idea is correct, then Venus must initially have received a substantial fraction of Earth's water endowment. This water is obviously not present on Venus today. The mixing ratio of water vapor in the lower atmosphere of Venus is approximately 10^{-4}; thus, the total amount of water present is only about 10^{-5} times the amount in Earth's oceans. The runaway greenhouse theory provides a convenient explanation for how the rest of Venus' water might have been lost.

A runaway greenhouse atmosphere may also have been present on Earth during at least part of the accretionary period. Matsui and Abe (1986a,b) and Zahnle *et al.* (1988) have shown that an impact-induced steam atmosphere could have raised the Earth's surface temperature to 1500 K, near the solidus temperature for typical silicate rocks. This implies the existence of a global magma ocean of unspecified depth. Although the continuous existence of such a steam atmosphere has been questioned (Stevenson 1987), an analysis of terrestrial neon isotope data (Kasting 1990) strongly supports the notion that such an atmosphere existed during some portion of the accretionary period.

A third reason for interest in runaway greenhouse atmospheres concerns their implications for the existence of habitable planets around other stars. Any planet that loses its water as a consequence of a runaway greenhouse effect is not likely to be able to support life as we know it. Thus, the idea that runaway greenhouses are possible sets limits on the width of the continuously habitable zone (CHZ) around our Sun and around other main sequence stars (Hart 1978, 1979; Kasting and Toon 1989). One of the most important reasons for studying runaway greenhouses is to try to estimate the chances of finding another Earth-like planet elsewhere in our galaxy.

Here, the theory of runaway greenhouse atmospheres is briefly reviewed, and the consequences for the three problems mentioned above are discussed.

RUNAWAY GREENHOUSE ATMOSPHERES

The concept of the runaway greenhouse atmosphere was introduced by Hoyle (1955) and has been further developed by Sagan (1960), Gold (1964), Dayhoff *et al.* (1967), Ingersoll (1969), Rasool and DeBergh (1970), Pollack (1971), Goody and Walker (1972), Walker (1975), Watson *et al.* (1984), Matsui and Abe (1986a,b), Kasting (1988), and Abe and Matsui (1988). The basic idea, as explained by Ingersoll (1969), is that there exists a critical value of the solar flux incident at the top of a planet's atmosphere above which liquid water cannot exist at the planet's surface. Intuitively,

one expects this to be the case. If Earth were by some means to be pushed closer and closer to the Sun, one would anticipate that at some point the oceans would be vaporized and the planet would be enveloped in a dense, steam atmosphere. The amount of water in Earth's oceans, 1.4×10^{24} g, is such that the surface pressure of this atmosphere would be about 270 bar. For comparison, this is ~50 bar greater than the pressure at the critical point of water (647.1 K, 220.6 bar).

It should be noted that the term "runaway greenhouse" has also been used to describe the positive feedback between the surface temperature of a planet and the amount of water vapor in its atmosphere. An increase in surface temperature causes an increase in the vapor pressure of water which, in turn, leads to an enhanced greenhouse effect and a further increase in surface temperature. Although this type of positive feedback certainly exists, there is no reason to believe that Earth's present climate system is unstable. Surface temperature is simply a monotonically increasing function of the incident solar flux. Thus, the phrase "runaway greenhouse" is best reserved to describe a situation in which a planet's surface is entirely devoid of liquid water.

The single most important characteristic of a runaway greenhouse atmosphere is the critical solar flux required to trigger it. Only recently have detailed estimates of this energy threshold been made (Kasting 1988; Abe and Matsui 1988). Even these estimates, which were obtained using elaborate radiative-convective climate models, cannot be considered reliable. The greatest uncertainty in performing such a calculation is the effect of clouds on the planetary radiation budget. Kasting (1988) has derived results for a fully saturated, cloud-free atmosphere. (Actually, clouds were crudely parameterized in this model by assuming an enhanced surface albedo.) The critical solar flux in his model is 1.4 S_o, where S_o is the present solar flux at Earth's orbit (1360 W m^{-2}). An Earth-like planet with Earth-like oceans was assumed. Abe and Matsui (1988) have performed a similar calculation for a case in which part of the energy required to trigger the runaway greenhouse was derived from infalling planetesimals. (Their study was specifically directed at the problem of atmospheric evolution during the accretion period.) The amount of accretionary heating required to trigger runaway conditions in their model, 150 W m^{-2}, is the same as the value derived by Kasting (1988) for an analogous simulation. (See his Figure 13.) Thus, the two existing detailed calculations of the energy threshold of the runaway greenhouse are in good agreement.

Although clouds cannot reliably be parameterized in such an atmosphere, their qualitative effect on the planetary radiation balance is not difficult to determine (Kasting 1988). An atmosphere rich in water vapor would probably exhibit at least as much fractional cloud cover as the current

Earth and possibly much more. Although clouds affect both the incoming solar and outgoing infrared radiation, the solar effect should dominate because a water vapor atmosphere would already be optically thick throughout the infrared. Thus, the main effect of increased cloud cover should be to reflect a greater proportion of the incident solar radiation, thereby diminishing the amount of energy available to sustain a steam atmosphere. It follows that the energy threshold for a runaway greenhouse is almost certainly higher than 1.4 S_o. How much higher is uncertain, but values as high as 5 S_o are within the realm of possibility (Kasting 1988, Figure 8c).

These rather speculative theoretical models should be weighed against an observational fact: our neighboring planet Venus has very little water in its atmosphere, less than 200 ppm by volume (Moroz 1983; von Zahn *et al.* 1983). As discussed further below, it is not clear whether this lack of water is innate to the planet or whether it is the result of an evolutionary process. One possibility, however, is that Venus was initially water-rich, and that it lost its water by photodissociation in the upper atmosphere followed by escape of hydrogen. (See references in opening paragraph.) If this theory is correct, then Venus must have experienced either a runaway greenhouse or a phenomenon akin to a runaway greenhouse at some time in the past. The solar flux at Venus' orbit is currently 1.91 S_o. Based on stellar evolution models, the Sun's output was some 25-30% lower (1.34-1.43 S_o) early in solar system history (Gough 1981). This implies that the critical threshold for losing water from a planet is no higher than 1.9 S_o and may well be considerably lower.

If this was all there was to the problem, one could reliably conclude that the energy threshold for the runaway greenhouse was between 1.4 S_o and 1.9 S_o. However, it has recently been demonstrated that there are other ways for a planet to lose water rapidly besides the runaway greenhouse. An alternative possibility, proposed by Kasting *et al.* (1984) and Kasting (1988) is that Venus experienced a so-called "moist greenhouse," in which the planet lost its water while at the same time maintaining liquid oceans at its surface. This turns out to be slightly favored from a theoretical standpoint; it also requires a significantly lower energy input than does the runaway greenhouse model. This alternative theory is described briefly below.

MOIST GREENHOUSE ATMOSPHERES

The concept of the moist greenhouse atmosphere stems from the analysis of moist adiabats by Ingersoll (1969). Ingersoll showed that the vertical distribution of water vapor in an atmosphere should be strongly correlated with its mass-mixing ratio $c_o(H_2O)$ near the surface. When water vapor is a minor constituent of the lower atmosphere [$c_o(H_2O) < 0.1$], its concentration declines rapidly with altitude throughout the convective

region as a consequence of condensation and rainout. This is the situation in Earth's atmosphere today, where $C_o(H_2O)$ declines from roughly 0.01 near the surface to about 3×10^{-6} in the lower stratosphere. When water vapor is a major constituent [$c_o(H_2O) > 0.1$], however, its behavior is quite different. The amount of latent heat released by condensation becomes so large that the temperature decreases very slowly with altitude, and the water vapor mixing ratio remains nearly constant. This allows water vapor to remain a major constituent even at high altitudes which, in turn, allows it to be effectively photodissociated and the hydrogen lost to space. (The important constraint here is that water vapor remain abundant above the cold trap, i.e. the maximum height at which it can condense. When this criterion is satisfied, hydrogen should escape at close to the diffusion-limited rate (Hunten 1973), as long as sufficient solar extreme ultraviolet (EUV) energy is available to power the escape.)

Kasting *et al.* (1984) and Kasting (1988) have applied the moist green-house model to the problem of water loss from an Earth-like planet. The most recent results (Kasting 1988) indicate that hydrogen escape becomes very rapid (i.e. $c_o(H_2O)$ becomes greater than 0.1) for incident solar fluxes exceeding 1.1 S_o. As before, this calculation was performed for a fully saturated, cloud-free atmosphere, so the actual value of the solar flux at which water loss becomes efficient is probably greater than this value. The calculation does demonstrate, however, that Venus could have lost most of its water without ever experiencing a true runaway greenhouse. Indeed, the solar flux at Venus's orbit early in solar system history (1.34 − 1.43 S_o) is so close to the minimum value required for runaway (1.4 S_o) that it seems likely that clouds would have tipped the balance in favor of the moist greenhouse scenario. Thus, if Venus were originally endowed with as much water as Earth, it may at one time have had oceans at its surface.

With the concepts of runaway and moist greenhouses in mind, let us now return to the three topics mentioned in the introduction.

LOSS OF WATER FROM VENUS

The real issue concerning Venus is not so much whether it could have lost its water but, rather, whether it had any appreciable amount of water initially. It is now well accepted that the D/H ratio on Venus is very high: approximately 100 times the terrestrial value. The original interpretation of this observation (Donahue *et al.* 1982) was that Venus was once wet. If Venus and Earth started out with similar D/H ratios, which seems even more likely now in view of the terrestrial D/H ratio observed in the tail of comet Halley (Eberhardt *et al.* 1987), this measurement implies that Venus once had at least 100 times as much water as it does now. The current water abundance on Venus, assuming a lower atmosphere mixing

ratio of 100 ppmv, is only 0.0014% of a terrestrial ocean. Thus, this minimal interpretation requires only that Venus start out with about 0.1% of Earth's water endowment. Even advocates of a dry early Venus would probably not dispute such a claim, given the potential for radial mixing of planetesimals during the accretion process (Wetherill 1985). If, however, some deuterium was lost along with the escaping hydrogen (which seems, indeed, to be unavoidable), the amount of water that was lost could be orders of magnitude greater. Consequently, proponents of a wet origin for Venus (Donahue *et al.* 1982; Kasting and Pollack 1983) have suggested that Venus may well have started out with an Earth-like water endowment.

The wet-Venus model has been challenged by Grinspoon (1987) and Grinspoon and Lewis (1988), who point out that the present D/H enrichment on Venus could be explained if the water abundance in Venus' atmosphere were in steady state. Loss of water by photodissociation and hydrogen escape could be balanced by a continued influx of water from comets. Grinspoon and Lewis's steady-state model has, in turn, been criticized (Donahue, private communication, 1988) on the grounds that they underestimated the amount of water vapor in Venus' lower atmosphere. The time constant for evolution of the D/H ratio in Venus' atmosphere can be expressed as (Grinspoon 1987)

$$\tau \sim R/(f\phi) \tag{1}$$

where R is the vertical column abundance of water vapor in the atmosphere, ϕ is the hydrogen escape rate, and f is the D/H fractionation factor (i.e. the relative efficiency of D escape compared to H escape). Best estimates for the values of ϕ and f, based on a weighted average of the hydrogen loss rates from charge exchange with H^+ and from momentum transfer with hot O atoms, are 2×10^7 H atoms cm^{-2} s^{-1} and 0.013, respectively (Hunten *et al.* 1989). This estimate draws upon a reanalysis of the charge exchange process by Krasnopolsky (1985). Grinspoon (1987) assumed that the Venus lower atmosphere contained only 20 ppmv of water vapor; this gives $R \approx 6 \times 10^{22}$ H atoms cm^{-2} and $\tau \approx 7 \times 10^9$ years. Even this value is somewhat longer than the age of the solar system, indicating that the steady-state model is marginally capable of explaining the observations. Small increases in the value of either ϕ or f could eliminate the time scale problem. If the H_2O mixing ratio is actually closer to 200 ppm, however, then $\tau \approx 7 \times 10^{10}$ years, and the steady-state model is in serious trouble.

Resolution of this question requires, at a bare minimum, that the present controversy regarding the H_2O content of the Venus lower atmosphere be resolved. Our present understanding of Venus' water inventory is further muddled by the fact that the H_2O mixing ratio apparently varies with altitude from about 20 ppmv near the surface to 200 ppmv just below

the clouds (von Zahn *et al.* 1983). Until this variation is explained theoretically, little confidence can be given to any of the measurements, and the H_2O abundance on Venus will remain an enigma.

Setting aside the problem of the initial water endowment, subsequent aspects of the evolution of Venus' atmosphere are now reasonably well understood (Kasting and Toon, 1989). If Venus had water originally, most of it was lost through either the runaway or moist greenhouse processes described above. One additional reason for favoring the moist greenhouse model is that it might have facilitated removal of the last few bars of Venus' water (Kasting *et al.*, 1984). If an ocean had been present on Venus for any significant length of time, it should have drawn down the atmospheric CO_2 partial pressure by providing a medium for the formation of carbonate minerals. A thinner atmosphere would, in turn, have provided less of a barrier to loss of water by photodissociation followed by hydrogen escape. For example, suppose that an initial 100-bar CO_2-N_2 atmosphere was reduced to 10 bar of total pressure by carbonate formation. The critical water abundance at which the cold trap became effective would then be reduced from 10 bar to 1 bar, based on the criterion $c_o(H_2O) < 0.1$. Only 1 bar of water would then need to be lost by the relatively inefficient hydrogen loss processes that would have operated after the cold trap had formed.

Once surface water was depleted, carbonate formation would have slowed dramatically, and CO_2 released from volcanos should have begun accumulating in the atmosphere. SO_2 concentrations would have likewise increased, and the modern sulfuric acid clouds would have started to form. Thus, regardless of its initial condition, Venus' atmosphere should eventually have approached its modern state.

STEAM ATMOSPHERES DURING ACCRETION

The possibility that Earth was enveloped in a dense steam atmosphere during the accretionary period was raised by Matsui and Abe (1986a,b, and earlier references therein). Their model elaborated on earlier studies (Benlow and Meadows 1977; Lange and Ahrens 1982) that predicted that infalling planetesimals would be devolatilized on impact once the growing Earth had reached about 30% of its present radius. The water contained in these impactors would thus have been emplaced directly into the protoatmosphere, instead of following the more traditional route of being first incorporated into the solid planet and being subsequently outgassed from volcanos.

A critical aspect of Matsui and Abe's model was that it considered the effect of the impact-induced steam atmosphere on the planetary radiation budget. Based on a relatively crude, grey-atmosphere, radiative-equilibrium

calculation, they argued that the surface temperature of such an atmosphere would rise to the approximate solidus temperature of crustal rocks, about 1500 K. The surface pressure would continue to rise until it was of the order of 100 bar. At this point dissolution of water in the partially molten planetary surface would balance the continued input of water from planetesimals and thereby stabilize the atmospheric pressure and temperature.

Matsui and Abe's fundamental predictions have been largely borne out by studies carried out using more detailed models (Kasting 1988; Zahnle *et al.* 1988; Abe and Matsui 1988). Given an accretionary time scale of 10^7 to 10^8 years (Safronov 1969), the rate of energy release from infalling material should indeed have been sufficient to maintain the atmosphere in a runaway greenhouse state (Kasting 1988, Figure 13). One objection that has been raised, however, is that none of these models have taken into account the stochastic nature of the accretion process (Stevenson 1987). If the latter stages of accretion were dominated by large impacts spaced at relatively long time intervals (Wetherill 1985), a steam atmosphere may have existed only for short time periods following these events.

Some light can be shed on this otherwise difficult question by an analysis of neon isotopic data. Craig and Lupton (1976) pointed out some time ago that the $^{20}Ne/^{22}Ne$ ratio in Earth's atmosphere (9.8) is lower than that in gases thought to originate in the mantle. Their database has now been expanded to include volcanic gases, along with trapped gases in diamonds and MORBs (mid-ocean ridge basalts) (Ozima and Igarashi 1989). The neon isotope ratio in gases derived from the mantle is generally between 11 and 14, with the lower values attributed to mixing with atmospheric neon (Ozima and Igarashi 1989). Thus, the $^{20}Ne/^{22}Ne$ ratio of mantle neon is similar to the solar ratio, which ranges from 13.7 in the solar wind to 11-12 in solar flares (Ozima and Igarashi 1989).

The neon isotopic data are most easily explained if Earth formed from material with an initially solar $^{20}Ne/^{22}Ne$ ratio, and if ^{20}Ne was preferentially lost from Earth's atmosphere during rapid, hydrodynamic escape of hydrogen (Kasting 1990). The requirements for losing neon are quite specific and can be used to set rather tight constraints on the composition of Earth's atmosphere at the time when the escape occurred. The minimum hydrogen escape flux required to carry off ^{20}Ne is 2 × 10^{13} H$_2$ mol cm^{-2} s^{-1} (Kasting 1989). If the background atmosphere at high altitudes was predominantly CO$_2$, the diffusion-limited escape rate of hydrogen is given by

$$\phi_{lim} \approx 3 \times 10^{13} \; f(H_2)/[1 + f(H_2)] cm^{-2} s^{-1} \qquad (2)$$

where f(H$_2$) is the atmospheric H$_2$ mixing ratio (Hunten 1973). By comparing this expression with the escape rate needed to carry off neon, one

can see that this is only possible if $f(H_2)$ exceeds unity, i.e. the atmosphere must be composed primarily of hydrogen. Such an atmospheric composition would have been very difficult to sustain during most of Earth's history. However, it is entirely reasonable in an impact-induced steam atmosphere, where copious amounts of H_2 could have been generated from the reaction of H_2O with metallic iron.

A second reason that the escape of neon must have occurred early is that this is the most favorable period from an energetic standpoint. The solar EUV energy flux required to power an escape rate of 2×10^{13} H_2 mol cm^{-2} s^{-1} is about 40 ergs cm^{-2} s^{-1}, or roughly 130 times greater than the present solar minimum EUV flux (Kasting 1989). EUV fluxes of this magnitude are expected only within the first 10 million years of solar system history (Zahnle and Walker 1982). Thus, the isotopic fractionation of neon must have taken place during the accretionary period, most likely in a steam atmosphere of impact-induced origin.

If this explanation for the origin of the atmospheric $^{20}Ne/^{22}Ne$ ratio is correct, it is possible to use this information to estimate the length of time that a steam atmosphere must have been present on the growing Earth. According to theory (Zahnle et al. 1988), the surface pressure of the steam atmosphere should have been buffered at a more or less constant value of ~ 30 bar. Application of the "constant inventory" model for hydrodynamic mass fractionation (Hunten et al. 1987) then predicts that the escape episode must have lasted at least five million years (Kasting 1989). Thus, even if large impacts were important and steam atmospheres were essentially a transient phenomenon, the neon isotope data implies that such conditions may have obtained during an appreciable fraction of the accretionary period.

An alternative theory for explaining the isotopic abundance pattern of atmospheric neon (and xenon) is that the fractionation occurred during the loss of a primordial H_2 atmosphere captured from the solar nebula (Sasaki and Nakazawa 1988; Pepin, manuscript in preparation). This theory appears equally viable in terms of its ability to explain the isotopic data. It differs from the steam atmosphere model in that it requires that the accretion process proceed in the presence of nebular gas. If the lifetime of the solar nebula was much less than the accretionary time scale, as predicted by Safronov (1969), then the steam atmosphere model is preferred.

THE CONTINUOUSLY HABITABLE ZONE

A third reason that runaway (and moist) greenhouse atmospheres are of interest is that they set constraints on the inner edge of the continuously habitable zone (CHZ) around the Sun (Kasting et al. 1988; Kasting and Toon 1989). The concept of the CHZ was introduced by Hart (1978, 1979).

He defined it as that region of space in which a planet could remain habitable (i.e. maintain liquid water at its surface) over time scales long enough for life to originate and evolve. Hart concluded, based on what seems in retrospect to have been an overly simplified model, that the CHZ extended from only about 0.95 AU to 1.01 AU. In Hart's model, the inner boundary of the CHZ was determined to be the distance at which an Earth-like planet would experience a runaway greenhouse at some time during the last 4.6 billion years. The outer edge of the CHZ was the position at which runaway glaciation would occur.

The climate models described earlier (Kasting 1988; Abe and Matsui 1988) show that the runaway greenhouse threshold is considerably higher than Hart had estimated. If the minimum solar flux needed for runaway is 1.4 S_o (see above), then the distance at which this would occur should be < 0.85 AU. On the other hand, the minimum solar flux required to lose water in the moist greenhouse model is only 1.1 S_o. The radial distance at which this might occur is thus 0.95 AU, in agreement with Hart's original estimate. Thus, it appears that Hart located the inner edge of the CHZ correctly, even though his reasoning was slightly flawed.

The outer edge of the CHZ, on the other hand, probably lies well beyond Hart's estimate of 1.01 AU. Hart erred because he ignored the important feedback between atmospheric CO_2 levels and climate (Walker *et al.* 1981). This story is told in detail elsewhere (Kasting *et al.* 1988; Kasting and Toon 1989) and will not be repeated here. A modern conclusion, however, is that the CHZ is relatively wide, and that the chances of finding another Earth-like planet elsewhere in our galaxy are reasonably good.

CONCLUSION

Runaway greenhouse atmospheres are much better understood than they were several years ago. Recent theoretical work has provided better estimates of the amount of heating required to trigger runaway and new ideas about where such conditions may have applied. Future advances in our understanding of the evolution of Earth and Venus will require continued theoretical work in conjunction with new data on the isotopic composition of noble gases on both planets and on the water vapor distribution in Venus' lower atmosphere.

REFERENCES

Abe, Y., and T. Matsui. 1988. Evolution of an impact-generated H_2O-CO_2 atmosphere and the formation of a hot proto-ocean on Earth. J. Atmos. Sci. 45:3081-3101.

Benlow, A., and A.J. Meadows. 1977. The formation of the atmospheres of the terrestrial planets by impact. Astrophys. Space Sci. 46:293-300.

Craig, H., and J.E. Lupton. 1976. Primordial neon, helium, and hydrogen in oceanic basalts. Earth. Planet. Sci. Lett. 31: 369-385.

Dayhoff, M.O., R. Eck, E.R. Lippincott, and C. Sagan. 1967. Venus: atmospheric evolution. Science 155:556-557.

Donahue, T.M., J.H. Hoffman, R.R. Hodges, Jr., and A.J. Watson, 1982. Venus was wet: a measurement of the ratio of D to H. Science 216:630-633.

Eberhardt, P., R.R. Hodges, D. Krankowsky, J.J. Berthelier, W. Schultz, U. Dolder, P. Lammerzahl, J.H. Hoffman, and J.M. Illiano. 1987. The D/H and $^{18}O/^{16}O$ isotopic ratios in comet Halley. Lunar Planet. Sci. XVIII:252-253.

Gold, T. 1964. Outgassing processes on the Moon and Venus. Pages 249- 256. In: Brancazio, P.J., and A.G.W. Cameron (eds.). The Origin and Evolution of Atmospheres and Oceans. Wiley, New York.

Goody, R.M., and J.C.G. Walker. 1972. Atmospheres. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Gough, D.O. 1981. Solar interior structure and luminosity variations. Solar Phys. 74:21-34.

Grinspoon, D.H. 1987. Was Venus wet? Deuterium reconsidered. Science 238:1702-1704.

Grinspoon, D.H., and J.S. Lewis. 1988. Cometary water on Venus: Implications of stochastic comet impacts. Icarus 74:430-436.

Hart, M.H. 1978. The evolution of the atmosphere of the Earth. Icarus 33:23-39.

Hart, M.H. 1979. Habitable zones around main sequence stars. Icarus 37:351-357.

Hoyle, F. 1955. Frontiers in Astronomy. William Heinemann, London.

Hunten, D.M. 1973. The escape of light gases from planetary atmospheres. J. Atmos. Sci. 30:1481-1494.

Hunten, D.M., R.O. Pepin, and J.C.G. Walker. 1987. Mass fractionation in hydrodynamic escape. Icarus 69:532-549.

Hunten, D.M., T.M. Donahue, J.C.G. Walker, and J.F. Kasting. 1989. Pages 386-422. Escape of atmospheres and loss of water. In: Atreya, S.K., J.B. Pollack, and M.S. Matthews (eds.). Origin and Evolution of Planetary and Satellite Atmospheres. University of Arizona Press, Tucson, in press.

Ingersoll, A.P. 1969. The runaway greenhouse: a history of water on Venus. J. Atmos. Sci. 26:1191-1198.

Kasting, J.F. 1988. Runaway and moist greenhouse atmospheres and the evolution of Earth and Venus. Icarus 74:472-494.

Kasting, J.F. 1990. Steam atmospheres and accretion: evidence from neon isotopes. In: Visconti, G. (ed.). Interactions Between Solid Planets and Their Atmospheres, in press.

Kasting, J.F., and J.B. Pollack. 1983. Loss of water from Venus. I. Hydrodynamic escape of hydrogen. Icarus 53: 479-508.

Kasting, J.F., and O.B. Toon. 1989. Pages 423-449. Climate evolution on the terrestrial planets. In: Atreya, S.K., J.B. Pollack, M.S. Matthews (eds.). Origin and Evolution of Planetary and Satellite Atmospheres. University of Arizona, Tucson.

Kasting, J.F., J.B. Pollack, and D. Crisp. 1984. Effects of high CO_2 levels on surface temperature and atmospheric oxidation state on the early Earth. J. Atmos. Chem. 1: 403-428.

Kasting, J.F., O.B. Toon, and J.B. Pollack. 1988. How climate evolved on the terrestrial planets. Scientific Amer. 258: 90-97.

Krasnopolsky, V.A. 1985. Total injection of water vapor into the Venus atmosphere. Icarus 62:221-229.

Lange, M.A., and T.J. Ahrens. 1982. The evolution of an impact-generated atmosphere. Icarus 51:96-120.

Matsui, T., and Y. Abe. 1986a. Evolution of an impact-induced atmosphere and magma ocean on the accreting Earth. Nature 319:303-305.

Matsui, T., and Y. Abe. 1986b. Impact-induced oceans on Earth and Venus. Nature 322:526-528.

Moroz, V.I. 1983. Summary of preliminary results of the Venera 13 and Venera 14 missions. Pages 45-68. In: Hunten, D.M., L. Colin, T.M. Donahue, and V.I. Moroz (eds.). Venus. University of Arizona Press, Tucson.

Ozima, M., and G. Igarashi. 1989. Pages 306-327. Terrestrial noble gases: constraints and implications on atmospheric evolution. In: Atreya, S.K., J.B. Pollack, and M.S. Matthews (eds.). Origin and Evolution of Planetary and Satellite Atmospheres. University of Arizona Press, Tucson.

Pollack, J.B. 1971. A nongrey calculation of the runaway greenhouse: implications for Venus' past and present. Icarus 14:295-306.

Rasool, S.I., and C. de Bergh. 1970. The runaway greenhouse and accumulation of CO_2 in the Venus atmosphere. Nature 226: 1037-1039.

Safronov, V.S. 1969. Evolution of the Protoplanetary Cloud and Formation of the Earth and the Planets. Nauka Press, Moscow, in Russian. Trans. NASA TTF-677, 1972.

Sagan, C. 1960. The Radiation Balance of Venus. JPL Tech. Rept. No. 32-34.

Sasaki, S., and K. Nakazawa. 1988. Origin of isotopic fractionation of terrestrial Xe: hydrodynamic fractionation during escape of the primordial H_2-He atmosphere. Earth Planet. Sci. Lett. 89:323-334.

Stevenson, D.J. 1987. Steam atmospheres, magma oceans, and other myths. Paper presented at the American Geophysical Union Meeting. San Francisco, Dec. 3 - 7.

von Zahn, U., S. Kumar, H. Niemann, and R. Prinn. 1983. Composition of the Venus atmosphere. Pages 299-430. In: Hunten, D.M., L. Colin, T.M. Donahue, and V.I. Moroz (eds.). Venus. University of Arizona Press, Tucson.

Walker, J.C.G. 1975. Evolution of the atmosphere of Venus. J. Atmos. Sci. 32:1248-1256.

Walker, J.C.G., P.B. Hays, and J.F. Kasting. 1981. A negative feedback mechanism for the long-term stabilization of Earth's surface temperature. J. Geophys. Res. 86: 9776-9782.

Watson, A.J., T.M. Donahue, and W.R. Kuhn. 1984. Temperatures in a runaway greenhouse on the evolving Venus: implications for water loss. Earth Planet. Sci. Lett. 68: 1-6.

Wetherill, G. 1985. Occurrence of giant impacts during the growth of the terrestrial planets. Science 228:877-879.

Zahnle, K.J., and J.C.G. Walker. 1982. The evolution of solar ultraviolet luminosity. Rev. Geophys. Space Phys. 20: 280-292.

Zahnle, K.J., J.F. Kasting, and J.B. Pollack. 1988. Evolution of a steam atmosphere during Earth's accretion. Icarus 74: 62-97.

The Oort Cloud

Leonid S. Marochnik, Lev M. Mukhin, and Roald Z. Sagdeev
Institute of Space Research

ABSTRACT

Views of the large-scale structure of the solar system, consisting of the Sun, the nine planets and their satellites, changed in 1950 when Oort (Oort 1950) demonstrated that a gigantic cloud of comets (the Oort cloud) is located on the periphery of the solar system. From the flow of observed comets of $\simeq 0.65$ yr.$^{-1}$ AU^{-1}, the number of comets in the cloud was estimated at $N_o \simeq 2 \cdot 10^{11}$. Oort estimated that the semi-major axes of the orbits of comets belonging to the cloud must lie within the interval $4 \cdot 10^4$ AU $\lesssim$ a $\lesssim 2 \cdot 10^5$ AU. This interval is now estimated to be $2-3 \cdot 10^4 <$ a $< 5-10 \cdot 10^4$ AU (see Marochnik et al. 1989).

The original estimate of the Oort cloud's mass was made on the hypothesis that the nuclei of all comets are spherical with a mean radius value on the order of R = 1 kilometer and a density of $\rho = 1^2$/cm^3. This produced an Oort cloud mass of $M_o = 0.1$ $M_\oplus$ (Oort 1950). Therefore, the comet cloud that occupies the outer edge of the solar system appeared to be in a dynamically zero-gravity state, having no effect on the mass and angular momentum distribution in it.

However, the estimate of the Oort cloud's mass was gradually increased (see below). We cannot rule out at this time the possibility that the Oort cloud has a concentration of mass comparable to the aggregate mass of the planets, in which the bulk of the solar system's angular momentum is concentrated (Marochnik et al. 1988).

THE OORT CLOUD'S MASS

The value of $N_o \simeq 2 \cdot 10^{11}$ (Oort 1950) was yielded without accounting for observational selection. Everhart (1967) was apparently the first to account for the effects of observational selection, having estimated the "true flux" of "fresh" comets as 4.7 yrs.$^{-1}$ AU^{-1}. This generated the estimate of $N_o \simeq 1.4 \cdot 10^{12}$. Monte Carlo modeling of the dynamics of the Oort cloud's comets (Weissman 1982; Remy and Mingrad 1985) also yielded figures within the interval $N_o = 1.2 - 2 \cdot 10^{12}$.

Weissman's analysis (1983) of the mass spectrum of comets in the Oort cloud demonstrated that the increase in the number of comets by an order in comparison with the original estimate of the Oort cloud is primarily due to comets of low mass and a large absolute value of H^{10}. According to Weissman's estimate (1983), $N_o \simeq 1.2 \cdot 10^{12}$ for comets whose absolute values are $H_{10} \leq 11.5$. With a density of nucleus matter $\rho = 1^2/cm^3$ and a surface albedo of $A = 0.6$, Weissman (1983) yielded $M_o \simeq 1.9$ M$_\oplus$, an average mass on the spectrum for a typical comet of $<M_w> 7.3 \cdot 10^{15}$ g and a corresponding radius of the nucleus of $<R_w> \simeq 1.2$ kilometers. In addition, $N_o{}^w / <M_w> 1.6 \cdot 10^{12}$.

Hughes (1987; 1988) however, demonstrated that Everhart's data (1967) had apparently been subjected to the effects of observational selection, since the index of the corresponding distribution function of long-period comets (LP) for absolute values (and, consequently, by mass; see below) is dependent upon perhelion distances and the epoch in which these comets are observed. If this is true, then doubt is cast over the estimate based on Everhart's data (1967) of the number of comets in the Oort cloud, generated by extrapolating the observed flux in the region of the largest values. In this case, we must return to the estimate of $N_o \simeq 2 \cdot 10^{11}$, obtained on the basis of direct observations, without taking into account the effects of observational selection. However, the mean mass of a typical "new" comet must also be estimated using direct observations, without extrapolation in the region of small dimensions and comet nucleus masses.

Direct observations of 14 bare nuclei of long-period comets (i.e., observations at great heliocentric distances) produced, according to Roemer (1966), a mean radius of $\overline{R}_{LP} = 4.2$ kilometers. Similarly, a mean radius of $\overline{R}_{LP} = 5.8$ kilometers was found for 11 comets with bare nuclei, selected by Svoren (1987) from a total number of 67 long-period comets. Both of these $\overline{R}_{LP}$ values were yielded on the hypothesis that the mean albedo of long-period comets $\overline{A}_{LP}$ is equal to $\overline{A}_{LP} = 0.6$, in accordance with the computation done by Delsemme and Rud (1973). A nucleus mass of $\overline{M}_{LP} \simeq 5 \cdot 10^{17}$ at $\rho = 1$ g/cm^3 corresponds to the average of these two values of $\overline{R}_{LP} = 5$ kilometers. For a more probable value of the density of the matter of a nucleus of $\rho \simeq 0.5$ g/cm^3 (Sagdeev *et al.* 1987)

$$\overline{M}_{LP} \simeq 2.5 \cdot 10^{17}. \tag{1}$$

Analysis of the mass spectrum of long-period comets using Hughes data (1987, 1988), i.e., without taking into account the effects of observational selection, generates a mean LP-comet mass for the spectrum (Marochnik *et al.* 1989) of,

$$< M_{LP} > \simeq 1.2 \cdot 10^{17} g, \tag{2}$$

this virtually (with an accuracy to a factor of ~ 2) coincides with (1) - and the mean values of $\overline{M}_{LP}$ according to observations of bare nuclei of LP-comets with the same albedo value of $\overline{A}_{LP} = 0.6$.

The closeness of the mean spectrum value $< M_{LP} >$ to the mean observed value $\overline{M}_{LP}$ is understandable in this case (as opposed to the case where the effects of observational selection are taken into account). Actually, in hypothesizing the effect of observational selection, we find the number of comets in the Oort cloud N_o to be an order greater than directly follows from the value for the flux of observed LP-comets (see above), due to low-mass comets of low luminosity. This should considerably reduce the value $< M_{LP} >$ as compared with $\overline{M}_{LP}$. The fact that this is true can be seen by comparing the value Weissman generated (1983) of $< M^W_{LP} > = 7.3 \cdot 10^{15}$ g with (1). At the same time, when we only use the flux of observed comets, it is clear that $< M_{LP} >$ and $\overline{M}_{LP}$ cannot differ so greatly, which follows from comparing (1) and (2).

At the same time, estimates of the Oort cloud mass M_O in both instances differ little since, despite the fact that the value of $< M_{LP} >$ according to (2) is an order greater than $< M^W_{LP} >$, N_o is an order less than N^W_o. A direct estimate based on (2) gives us:

$$M_o = N_o \cdot < M_{LP} > = 2 \cdot 10^{11} \cdot 1.2 \cdot 10^{17} g \simeq 4M, \tag{3}$$

which is approximately twice as large as M_o generated by Weissman (1983). It is, however, a value of the same order. Therefore, if we refrain from "battling" for exactness in the coefficient values on the order of two (which is completely unjustified with the framework of ambiguities in observed data), we can then conclude that both approaches (accounting for and not accounting for the effect of observational selection) produce values of one order for the Oort cloud mass of $M_o \simeq 2\text{-}4\ M_\oplus$, with a mean albedo of the nuclei of long-period comets of $\overline{A}_{LP}\ 0.6$.

At the same time, direct measurements of the albedo of Halley's comet give us an albedo value of $A_H = 0.04^{+0.02}_{-0.01}$ (Sagdeev *et al.* 1986). If we hypothesize that comets in the Oort cloud have an albedo which on the average approaches A_H, then this must lead to an appreciable

overestimation of the mass of M_o. Since the mass of the nucleus of a comet is $M \sim A^{-3/2}$, reduction of albedo by $0.04/0.6 = 1/15$ times triggers an increase in the mass of the average comet and consequently, the mass of the entire Oort cloud by $(15)3/2 \simeq 58$ times. Naturally, this must produce radical cosmogonic consequences. We will note that this circumstance was first noticed by Weissman (1986). Having assumed that $\overline{A}_{LP} = 0.05$, he found that $M_o \simeq 25 \ M_\oplus$. The corresponding estimate by Marochnik *et al.* (1988) produced $M_o \simeq 100 \ M_\oplus$.

What is the reasoning for hypothesizing that the values of $\overline{A}_{LP}$ and A_H are approximately equal?

We will first of all note that the mass of the "mean" short-period (SP) comet cannot be greater than the mass of the "mean" LP-comet. That is, the following ratios must be fulfilled:

$$\overline{M}_{LP} \gtrsim \overline{M}_{SP}; < M_{LP} > \gtrsim < M_{SP} >, \tag{4}$$

if, of course, we do not presuppose that LP- and SP-comets have varying origins (Marochnik *et al.* 1988). This study demonstrated that the measurements made during the Vega mission of the mass and albedo of Halley's comet (M_H and A_H) are typical for SP-comets, and approach the mean values of:

$$A_H \simeq \overline{A}_{SP} \simeq 0.04, \tag{5}$$

$$M_H \simeq \overline{M}_{SP} \simeq 3 \cdot 10^{17}.$$

At the same time, an estimate of the loss of mass by Halley's comet during its lifetime has demonstrated that its initial mass was, probably, an order greater than its contemporary mass (Marochnik *et al.* 1989) and this (owing to M_H's convergence with typical mass values for SP-comets) allows us to hypothesize that the mean mass of a comet in the Oort cloud must be, apparently, at least an order greater than the present values of $\overline{M}_{SP}$ and $< M_{SP} >$ in accordance with (4).

On the other hand, according to Hughes (1987; 1988), the functions of comet distribution by their absolute values for LP- and SP-comets are homologous. In other words, the cumulative number of comets $N_{cum}(H_{10})$ (i.e., the aggregate number of comets whose absolute values of $\leq H_{10}$) for LP- and SP-comets have the appearance in the logarithmic scale of straight lines of equal inclination up to the corresponding inflection points in the spectra. These "knees" in the spectra of LP- and SP-comets have values of $H_K^{LP} = 5.8$ and $H_K^{SP} = 10.8$, respectively (Hughes 1987). In the regions of $H_{10}^{SP} > H_K^{SP}$ and $H_{10}^{LP} > H_K^{LP}$, "saturation" occurs: the curves acquire a very gentle slope. The values of H_K^{LP} and H_K^{SP} are close to the mean

value for the spectra, and in the regions $H_{10}^{LP} < H_K^{LP}$ and $H_{10}^{SP} < H_K^{SP}$ the effects of observational selection are minor.

The relationship between the absolute values of H_{10} and the masses of LP- and SP-comets were explored by a number of authors (Allen 1973; Öpik 1973; Newburn 1980; Whipple 1975; Weissman 1983).

We will use Weissman's data (1983) who produced the following dependency from Roemer's data (1966) for LP- and SP-comets:

$$\log M_{LP} = 19 - 0.4 H_{10} + \frac{3}{2} \log \left(\frac{\overline{A_{LP}}}{0.6} \right) + \log(\rho/lg/cm^3) \qquad (5a)$$

$$\log M_{SP} = 20.5 - 0.3 H_{10} + \frac{3}{2} \log \left(\frac{A_{SP}}{0.05} \right) + \log(\rho/lg/cm^3). \qquad (5b)$$

From (5a) we find the ratios of masses corresponding to the "knees" in the LP- and SP-comet spectra to be equal to:

$$\frac{M_K^{LP}}{M_K^{SP}} = 0.18 \cdot \left(\frac{A_{LP}}{0.6} \right)^{-3/2} \left(\frac{A_{SP}}{0.04} \right)^{3/2}. \qquad (6)$$

It clearly follows from (6) that the albedo value of $\overline{A}_{LP} = 0.6$, assumed for LP- comets, directly contradicts (4). Formula (6) can be rewritten as:

$$\frac{M_K^{LP}}{M_K^{SP}} = 10.8 \cdot (A_{SP}/A_{LP})^{3/2}. \qquad (7)$$

Therefore, the mean mass of LP-comets can only be an order greater than the mean current mass of SP-comets on the condition that

$$\overline{A}_{LP} \simeq \overline{A}_{SP}. \qquad (8)$$

We will note that there are also physical reasons for hypothesizing the close values of $\overline{A}_{LP}$ and $\overline{A}_{SP}$. A low albedo is a consequence of the formation of a thin layer of dark material on the surface of the comet nucleus. Data from laboratory experiments on irradiation by energy protons of low-temperature ices (that contain H_2O, CH_4, and organic residues) demonstrate the formation of a black graphite-like material (Strazzula 1986).

As Weissman has pointed out (1986b), the effect of galactic cosmic rays on comet nuclei in the Oort cloud (before their appearance in the region of the planetary system) must, for the aforementioned reason, lead to the formation of a sufficiently thick crust from the dark, graphite-like

polymer. The latter acts as a "cometary paste" binding the nucleus surface against sublimation.

It is our view that owing to the low heat conductivity of this polymer layer, a low albedo of the surface of comet nuclei can be maintained by a layer thickness of several centimeters. Due to the low volatility of this layer and its "sticky properties," the latter must also be conserved as the comet shifts into a short-period orbit.

Therefore, if we are to propose that the hypothesis (8) is correct, we can estimate the mass of the Oort cloud to be a value of $M_o \simeq 100 \, M_\oplus$ (with an accuracy of up to a factor on the order of two).

HILL'S CLOUD MASS

It has currently been deemed likely that the canonical Oort cloud is only a halo surrounding a dense, internal cometary cloud. This cloud contains one to two orders of cometary nuclei greater than the halo with an outer boundary corresponding to the semi-major axis, $a_e^c = 2-3 \cdot 10^4$ AU (Hills 1981; Heisler and Tremaine 1986). It is a source which delivers comets to the halo as the latter is depleted when the Sun approaches closely passing stars and gigantic molecular complexes in the galaxy (GMC) and under the impact of the galactic tidal effects. The internal cometary cloud is sometimes called the Hills cloud. The outer boundary of the Hills cloud is defined quite clearly, as Hills demonstrated (1981), since comets with semi-major axes of $a < a_e^c$ do not fill the loss cone in the velocity space delivering them to the planetary system region of the solar system, where they have been recorded through observation. According to Hills, the value of a_e^c is weakly dependent on the parameters input into the formula to determine this value (an indicator of the degree of 2/7). Therefore, the value of $a_e^c \simeq 2 \cdot 10^4$ AU is defined with sufficient certainty. Bailey (1986) also later generated the same value for the outer inner cometary cloud (ICC) boundary prior to this; he considered interaction with GMC instead of convergence with stars, as Hills had done. Furthermore, if the tidal effect of the "galaxy's vertical gravitational pull" is taken into account, we have, according to Heisler and Tremaine (1986) an estimate of $a_e^c \simeq 3 \cdot 10^4$ AU.

The location of the inner boundary of a_i^c is considerably less definite. An extreme estimate, performed by Whipple (1964) produces $a_i^c \gtrsim 50$ AU. At the same time, by hypothesizing that comets are formed in the outer regions of the protosolar nebula, Hills (1981) estimated the inner boundary of the core as $a_i^c \simeq 3 \cdot 10^3$ AU.

What is the mass of the Hills cloud? Let us designate the number of comets in it as N_{core}, so that

$$N_{core} = \beta \cdot N_o. \tag{9a}$$

Then

$$M_{core} = \beta M_o, \tag{9b}$$

where the value of β is not clearly known.

What can be said about the value of β? Simple extrapolation for the core of the law of comet distribution around the semi-major axis in the halo for the original Oort model and a somewhat refined version produce, according to Hills' estimate (1981), $\beta = 20$ and $\beta = 89$, respectively.

Proposing that comet formation occurs in the Uranus-Neptune zone, Shoemaker and Wolfe (1984) and Duncan *et al.* (1988) generated $\beta = 10$ and $\beta = 5$, respectively, in their numerical experiments. The internal boundary of the core in the latter instance was equal to $3 \cdot 10^3$ AU; this fits with Hills' estimate (1981).

However, it was proposed in these computations that the total mass of comets scattered by Uranus and Neptune is minor when compared with the masses of the planets. Clearly, this is not true if the reasoning put forward in this paper is sound. For this reason, the results generated by the authors mentioned here apparently require clarification.

Assuming, nevertheless, the region of parameter alteration as:

$$\beta = 5 - 10, \tag{10}$$

we find the mass of the Hills cloud approximately

$$M_{core} \simeq 500 - 1000 M_\oplus. \tag{11}$$

Figure 1 represents schematically the probable mass distribution in the solar system for $\beta = 10$, in the case of a massive Oort cloud.

ANGULAR MOMENTUM DISTRIBUTION IN THE SOLAR SYSTEM

If the Hills and Oort clouds are truly as massive as follows from the above estimations, then: (1) comet formation could hardly have occurred in the Uranus-Neptune zone, as is frequently considered, since as such a large mass was ejected to the periphery of the solar system, the planets should have moved considerably closer to the Sun (Marochnik *et al.* 1989); (2) since comet formation apparently took place in the rotating protoplanetary disk (if, of course, we rule out the hypothesis of cometary cloud capture during the Sun's formation through GMC) (Clube and Napier 1982), then since the angular momentum is conserved, the greater portion of it must,
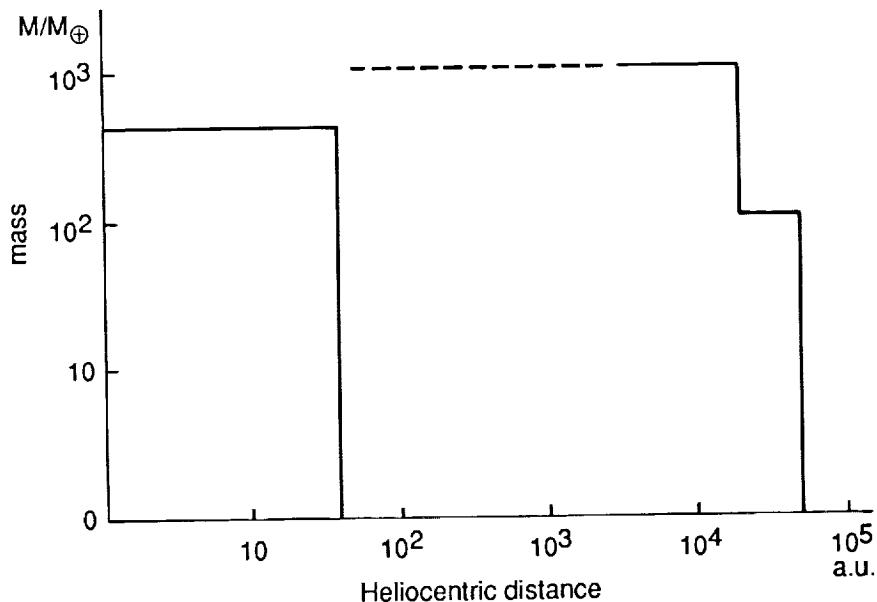
FIGURE 1   Histogram of the probable mass distribution in the solar system. A planetary system with a mass of ΣM_{planet} = 448 $M_\oplus$ is located in the region of heliocentric distances where r $\leq$ 40 AU. The Oort cloud with a mass of $M_o \simeq$ 100 $M_\oplus$ is located in the zone of $2 \cdot 10^4 \leq$ r $\leq$ $5 \cdot 10^4$ AU. A Hills cloud with a mass of $M_{core} \simeq 10^3$ $M_\oplus$ is located in the region where r $\leq$ $2 \cdot 10^4$ AU. The internal boundary $r_i{}^c$ of the Hills cloud is ambiguous. According to data from Hills (1981) and Duncan *et al.* (1988), $r_i{}^c \simeq 3 \cdot 10^3$ AU. However, neither can we exclude the value $r_i{}^c \simeq$ 50 AU (Whipple 1964).

apparently, be concentrated in the massive Oort and Hills clouds, and not the planets.

According to Marochnik *et al.* (1988) the Oort cloud's angular momentum can be written as

$$J_o = 4/3 M_o (GM_\odot \alpha_{min})^{1/2}(1 + \alpha^{-1/2}) - 1, \qquad (12)$$

where the original Oort model is used (Oort 1950) for the function of comet distribution by energies (n = 2); $\alpha = \alpha_{max}/ \alpha_{min}$; α_{min} and α_{max} denote the minimum and maximum possible semi-major axes of cometary orbits. Since the Oort cloud is thermalized by passing stars, integration occurs in (12) for all possible eccentricities (0 $\leq$ e $\leq$ 1).

Assuming that M_o = 100 $M_\oplus$, α_{min} = $2 \cdot 10^4$ AU, α_{max} = $5 \cdot 10^4$ AU, we find that

$$J_o = 3 \cdot 10^{51} g \cdot cm^2/s. \qquad (13)$$

For the assumed parameter values, the angular momentum of the halo is on the same order as the minimum possible angular momentum of the protosolar nebula before it loses its volatiles (Hoyle 1960; Kusaker *et al.* 1970; Weidenschilling 1977) and an order greater than the present angular momentum of the planetary system. $\Sigma J_{planet} \simeq 3 \cdot 10^{50}$g cm^2/s. Estimate (13) fits with the hypothesis of the *in situ* formation of comets, and thus generates the upper limit of the Oort cloud's possible angular momentum.

The lower limit will clearly be seen if we suppose that comet formation occurred in the Uranus-Neptune zone. Assuming, for example, that $\alpha_{min} = 25$ AU, $\alpha_{max} = 35$ AU, and supposing that the initial comet orbits in this case are nonthermalized and circular, we find that

$$J_o = 1.5 \cdot 10^{50} gcm^2/s.$$

Therefore, the interval in which Oort cloud angular momentum may lie can be written as:

$$1.5 \cdot 10^{50} gcm^2/s < J_o < 3 \cdot 10^{51} gcm^2/s. \qquad (14)$$

In any case, as we have seen:

$$J_o \gtrsim \Sigma J_{planet}. \qquad (15)$$

Let us now estimate the angular momentum for the Hills cloud. Since it is apparently nonthermalized, its angular momentum J_{core} is equal to (Marochnik *et al.* 1989):

$$J_{core} = 2M_{core}[GM_\odot \alpha^c_{min}(1 - e^2)]^{1/2}(1 + \alpha_c^{-1/2})^{-1}, \qquad (16)$$

where, as in conclusion (12), the classical Oort model is used: (n = 2), α^c_{min} and α_c modify the core.

Let us consider two extreme cases: (a) all the comets in the Hills cloud move along circular Kepler orbits (e = 0), and (b) all the comets in it move along sharply elongated, circumparabolic orbits (e < 1). In the case where e = 0, assuming $M_{core} \simeq 10^3$ M$_\oplus$, $\alpha^c_{min} = 3 \cdot 10^3$ AU, $\alpha^c_{max} = 2 \cdot 10^4$ AU, we find:

$$J_{core} \simeq 2 \cdot 10^{52} g \cdot cm^2/s. \qquad (17).$$

Estimate (17) agrees with the suggestion that comets are formed *in situ* and gives us an upper limit for the value J_{core} (for estimating the momentum, present values of α^c_{min} and α^c_{max} are used). In the case where e $\lesssim$ 1, we need to rewrite (16) in terms of perihelion distances of q = a(1 − e), that is, using the ratio

$$\alpha^c_{min}(1 - e^2) \simeq 2q_{min}, \qquad (18)$$
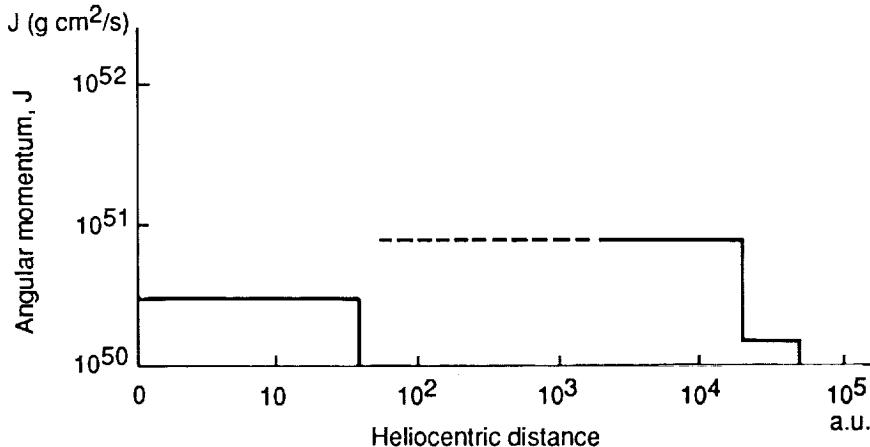
FIGURE 2 Histogram of the probable distribution of angular momentum in the solar system. A cometary system with a total angular momentum of $\Sigma J_{planet} = 3 \cdot 10^{50}$ is located in the region of heliocentric distances of $r \leq 40$ AU. An Oort cloud with a mass of $M_o = 100$ $M_\oplus$ and angular momentum of $J_o = 1.5 \cdot 10^{50}$ g cm^2/s is situated in the zone of $2 \cdot 10^4 \leq r \leq 5 \cdot 10^4$ AU. A Hills cloud with a mass of $M_{core} = 500$ $M_\oplus$, an internal boundary that is ambiguous ($50 \lesssim r_i{}^c \lesssim 3 \cdot 10^3$ AU) and an angular momentum of $J_{core} = 3 \cdot 10^{51}$ g·cm^2/s is located in the zone of $r \leq 2 \cdot 10^4$ AU.

where q_{min} denotes the perihelion distances of the comets' cores, which have minimum semi-major axes $a_{min}{}^c$. The lower limit for J_{core} can be produced, supposing that the formation of the comets of the Hills cloud occurred in the Uranus-Neptune zone. Assuming that $q_{min} \simeq 25$ AU, we find from (16), and taking into account (18), that:

$$J_{core} \simeq 3 \cdot 10^{51} g cm^2/s.$$

Therefore, for the angular momentum of J_{core}, we can write the following estimate:

$$3 \cdot 10^{51} < J_{core} \left(\frac{M_{core}}{10^3 M_\oplus}\right)^{-1} < 2 \cdot 10^{52} \frac{g \cdot cm^2}{s}. \tag{19}$$

The angular momentum of J_{core} is, therefore, very large: one to two orders greater than the contemporary angular momentum of the entire planetary system ΣJ_{planet}. However, its value still does not exceed the limits of the upper estimate of the possible initial angular momentum of the protosolar nebula (Marochnik et al. 1988).

Figure 2 shows angular momentum distribution in the solar system in the case where the Oort and Hills ($\beta = 5$) clouds are not too massive.

THE COMETARY CLOUD AROUND OTHER STARS

The picture we have described of the structure of the solar system apparently does not contradict IRAS data on observations of infrared excesses in stars of the circumsolar vicinity of the galaxy.

As Backman notes (personal communication), IRAS data shows that thin clouds of solid particles are spread out over distances of up to 700 AU near the stars α Lyrae and β Pictoris, and, possibly, up to 10^4 AU, respectively.

According to Smith's data (1987), optical observations of β Pictoris point to the presence around this star of an elongated (radius $\simeq 11^{50}$ AU) and a thin (projected thickness is h $\simeq$ 50 AU) disk. According to data from Smith and Terrile (1984), optical observations of β Pictoris may also be evidence of the presence of a zone of transparency with a radius of about 30 AU around this star.

Of the 150 main sequence stars in Glize's catalogue, and which were examined by Backman (1987), 18% demonstrated infrared excess at a level of 5 sigma. This exceeds the extrapolated photospheric flow. According to Backman's analysis (1987), this may indicate the presence of thin clouds of solid particles spread out over distances of 10 + 1000 AU from the respective stars.

The presence of elongated disks of fine solid particles spread out over distances of hundreds and thousands of AU from the respective stars is most likely evidence of the presence in these regions of bodies of cometary dimensions for which the sublimation of volatiles and mutual collisions may offset the accretion of dust grains from these regions by light pressure and the Poyting-Robertson effect during the lifespan of a star (Weissman 1984; Harper *et al.* 1984; O'Dell 1986). As Beichman has noted (1987), the masses of disks around the stars are apparently the most difficult values to determine. A huge spread in estimates exists for β Pictoris: from $M_{disk} \simeq 10^{-2}$ $M_\oplus$ for dust grains of equal dimensions to $M_{disk} \simeq 3 \cdot 10^3 M_\oplus$ for an asteroid-like distribution of them (Aumann *et al.* 1984; Weissman 1984; Gillet 1986).

Therefore, the data of infrared observations of stars near the solar vicinity of the galaxy do not apparently clash with the proposed model of the structure of the solar system. Attempts to make some stronger assertions would be too speculative at this point.

CONCLUSION

Assuming as typical for long-period comets the albedo value of Halley's comet, we come to the conclusion that the Oort cloud must be extremely

massive ($M_o \simeq 100\ M_\oplus$). This mass must be located in the region of the semi-major axes of the orbits:

$$2 - 3 \cdot 10^4 AU \lesssim \alpha \lesssim 5 - 10 \cdot 10^4 AU.$$

The hypothetical Hills cloud is located in the region $\alpha < 2 - 3 \cdot 10^4$ AU. It has a mass of $M_{core} = \beta M_o$, where the value β may generally be included in the ranges of $0 \leq \beta \leq 100$.

The extreme case of $\beta = 0$ fits with the general hypothesis of the absence of the Hills cloud. The case where $\beta = 100$ agrees with the other extreme hypothesis of the highly massive core.

In limiting ourselves to the values of $\beta = 5 - 10$, we find the Hills cloud mass of $M_{core} = 500 - 1000\ M_\oplus$, which must have a value of $a_e^c = 2 - 3 \cdot 10^4$ AU as an outer border. The internal boundary of the Hills cloud is indefinite.

Therefore, the hypothesis of the low value for the albedo of comets in the Oort cloud brings us to the conclusion that at the outer edge of the solar system there may be an invisible material in the form of cometary nuclei whose mass, ΣM_{comet}:

$$\Sigma M_{comet} \gtrsim \Sigma M_{planet},\qquad(20)$$

where ΣM_{planet} denotes the total mass of the planetary system.

The second conclusion states that the cometary population, and not the planetary system accounts for the bulk of the solar system's angular momentum. On the basis of (14) and (19) we can write the following, which is analogous to (20):

$$\Sigma J_{comet} \gtrsim \Sigma J_{planet},\qquad(21)$$

where ΣJ_{comet} denotes the total angular momentum of the cometary population. Let us note in conclusion that if (20) and (21) are correct, the cosmogonic scenarios for the solar system's origin call for considerable refinement.

ACKNOWLEDGMENT

REFERENCES

Allen, C.W. 1973. *Astrophysical Quantities*. Athlone Press, London.

Aumann, H.H., F.C. Gillett, C.A. Beichman, T. de Jong, J.R. Houck, *et al.* 1984. Astrophys. J. Lett. 278:L23.

Backman, D. 1987. IRAS Statistics in IR excess and models of circumstellar discs. In: IAU. Coll. 99, Book of Abstracts, Hungary.

Bailey, M.E. 1986. Mon. Not. Roy. Astron. Soc. 218:1.

Beichman, C.A. 1987. Ann. Rev. Astron. Astrophys. 25:521.

Clube, S.V.M., and W.M. Napier. 1982. Quart. J. Roy. Astron. Soc. 23:45.

Delsemme, A.H., and D.A. Rud. 1973. Astron. Astrophys. 28:1.

Duncan. M., T. Quinn, and S. Tremaine. 1988. The Formation and Extent of the Solar System comet cloud. Preprint.

Everhart, E. 1967. Astron. J. 72:1002.

Gillet, F.C. 1986. Pages 61-69. In: Israel, E.P. (ed.). Light on Dark Matter.

Harper, D.A., R.F. Loewenstein, and J.A. Davidson. 1984. Astrophys. J. 285:808.

Heisler, J., and S. Tremaine. 1986. Icarus 65:13.

Hills, J.G. 1981. Astron. J. 86:1730.

Hoyle, F. 1960. Quart, J. Roy. Astron. Soc. 1:28.

Hughes, D.W. 1987. Mont. Not. Roy. Astron. Soc. 226:309.

Hughes, D.W. 1988. Icarus 73:149.

Hughes, D.W., and P.A. Daniels. 1980. Mont. Not. Roy. Astron. Soc. 191:511.

Kusaka, T., T. Nakano, and C. Hayashi. 1970. Progr. Theor. Phys. 44:1580.

Marochnik, L.S., L.M. Mukhin, and R.Z. Sagdeev. 1988. Science 242:547.

Marochnik, L.S., L.M. Mukhin, and R.Z. Sagdeev. 1989. Astrophys. and Space Phys. Rev., in press.

O'Dell, C.R. 1986. Icarus 67:71.

Okamoto, J. 1969. Publ. Astron. Soc. Japan 21:1.

Oort, J. 1950. Bull. Astron. Inst. Neth. 11:91.

Opik, E.J. 1973. Astrophys. Space Sci. 21:307.

Remy, F., and F. Mingrad. 1985. Icarus 63:1.

Roemer, E. 1966. The dimensions of cometary nuclei. Pages 23-28. In: Nature of Origine des Cometes. Publ. Inst. d'Astrophys. Liege, Belgium.

Sagdeev, R.Z., V.I. Moroz, and P.E. Eliasberg. 1987. Astr. Zh. Letters 13:621.

Sagdeev, R.Z., J. Blamont, A.A. Galeev, V.I. Moroz, V.D. Shapiro, V.I. Shevchenko, and K. Szego. 1986. Nature 321:259.

Shoemaker, E.M., and R.F. Wolfe. 1984. Abstract. Lunar and Planet Sci. Conf. XV:780.

Smith, B. 1987. Beta Pictoris: What's new? In: IA.U. Coll. 99. Book of abstracts.

Smith, B., and R.J. Terrile. 1984. Science 226:1421.

Spinrad, H. 1987. Ann. Rev. Astron. Astrophys. 25:231.

Strazzula, G. 1986. Icarus 67:63.

Svoren, J. 1987. Consequences of the size determination of p/Halley by space probes on the scale of sizes of cometary nuclei. Pages 707-712. In: Proc. Int. Symp. on the diversity and similarity of comets 6-9 April. Brussels, Belgium.

Weidenschilling, S. 1977. Astrophys. Space Sci. 51:153.

Weissman, P.R. 1982. Page 637. In: Wilkening, L. (ed.). Comets. University of Arizona Press, Tucson.

Weissman, P.R. 1983. Astron. Astrophys. 118:90.

Weissman, P.R. 1984. Science 224:987.

Weissman, P.R. 1986a. Bull. Am. Astron. Soc. 18:799.

Weissman, P.R. 1986b. ESA SP-249.

Whipple, F.L. 1964. Proc. Nat. Acad. Sci. USA 51:711.

Whipple, F.L. 1975. Astron. J. 80:525.

The Chaotic Dynamics of Comets and the
Problems of the Oort Cloud

ROALD Z. SAGDEEV AND G.M. ZASLAVSKIY
Institute of Space Research

ABSTRACT

This paper discusses the dynamic properties of comets entering the planetary zone from the Oort cloud. Even a very slight influence of the large planets (Jupiter and Saturn) can trigger stochastic cometary dynamics. Multiple interactions of comets with the large planets produce diffusion of the parameters of cometary orbits and a mean increase in the semi-major axis of comets. Comets are lifted towards the Oort cloud, where collisions with stars begin to play a substantial role. The transport of comets differs greatly from the customary law of diffusion and noticeably decelerates the average comet flow. The vertical tidal effect of the galaxy in this region of motion is adiabatic and cannot noticeably alter cometary distribution. A study of the sum of forces operating in the region to a $\sim 10^4$ AU does not permit us to explain at this time the existence of a sharp maximum, where a $\sim 10^4$ AU in the distribution of long-period comets. This is an argument in favor of the suggestion that it was caused by the close passage of a star several million years ago.

INTRODUCTION

The solar system's new object, the Oort cloud, arose as a source of long-period comets (Oort 1950) in the planetary system's visible portion ($r < 2$ AU). Experimental material generated by processing the trajectory of a large number of long-period comets (Marsden et al. 1978; Marsden and Roemer 1982) determined for these comets the region in which they exist, which reaches a size of up to $\lesssim 2 \cdot 10^5$ AU. Oort proposed that collisions

with stars passing fairly close to the Sun may be one of the primary causes for which comets attain the visible region. Research during the ensuing years greatly complicated the Oort cloud model, inputting Hills cloud (Hills 1981) into the analysis (with an upper boundary of $r \sim 2 \cdot 10^4$ AU) and the action of various forces such as the galactic tidal effect (Hills 1981; Heisler and Tremaine 1986; Morris and Muller 1986; Bailey 1986), collision with molecular clouds (Biermann and Lust 1978; Hut and Tremaine 1985), and interaction with the planets (Oort 1950; Khiper 1951).

Numerical analysis within the framework of the simplest, initial Oort model demonstrated the possibility of qualitatively explaining the reason for which comets enter the visible zone due to the effect of near stellar passages (Wiesmann 1982). Subsequent analysis showed that the action of the galaxy's vertical tidal effect may somewhat modify Oort cloud and Hills cloud parameters and the number of comets in them (Fernandez and Ip 1987; Duncan et al. 1988). Oort cloud mass and momentum may fluctuate more significantly if we assume certain, typical estimates for them, made after processing the results of the Halley's comet mission (Marochnik et al. 1988). The large mass of the Oort cloud ($M_o \sim 100$ $M_\oplus$, $r > 2 \cdot 10^4$ AU) must affect in the most serious way models of the formation of the solar system.

It should be added here that the increase in the mass of the Hills cloud must also bring about an increase in its angular momentum (Marochnik et al. 1988). This must be reconciled with the approximate equality of the number of prograde and retrograde new comets. If it was not a question of new comets, this equality would be a sufficiently obvious consequence of the impact of random collisions of stars with comets with highly eccentric orbits. However, numerical simulation, where the initial angular momentum value of the cometary protocloud is taken into account, also reveals the considerable impact it exerts on the size of the final Oort cloud and on the number of comets in it (Lopatnikov et al. 1989). The obvious reason for this is tied to the different impacts of stellar collisions on circular and eccentric orbits.

In accounting for the final angular momentum of cometary distribution in the Hills and Oort clouds, an anisotropy is created in cometary dynamics at virtually all of its stages. Anisotropy in the distribution of cometary aphelia has been experimentally discovered (Delsemme 1987). It indicates the correlation between cometary distribution and the effect of galactic tidal forces. We can consider that these forces exert an influence on both cometary dynamics in the aphelion region (Heisler et al. 1987) and on their dynamics in the planetary zone. Hence, all characteristic regions of cometary orbits must participate in an interrelated manner in the formation of cometary zones. This makes it necessary to analyze more carefully all of the processes by which comets interact with the planets and the stars.

The interaction of comets with the stars is statistical. Therefore, the impact of individual collisions is averaged out, while the mechanism itself by which they have an impact on the zone of cometary aphelia is weakly dependent on individual details.

The passage of comets through the planetary zone is quite different. The influence of the large planets, Jupiter and Saturn, on cometary dynamics has proven more subtle. A significant portion of comets (about 50%) which pass close to Jupiter are thrown into hyperbolic orbits as early as the first passage. However, the phase space occupied by comets with a perhelion $\lesssim$ 2 AU is not very large. The phase space occupied by comets with a perihelion similar to the radius of Jupiter's orbit is substantially greater. The phenomena of chaos may arise for such comets (Petrosky 1986; Sagdeev and Zaslavskiy 1987; Petrosky and Broucke 1988; Sagdeev *et al.* 1988). They consist of the following: in the strictly dynamic, three-body problem (Sun-Jupiter-comet), the movement of the comets in varying conditions becomes unstable. This instability is seen, in particular, in the fact that Jupiter's phase at the moment when a comet passes through its perihelion is a sequence of random numbers. As a result, a mechanism accelerating comets begins to operate. This mechanism is analogous to Fermi's method of stochastic acceleration (Sagdeev *et al.* 1988). It produces diffusive alteration of all of the comet's parameters, a mean increase of the semi-major axis of cometary orbit, and the expulsion of the comet from the solar system. The process of stochastization of cometary movement is considered in detail in Natenson *et al.* (1989). Numerical analysis demonstrated that the region of cometary eccentricity values for which chaos arises is very broad. Orbits with $\epsilon \sim 0.5$ may already become stochastic. This circumstance should noticeably modify views of cometary interaction with the planetary zone.

We will discuss below the conditions in which the dynamics of comets with long-period orbits become stochastic, the role of such comets in the overall model of the Oort cloud, and the influence of the galactic tidal forces on cometary dynamics.

THE DYNAMIC CHAOS OF COMETS

We can generate a straightforward idea of this chaos by looking at how a ball falls on a heavy plate in a gravity field (Figure 1, Zaslavskiy 1985) and if we consider their collision to be absolutely elastic. If the plate oscillates with an oscillation amplitude of $\bar{a}$ and a velocity amplitude of $\bar{v}$, then on the condition that:

$$2\bar{v}^2 \gtrsim \bar{a}g,$$

(g denotes acceleration in the gravity field), the oscillation phase of the
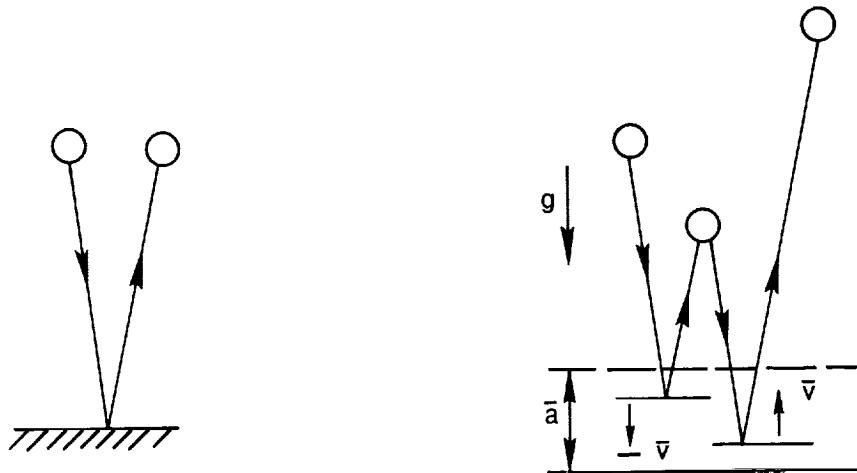
FIGURE 1   Illustration of the action of the "gravitational machine" (Zaslavskiy 1985), triggering the stochastic increase of the energy of a ball bouncing on a periodically oscillating plate.

plate at the moment of collision is random. The ball bounces irregularly over the plate and, on the average, rises increasingly higher. Its mean energy at the moment of impact behaves asymptotically, as in $<v^2> \sim t^{2/3}$, and the mean height of lift correspondingly increases, as in $<h^1> \sim t^{2/3}$. The time for the ball to return back to the plate, naturally, increases. However, the acceleration process does not stop.

This example somewhat clarifies what occurs with long-period comets whose perihelion is in the sphere of influence of, for example, Jupiter. Let M denote the orbital momentum of a comet, and ψ be the phase of a comet's location relative to Jupiter in the comet's orbital plane. If the orbits of Jupiter and the comet lie in the same plane (the so-called flat, limited three-body problem), the Hamiltonian for the comet is equal to

$$H = \frac{1}{2}\left(p_r^2 + \frac{1}{r^2}p_\varphi^2\right) - \sigma p_\varphi - \frac{1-\mu}{r_1} - \frac{\mu}{r_2} \tag{1}$$

$\sigma = \pm 1$ defines here either prograde or retrograde comets; μ denotes the ratio of Jupiter's mass to that of the Sun;

$$r_1 = [r^2 + 2\mu r\ cos(\varphi - \sigma w_J t) + \mu^2]^{1/2}$$

$$r_2 = [r^2 - 2(1 - \mu)r\ cos(\varphi - \sigma w_J t) + (1 - \mu)^2]^{1/2},$$

where the radius of Jupiter's orbit, r_J, is assumed to be equal to the

unit, w_J (Jupiter's rotational frequency). In this two-dimensional motion, $\psi = \psi - \sigma w_J t$, and a Jacoby motion integral exists:

$$G_o = H - \sigma w_J M. \tag{2}$$

Using ratio (2), we can make problem (1) a two-variable, nonstationary problem. We can select a canonically coupled pair (M, ψ) as the variables. Let, for example, t_n denote the moment in time when the comet passes through its aphelion, M_n be the orbital momentum during passage through aphelion, and ψ_n denote the value of the Jovian phase during cometary passage through perihelion (preceding the time t_n). The relationship between the values (M_{n+1}, ψ_{n+1}) and M_n, ψ_n) then defines the expression (Petrosky 1986; Sagdeev and Zaslavskiy 1987)

$$M_{n+1} = M_n + \Delta M \ sin\psi_n \tag{3}$$

$$\psi_{n+1} = \psi_n + 2\pi\sigma \frac{w_j}{w(E_{n+1})},$$

where ΔM denotes fluctuation in the orbital momentum over one cometary rotation, while the comet's energy, E_n is determined using the motion integral (2):

$$E_n = G_o + \sigma w_J M_n. \tag{4}$$

The variables (M_n, ψn) are canonically coupled. Expression (3) is only defined in the region of negative values of a comet's energy, H = E < 0, that is , according to (2)

$$G_o + \sigma w_J M < 0. \tag{5}$$

Inequality (5) is violated and expression (3) becomes meaningless when a comet is thrown into hyperbolic orbit. The value ΔM in (3) is defined by the expression

$$\Delta M = max \int_{t_n}^{t_{n+1}} dt \frac{\partial H}{\partial \psi}, \tag{6}$$

where t_n, and t_{n+1} denote the moments of time of two sequential passages of the apocenter by a comet. Estimates of the value of ΔM in varying cases are provided in Petrosky 1986; Sagdeev and Zaslavskiy 1987; Petrosky and Broucke 1988; Natenson *et al.* 1989.

Expression (3) has a frequently encountered form, described in detail by Sagdeev *et al.* (1988) and Zaslavskiy (1985). If a comet does not pass too far from Jupiter, the duration of its interaction with Jupiter is on the order of a Jovian period of $2\pi/w_J$. This time scale determines the duration

of a "collision." It is a great deal less than the time period between two "collisions" for long-period comets of $2\pi/w(E)$. In view of this circumstance, we can write a simple form of expression (3), in which the second formula simply describes alteration of the ψ phase during the time between two sequential collisions. A similar expression also occurs for the model with the ball in Figure 1. The velocity, v, plays the role of a generalized pulse, M, while the collision frequency is proportional to $1/v$. In this case, $w(E) = |2E|^{3/2}$.

We can produce a straightforward assessment of the stochastic dynamics of the comet for problem (3) from the condition that (Sagdeev *et al.* 1988; Zaslavskiy 1985):

$$K = \left| \frac{\partial \psi_{n+1}}{\partial \psi_n} - 1 \right| \gtrsim 1. \tag{7}$$

This gives us

$$K = 2\pi \frac{w_J}{w^2(E)} \left| \frac{\partial w(E)}{\partial E} \right| \cdot |\Delta M \cos \psi| > 1. \tag{8}$$

Since the perihelion changes little as a result of the collision ($\Delta M \ll M$), while the comet's rotation frequency $w(E) = |2E|^{3/2} \to 0$ where $|E| \to 0$, condition (8) can be fulfilled for comets with sufficiently low binding energy ($|E| \to 0$) with a fixed value M. The phase portrait of cometary movement, corresponding to the Hamiltonian (1), with a fixed Jacoby integral value is in Figure 2 (Natenson *et al.* 1989). It was produced for true trajectories and demonstrates the complex structure of phase space with a large number of regions of stability. The region of global chaos is defined by the estimate in (7) and (8). A boundary with $\epsilon \approx 0.55$ and a semi-major axis of a ≈ 16.5 AU corresponds to this. The comet's perihelion originally had a value of q ≈ 7.5 AU.

Points in the region of global chaos in Figure 2, that is where $0 <$ E $< - 0.03$ AU, belong to one trajectory. If any initial condition in this region is selected, with the same Jacoby integral value, the corresponding movement of a comet will also be stochastic. This is where the significance of this region of stochasticity is manifested.

Two important consequences stem from the results of study (Natenson *et al.* 1989). The region of chaos is very significant, and even Halley's comet enters the zone where conditions of chaos are applicable. An analogous comment regarding Halley's comet, based on the use of representation (3), was made by Chirikov and Vecheslavov (1985). The region of chaos in Figure 2 also applies to medium-period comets. Therefore, the phase magnitude of comets with stochastic dynamics is an order more than the phase magnitude of comets appearing in the visible portion with r < 2 AU.

C- 4
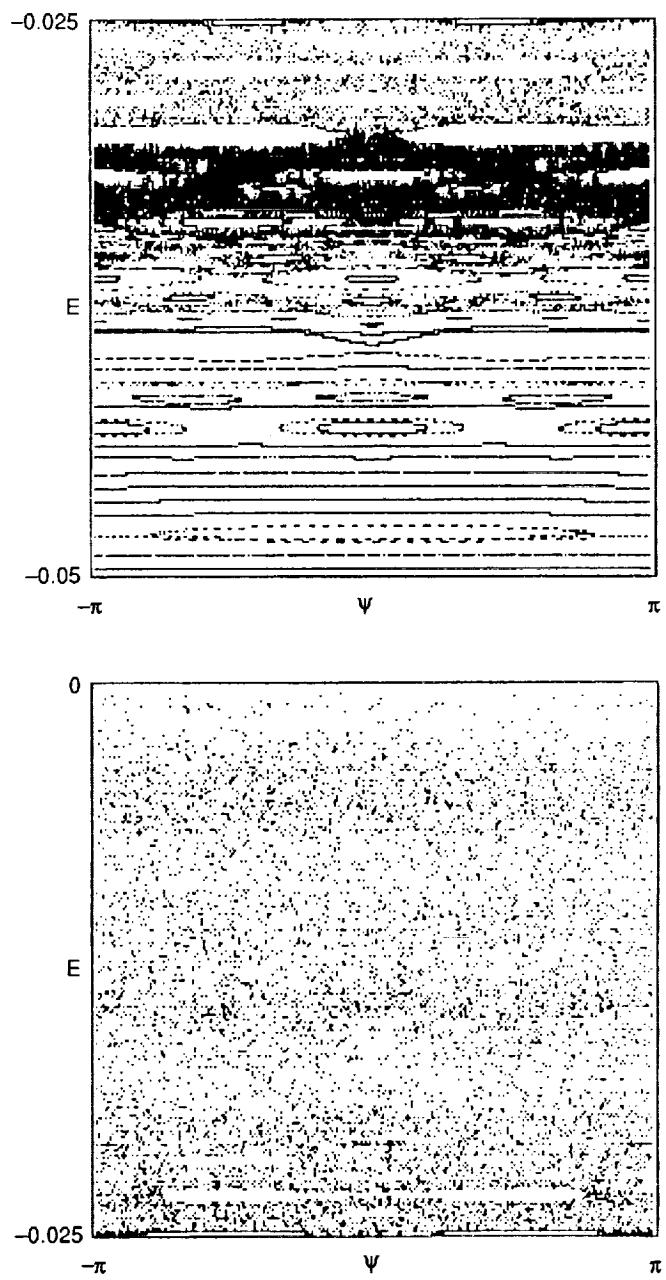
FIGURE 2   Phase portrait on the trajectory plane (ψ, E = $-1/2a$) of a comet. The orbit points are plotted at the moment in time when a comet passes through the aphelion point. For the sake of convenience, the portrait has been broken into two parts.

The second consequence is related to the nature of cometary diffusion. This process is extremely important, since it is the mechanism by which comets attain the region of semi-major axes. The usual diffusion formula (Yabushita 1980)

$$\frac{\partial f(M,t)}{\partial t} = \frac{1}{4\pi}\frac{\partial}{\partial M} < (\Delta M)^2 w(M) > \frac{\partial f(M,t)}{\partial M}, \qquad (9)$$

is only justified in a region sufficiently remote from the chaos boundary. The impact of the chaos boundary and regions of stability (see Figure 2) significantly decelerates diffusion at the initial stage in comparison with the diffusion defined by formula (9) (Natenson et al. 1989).

GENERAL COMETARY DYNAMICS AND
THE INFLUENCE OF THE GALACTIC TIDE

The existence of a mechanism of dynamic chaos makes it necessary to reconsider the general dynamics of comets as they move from the Oort cloud to the visible zone. The customary route is that collisions with stars operate in the zone of aphelion of a comet's orbit. Those comets in the loss cone region enter the visible zone, originally having a $\gtrsim 10^4$ AU. Jupiter's influence throws about one half of these comets into hyperbolic orbit. Only a small portion of the comets may subsequently return again directly to the loss cone, in order to set out on the new path from the Oort cloud to the visible zone.

However, another avenue also exists.

A rather large portion of comets, those that first entered the invisible zone and have a perihelion comparable to the radius of Jupiter's orbit, enter the region of stochastic dynamics. The comet begins a long, diffusive path to the loss cone region. Therefore, an independent way of filling the loss cone is defined.

Other large planets of the solar system may also play the same role as Jupiter. Therefore, the portion of comets moving stochastically should be insignificant. The planetary "barrier," expelling part of the comets, concurrently makes the dynamics of others stochastic, thereby providing their route to the loss- cone. At the same time, the cometary perihelion changes very slightly, and the comets' orbits are as if "attached" in the zone of their perihelion.

Vertical tidal galactic forces must play an important role in the process of cometary stochastic transport described above (Bahcall 1984; Heisler and Tremaine 1986)

$$F_z = 1\pi m G \rho_s z, \qquad (10)$$

where m denotes comet mass, G is the gravitational constant, $\rho s = 0.186$ $m_\odot/pc^3$ is stellar density, and z is the vertical coordinate in the galactic system of coordinates. The tidal force (10) considerably alters the perihelion. Therefore, it creates the drift of comets from the planetary zone during one phase of cometary orbit and, inversely, causes cometary perihelia to flow into the planetary barrier in another orbital phase. These two fluxes are approximately equal.

We will find the region, on the nonadiabatic influence of the tidal force F_z, from the condition that perhelion variation under its influence must be fairly strong. We will assume 10 AU as an example of the characteristic size of the planetary zone. Then the nonadiabatic condition means that

$$\Delta q > 10 AU, \tag{11}$$

where Δq denotes perihelion alteration under the influence of F_z in a period of cometary orbit. For eccentric orbits

$$\Delta q \sim M^2/2m^2 m_\odot G.$$

Thus

$$\Delta q \sim M \Delta M/m^2 m_\odot G.$$

Assuming

$$\Delta M \sim F_z aT \sim 4\pi mG\rho_* a^2 T \sim nm_\odot Ga^2 T/pc^3,$$

after substitution of all of these expressions in (11), we yield:

$$a \gtrsim 10^4 AU.$$

This estimate (Duncan *et al.* 1988) can also be clarified using a more careful input of numbers. However, it is clear that the effect of tidal forces in the region of a significant portion of cometary orbits is adiabatic. Therefore, the tidal forces cannot substantially alter cometary distribution in the region $\lesssim 10^4$ AU. However, they exert considerable influence on the near-boundary processes, where the planetary barrier operates, and along the border of the Oort cloud, where effective collisions with stars fill the loss cone (Fernandez and Ip 1987; Duncan *et al.* 1988).

CONCLUSION

The dynamic chaos of comets with fairly eccentric orbits, moving in the Sun's field and perturbed by the fields of Jupiter and Saturn (or by the fields of other remote planets), forces us to reconsider individual elements

of the Oort cloud theory. Cometary stochastization creates a mechanism by which comets move away from the planetary belt towards the Oort cloud, and may be an additional source by which the loss cone is filled. The process of diffusive cometary transport differs from the usual process of diffusion and retards the characteristic time scale for the flux of comets toward the semi-major axes a. The action of the tidal forces does not alter this time scale significantly and is adiabatic (with the exception of the regions near the belt of the large planets and near the inner boundary of the Oort cloud). Therefore, the internal mechanisms of cometary dynamics cannot explain the existence of a sharp maximum in the distribution of the observed comets from the Oort cloud (Marsden *et al.* 1978; Weismann 1982) with a period on the order of several million years. This gives us reason to suggest that the reason for the appearance of such comets may have been the last near-Earth passage of a star. This conclusion correlates with the conclusions of studies Biermann *et al.* 1983; Lust 1984) on the possible passage of a star or another large object in the region of cometary orbit with a $\sim 10^4$ AU, triggering the appearance of a coherently moving cometary cluster.

The global modeling of the dynamics of long-period comets must include the multiple interactions of comets with the large planets, if the perihelion of the comets does not greatly exceed the radii of planetary orbits. These issues and the existence of asymmetry in cometary cloud distribution are discussed in greater detail in Lopatnikov *et al.* (1989); Natenson *et al.* (1989).

ACKNOWLEDGEMENT

REFERENCES

Bahcall, J.N. 1984. Astrophys. J. 276:169.
Bailey, M.E. 1986. Nature 324:350.
Biermann, L., W.F. Huebner, and R. Lust. 1983. Proc. Natl. Acad. Sci. USA 80:5151.
Biermann, L., and R. Lust. 1978. Sitz. ber. Bayer. Akad. Wiss. Mat.-Naturw. K1.
Chirikov, B.V., and V.V. Vecheslavov. 1986. Preprint:86-184. Institute of Nuclear Physics. Novosibirsk.
Delsemme, A.H. 1987. Astron. Astrophys. 187:913.
Duncan M., T. Quinn, and S. Tremaine. 1988. University of Toronto.
Fernandez, J.A., and W.H. Ip. 1987. Icarus 71:46.
Heisler, J., and S. Tremaine. 1986. Icarus 65:13.
Heisler, J., S. Tremaine, and C. Alcock. 1987. Icarus 70:269.
Hills, J.G. 1981. Astron. J. 86:1730.
Hut, P., and S. Tremaine. 1985. Astron. J. 90:1548.
Khiper, G.P. 1951. Page 357. In: Hynek, J.A. (ed.). Astrophysics. McGraw-Hill, New York.

Lopatnikov, A., L.S. Marochnik, D.A. Usikov, R.Z. Sagdeev, and G.M. Zaslavskiy. 1989. Space Research Institute, Moscow, in press.

Lust, R. 1984. Astron. Astrophys. 141:94.

Marochnik, L.S., L.M. Mukhin, and R.Z. Sagdeev. 1988. Science 242:547.

Marsden, B.G., Z. Sekanina, and E. Everhart. 1978. Astron. J., 83:64.

Marsden, B.G., and E. Roemer. 1982. Page 707. In: Wilkening, L. (ed.). Comets. The University of Arizona Press, Tucson.

Morris, D.E., and R.A. Muller. 1986. Icarus 65:1.

Natenson, M.Ya., A.I. Neishtadt, R.Z. Sagdeev, G.K. Seryakov, and G.M. Zaslavskiy. 1989. Space Research Institute, Moscow, in press.

Oort, J.H. 1950. Bull. Astron. Inst. Neth. 11:91.

Petrosky, T.Y. 1986. Phys. Lett. A. 117:328.

Petrosky, T.Y., and R. Broucke. 1988. Celestial Mechanics 42:53.

Safronov, V.S. 1969. Evolution of the protoplanetary cloud and the formation of the Earth and the planets. Nauka, Moscow.

Sagdeev, R.Z., D.A. Usikov, and G.M. Zaslavskiy. 1988. Nonlinear Physics. Harwood Academic Publishers, Chur.

Sagdeev, R.V., and G.M. Zaslavskiy. 1987. Nuovo Cimento 97:119.

Weismann, P.R. 1982. Page 637. In: Wilkening, L.L. (ed.). Comets. Arizona University Press, Tucson.

Yabushita, S. 1980. Astron. Astrophys. 85:77.

Zaslavskiy, G.M. 1985. Chaos in Dynamic Systems. Harwood Academic Publishers, Chur.

Progress in Extra-Solar Planet Detection

ROBERT A. BROWN
Space Telescope Science Institute

INTRODUCTION

The solar system's existence poses this fundamental question: Are planetary systems a common by-product of star formation? One supporting argument is that flattened disks appear to be abundant around pre-main sequence stars (Strom *et al.* 1988). Perhaps the planetary orbits in the solar system preserve the form of such a disk that existed around the young Sun. Such heuristic evidence notwithstanding, real progress on the general question requires determining the frequency of occurrence of extra-solar planetary systems and measuring their characteristics (Black 1980).

At the current time (the beginning of 1989) no investigator has announced an extra-solar planet detection that is unqualified or that has been generally accepted as such. Indeed, the very definition of "planet" is ambiguous. The quest for planets is an arduous challenge—the classic astronomical grail.

This paper reviews progress to date. Several observing programs have measured direct light from sub-stellar masses orbiting other stars. Those observations are helpful in understanding why planets have not been found by the same techniques: their visibility is very low as compared with more luminous bodies like brown dwarfs.

Three investigator groups claim to have found evidence for smaller bodies, perhaps planets, by studying perturbations in star motions. Those observations are instructive about the specific strengths and weaknesses of indirect techniques for detecting planets with various masses and orbits.

More capable extra-solar planet searches are being planned for the
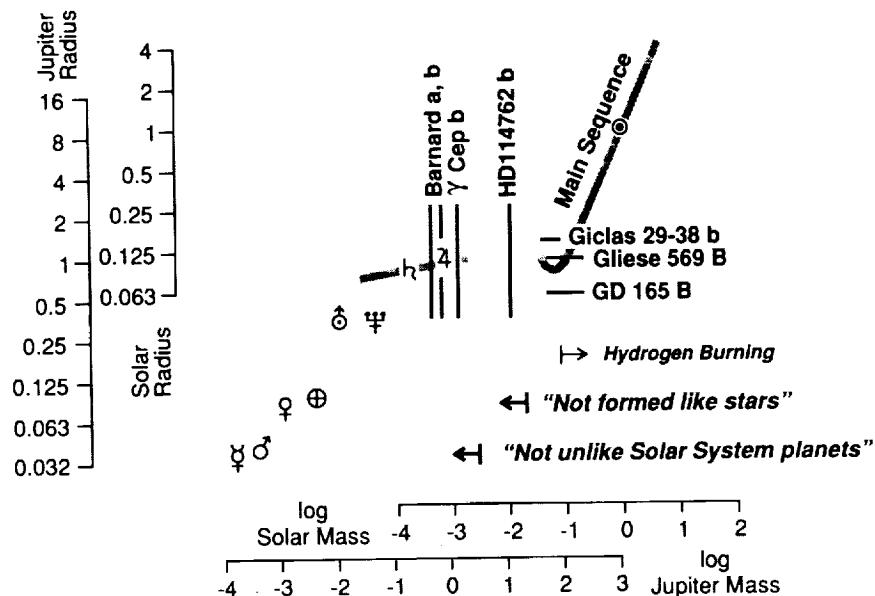
FIGURE 1 Mass-radius spectrum for planets, brown dwarfs, and stars. Solar system objects are indicated by their customary symbols. The vertical lines show the masses of claimed planet detections, but the radii are indeterminate. The horizontal lines show the radii of selected low-mass stars or brown dwarfs, but the masses are uncertain.

future. In the course of time, such observing programs will illuminate planet formation as an embedded process in star formation.

WHAT IS A PLANET?

First, what is a star? The mass range for stars is customarily stated as $M_* \geq 0.08 M_\odot$, where nuclear energy generation dominates gravitational contraction over the stellar lifetime (Bahcall 1986). Smaller astronomical objects, if they are not planets, are "brown dwarfs".

Figure 1 shows the mass and size of solar system bodies greater than $10^{-7} M_\odot$. In a restricted sense, Jupiter is a maximal planet and, indeed, has some stellar characteristics. Its density is approximately solar, and its internally generated luminosity adds about 70% to the sunlight it re-radiates thermally. If increased in mass, Jupiter would grow hotter and more self-luminous. Starlight would exert less influence over the object's outer characteristics.

In the solar system, the planets' surfaces and atmospheres have properties that are determined primarily by their distances from the Sun. Based

narrowly on such *intrinsic* characteristics, an object larger than a few Jupiter masses *could* be labeled a brown dwarf because it no longer resembles a solar system planet.

From the broader perspective motivated in the very first sentence, a planet is better defined by its *origin* (Black 1986). Multiple-star systems are thought to form by the fragmentation of spinning protostars. Generally, this mechanism produces elliptical orbits, and the process is not capable of producing bodies below a minimum mass—about $0.02M_\odot$ for solar metallicity (Boss 1986). A less massive secondary could form initially only by solid-body accretion in the special environment of a dense circumstellar disk (Wetherill and Stewart 1989). By the broad view then, any object less than $0.02M_\odot$ is a presumptive planet.

Hubbard (1984) has discussed the class of objects sufficiently massive $(M \geq 0.1M_\odot)$ to stabilize by deuterium burning for a brief period ($\sim 10^7$ years), but which are not stars due to the fact that they continue gravitational collapse. Van de Kamp (1986) uses a similar criterion to set Jupiter's mass as the upper limit to planets and the lower limit to brown dwarfs.

Further, dynamical and physical aspects of planets are central to current theories of planetary system formation, such as circular orbits and aligned spins. These *systemic* aspects could provide other definitions of "planet" that are more directly based on the origins concept. For example, two $0.02M_\odot \leq M \leq 0.08M_\odot$ masses in co-planar, circular orbits around a star could be called a planetary system by the common-origins criterion. The bodies then, would be planets. Further, they might also be called brown dwarfs depending on scientific motivation and on whether an "origins" or "intrinsic" criterion were preferable for *that* definition.

Finally, life originated in the solar system, and its occurrence poses a second fundamental question that is related hierarchically to the first posed above: Is the appearance of life a common by-product of planet formation? Perennial interest in that question suggests further restrictive criteria on "planets," such as a benign primary star, orbital stability, and sufficiently cool temperature so as not to break chemical bonds.

Mass is the critical issue for current planet search programs, and this review uses mass as the discriminating factor for planets. In Figure 1, the mass range $0.003M_\odot \leq M \leq 0.02M_\odot$ is the transition zone from the narrow definition, meaning "not unlike solar system planets" to the broad definition, meaning "could not have formed like stars."

PLANET SIGNALS ARE WEAK

Astronomical light carries six dimensions of information: one each, spectral and temporal, and two each, spatial and polarization. In principle, a strategy based on any combination of these could provide evidence for

planets around other stars, but the variations in time are most powerful: the *fact* of the orbit is confirming evidence, even if it is not the source of the observable effect itself.

Any experimental design for a search program presents particular opportunities and impediments to the astronomer. However, all approaches face one problem in common: the planet's signal is always very weak, in both absolute and relative terms.

Planet search techniques are either direct or indirect. Direct techniques use light from the planet itself. Indirect searches seek variations in starlight that imply the planet's presence.

Except for the most massive planets, the main source of difficulty for direct detection is the planet's low intrinsic luminosity. Starlight can easily overwhelm planet light. (Figure 2 compares the spectral luminosity of Jupiter and the Sun.) For indirect searches, the problem is the planet's small mass or small radius. Reflex motions are proportional to the planet/star mass ratio, and occultation effects vary with the planet/star radius ratio squared. Finally, the distance from the observer gives the planet orbit a small angular size, which is a problem for spatial techniques, either direct or indirect.

The Jupiter-Sun system is often used as a standard test example for planet detection schemes. Viewed from a distance of 5pc,[1] Jupiter would be 26th magnitude in the visual, and, at maximum orbital separation, it would be located only one arc second from a 4th-magnitude star. This poor flux ratio ($\sim 10^9$) improves to 11.5 magnitudes (4×10^4) at the wavelength of Jupiter's thermal spectrum peak, $\lambda = 20\mu$m. However, a diffraction-limited telescope operating at $\lambda = 20\mu$m must be 40 times larger than an optical telescope in order to separate the planet and star images as effectively. With regard to indirect detection, the solar reflex displacement in the Sun-Jupiter test case is only $1R_\odot$ or 0.001 arc seconds at 5pc distance. The reflex speed is only 13 meters per second or 0.6% of the Sun's equatorial rotation speed. (Intrinsic stellar phenomena can also produce observational effects that mimic reflex motion.) Finally, if Jupiter passes in front of the Sun, as it does for 0.1 % of the celestial sphere for just 30 hours every 12 years, the apparent solar flux is diminished by only 1%.

Low signal, high background, and low information rate: these are the trials awaiting those who would quest for extra-solar planets. Programs to detect planets must be exquisitely sensitive, robust, and patient.

[1] Only about 50 stars are nearer than 5 pc, and none is closer than 1 pc.
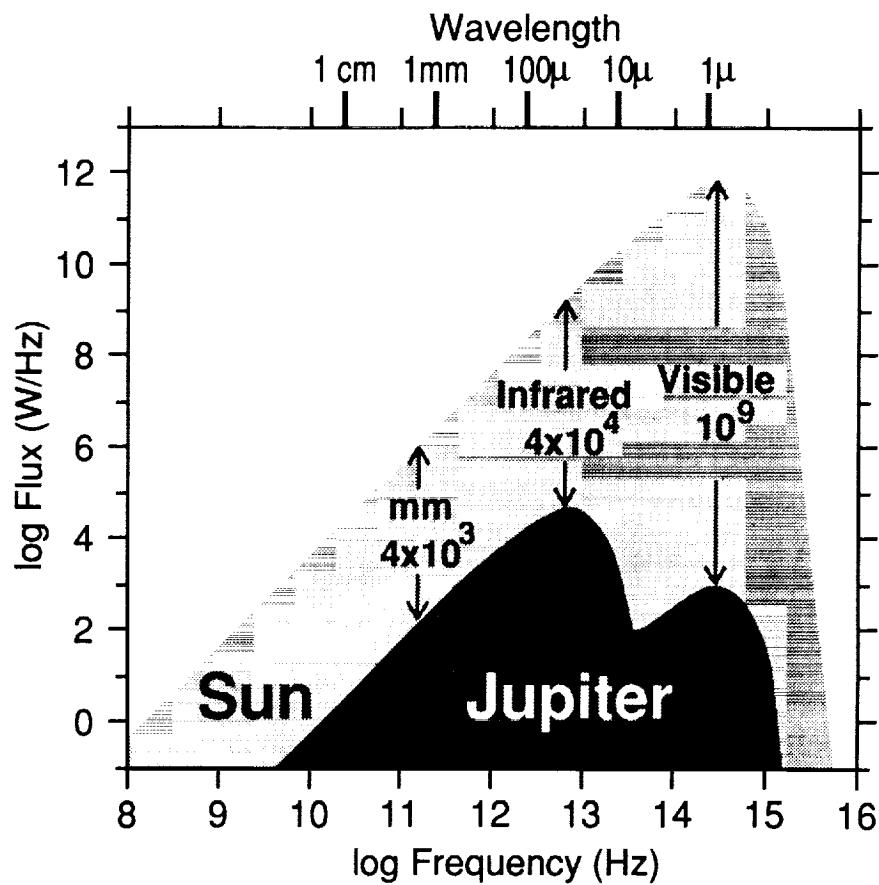
FIGURE 2   The specific fluxes of the Sun and Jupiter versus wavelength and frequency of light.

DIRECT TECHNIQUES: IMAGING AND SPECTRAL DETECTION

With a single exception that will be discussed immediately, no investigator has claimed a direct detection of an extra-solar planet. Therefore, this section's approach is to discuss the basic observational difficulties and the future prospects. The least massive objects that *have* been directly detected are brown dwarfs, and those observations are already described elsewhere.

The exception is "VB 8B". In 1985, McCarthy *et al.* reported the detection, via infrared speckle interferometry, of a cool (1360K) companion to the M dwarf star, VB 8. They stated that the observation might be the first direct detection of an extra-solar planet. However, subsequent star

evolution models have shown that the described object could not have mass less than $0.04M_\odot$ (Nelson *et al.* 1986; Stringfellow 1986). More recent observations have failed to confirm the existence of the companion to VB 8 (Perrier and Mariotti 1987).

In principle, diffraction effects limit even an ideal telescope's ability to separate the planet image from the star image. The imperfect optics in real telescopes also scatter starlight into the planet image, masking its signal to some additional degree. For ground-based telescopes, the atmosphere aggravates the problem by refracting a further amount of light from the image core into the wings; an effect called "seeing."

For planet searches by direct imaging, the critical instrumental factor is the contrast in surface brightness, which is the ratio of the brightness of the planet's image core to the starlight in the same region of the telescope focal surface. In concert with the absolute planet flux, this contrast ratio governs the fundamental rate at which information can accumulate about the planet's presence or absence. If the information rate is too low, systematic errors will prevent the planet's detection (Brown 1988).

No existing long-wavelength or ground-based system offers sufficient contrast even to approach the direct imaging problem for extra-solar planets. Either the Airy diffraction pattern is too wide, or seeing is, or both are. In the foreseeable future, only space-based telescopes operating at visible and near- infrared wavelengths offer a reasonable chance of success. (Strongly self- luminous companions, very low-mass stars and brown dwarfs, are an exception discussed below.)

Free from seeing, space-based telescopes will improve the planet/star contrast ratio at shorter wavelengths by offering narrow, diffraction-limited cores. Even so, special procedures will still be needed to reduce the *wings* of the point-spread function so that direct-imaging planet detection will be feasible. (Very large, very young planets are an exception and are discussed below.)

Figure 3 illustrates the case of the Hubble Space Telescope (HST). Brown and Burrows (1989) computed the expected HST point-spread function and applied it to the test case, the Jupiter-Sun system viewed from 5pc. Because HST operates only from the ultraviolet to the near-infrared, it can detect only reflected starlight and not thermal radiation from a Jupiter-size planet. At $\lambda = 0.5\mu m$, figure-error scattering from the HST mirrors and Airy aperture diffraction contribute approximately equally to the unwanted background at the planet-image position.[2] In this example, the predicted Jupiter/Sun contrast ratio is 6×10^{-5}, which is unfavorable. Because of the

---

[2] Figure error scattering dominates at larger angles or shorter wavelengths.

DIRECT EXTRASOLAR PLANET IMAGING USING HST

- HST Parameters
  - 2.4m aperture
  - "diffraction-limited" for $\lambda > 3600$Å

- Jupiter and the Sun at 5pc distance
  - 5 AU subtends 1"
 - 10% optical bandwidth: 4750-5250Å
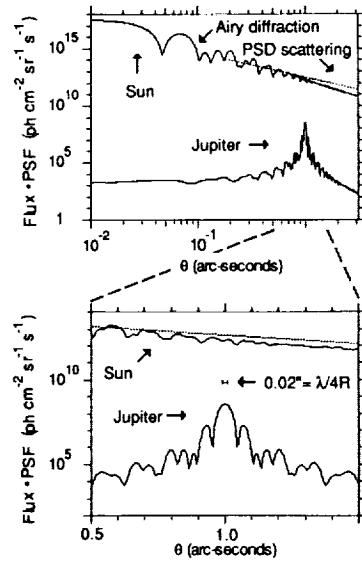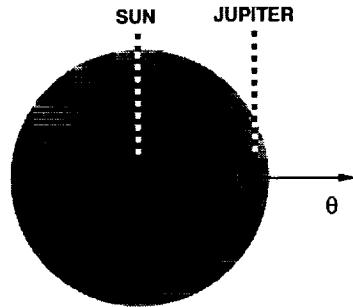
FIGURE 3   The contrast problem in detecting an extra-solar planet in reflected light. The example is Jupiter and the Sun as seen from 5 pc. Using HST, the predicted Jupiter/Sun contrast ratio is 6×10^{-5}, which is unfavorable.

lengthy integration times required by information theory, and the systematic problems they introduce, Brown and Burrows concluded that planet detection in reflected starlight is technically infeasible for HST.

The following discussion of low-mass stars and brown dwarfs is not complete. Its purpose is to demarcate the frontier for direct observations of sub-stellar objects.

Because of their self-luminosity, the very low-mass stars are now detectable in multiple star systems using near-infrared array detectors. Becklin and Zuckerman (1989) have imaged an example next to the white dwarf GD 165. Though GD 165 is about six times hotter than the discovered secondary, 12,000 K vs. 2,130 K, the white dwarf is about six times smaller than its companion, GD 165 B. Based on the temperature and the measured flux, the radius of GD 165 B is $0.061R_\odot \pm 0.015R_\odot$ versus $0.011R_\odot$ for GD 165 A. The increased surface area compensates for the temperature difference, and the two sources appear about equally strong in the near infrared.

The mass and nature of GD 165 B are uncertain. Classically, stellar spectrophotometry is translated into mass using a theoretical model of luminosity and *effective* temperature versus mass and age; evolutionary

tracks on the Hertzsprung-Russell (H-R) Diagram. Currently, though, it is not possible to do this confidently in the mass range $0.05M_\odot \le M \le 0.2M_\odot$. Models predict that heavy brown dwarfs dwell for a long time ($\sim 10^9$ years) in the region of the H-R Diagram near the least massive stars (Nelson et al. 1986; D'Antona and Mazzitelli 1985). Furthermore, existing models disagree as to where the evolutionary tracks actually lie.

Observational factors compound the confusion in using a theoretical mass-luminosity relationship for cool objects. Because the spectral characteristics of the cool emitting atmosphere are poorly understood, the reduction of the observed color temperature into an effective temperature is somewhat uncertain (Berriman and Reid 1984).

Cool companions to white dwarfs can also be directly detected by spectroscopy even when the image cannot be isolated. For example, Zuckerman and Becklin (1987) have found excess flux in the near-infrared spectrum of Giclas 29-38. At $\lambda = 1\mu m$, the two components have approximately equal signals, but the cooler source is 10 times more luminous than the white dwarf at $\lambda = 5\mu m$. The color temperature of the excess flux is 1200K. In this case, since the secondary source has not been separately imaged, the observations do not rule out dispersed dust as a possible source. Nevertheless, Zuckerman and Becklin (1987) favor the condensed source interpretation, "Giclas 29-38 b," for which the estimated photometric radius is $0.15R_\odot$.

No star could conceivably be as cool as 1200K. Giclas 29-38 b would be a definite brown dwarf, but its mass is indeterminate in the range $0.04M_\odot \le M \le 0.08M_\odot$. For the age of the white dwarf however, the radius of Giclas 29-38 b is in conflict with existing models, which predict the radius should be 50% smaller than observed.

For GD 165 A/B and Giclas 29-38 a/b, the primary and secondary are comparably bright in a limited spectral region because the objects are very different. The same situation occurs, of course, in cases where the objects are *similar*, for example, very unmassive. In just such a case, Forrest et al. (1988) have used an infrared array detector to image a cool companion to the red dwarf star Gliese 569. The colors and fluxes measured by Becklin and Zuckerman (1989) place Gliese 569 at a hotter (2775K), more luminous point ($0.11R_\odot$) on the H-R diagram than GD 165 B. Because brown dwarfs theoretically spend much less time in their hot, luminous stage, Gliese 569 B is more likely to be a star than GD 165 B. Nevertheless, a young age, a lower mass, and a brown dwarf label for Gliese 569 B are not ruled out.

The radii of GD 165 B, Giclas 29-38 b, and Gliese 569 B are plotted in Figure 1. The significance of these observations for planet searches is two-fold. First, they exemplify the breadth and intensity of current interest in probing the environments of stars. Second, they show the advancing state of the art in cool-object spectrophotometry and the benefits of the new infrared array detectors. However, they have not demanded the major

improvements in telescope imaging characteristics required by the extra-solar planet problem.

If Jupiter could not be imaged at a distance of 5pc using current telescope systems, but a brown dwarf could, what about a *large, young* planet (Black 1980)? Consider the pair $M = 0.02M_\odot$ and $M_* = 0.35M_\odot$ at a mutual age of 10^7 years. The planet flux would be 2% or more of the stellar flux longwards of $\lambda = 1\mu m$, and with an appropriate detector, HST could easily detect this planet. Outside the first bright Airy ring ($\theta >$ 0.2 arc seconds at $\lambda = 1\mu m$), HST will suppress the image wings by $>10^3$ with respect to the core. Under those conditions, the contrast ratio for this planet-star pair would be a favorable 20-to-1.

Relatively *unobscured* T Tauri stars would be prime targets for HST to examine for large, young planetary companions. They have the right age, and they permit viewing into the immediate stellar vicinity. The young stars in the Taurus-Auriga dark cloud complex are at a distance 150pc, where 0.2 arc seconds corresponds to 30AU. For these stars, HST would be expected to image very large planets within 10-20AU.

INDIRECT TECHNIQUES: REFLEX MOTION AND OCCULTATIONS

Three investigator groups have claimed indirect detections of what may be extra-solar planets based on observed stellar reflex motions. This section discusses the general methods involved and the planet findings. Indirect detections of low-mass stars or brown dwarfs are not discussed.

A star with a single planetary companion executes a reflex orbit that is isomorphic, co-planar, and synchronous with the planet orbit. The star orbit is smaller than the planet's by the ratio of the planet to star mass. If it can be detected, the star's miniature orbit implies that a second body exists. Further, if the star's orbit can be measured, and the star's mass estimated, the companion's mass and orbital radius are discovered.

For multiple planet systems, the reflex motions are independent and additive in the short term. The following discussion treats the restricted case of a single planet in a circular orbit.

The reflex orbit's two measurable aspects are first-order changes in the star's line-of-sight speed and second-order variations in its apparent position. (The lower-order terms are the components of normal inter-star motion: constant radial velocity and proper motion.) Figures 4 and 5 explain the basic geometry, physics, and parameterizations for the two types of planet search based on stellar reflex motion: the astrometric search and the radial velocity search.

The radial velocity and astrometric techniques produce respectively one- and two-dimensional data records versus time. The objective is to discover periodic variations in those records.

RADIAL VELOCITY PLANET SEARCH
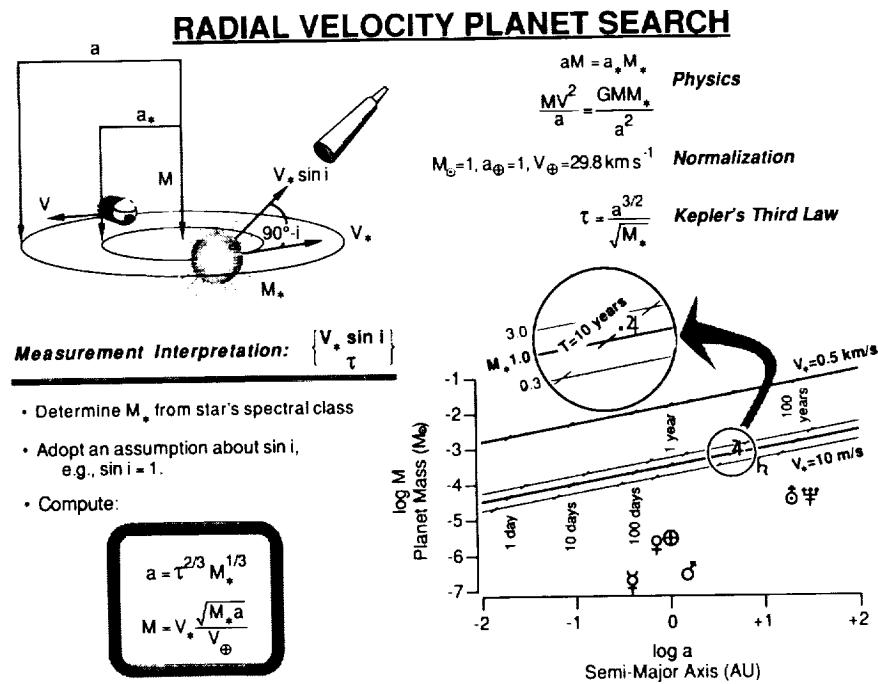
FIGURE 4   A tutorial on the radial velocity technique for indirectly detecting extra-solar planets.

The Scargle periodogram (1982; Horne and Baliunas 1986) is a standard procedure to discover and assign statistical significance to periodic signals. Black and Scargle (1982) have discussed it in the context of astrometry, and their mathematical results also apply to the radial velocity approach. The detection efficiency, for example, is the key to knowing the minimum detectable signal and for interpreting null results.

For long periods, where the observations may cover only about one cycle, the periodogram's performance needs to be better understood in purely mathematical terms.[3] Black and Scargle (1982) have also identified a potential source of systematic error in the long-period regime due to incorrect accommodation of linear drifts. Because long periods are associated with wide orbits, these factors further impede drawing valid early results from planet searches.

In practice, systematic rather than random errors may determine the

---

[3]Horne and Baliunas remark on page 761, "clear arrow signals with period slightly longer than T can sometimes be detected, but with poor resolution."

ASTROMETRIC PLANET SEARCH

$a M = a_* M_*$ *Physics*

$\theta_* = \dfrac{a_*}{D}$ *Geometry*

$M_\odot = 1, a_\oplus = 1, D_{1pc} = 1, \theta_1 = 1$ *Normalization*

$\tau = \dfrac{a^{3/2}}{\sqrt{M_*}}$ *Kepler's Third Law*

Measurement Interpretation: $\begin{Bmatrix} \theta_* \\ \tau \end{Bmatrix}$

• Determine M_* from star's spectral class

• Determine D from the star's annual parallax
  (assume D=10pc for ~100 candidate stars)

$a = \tau^{2/3} M_*^{1/3}$

$M = \theta_* \dfrac{M_* D}{a}$
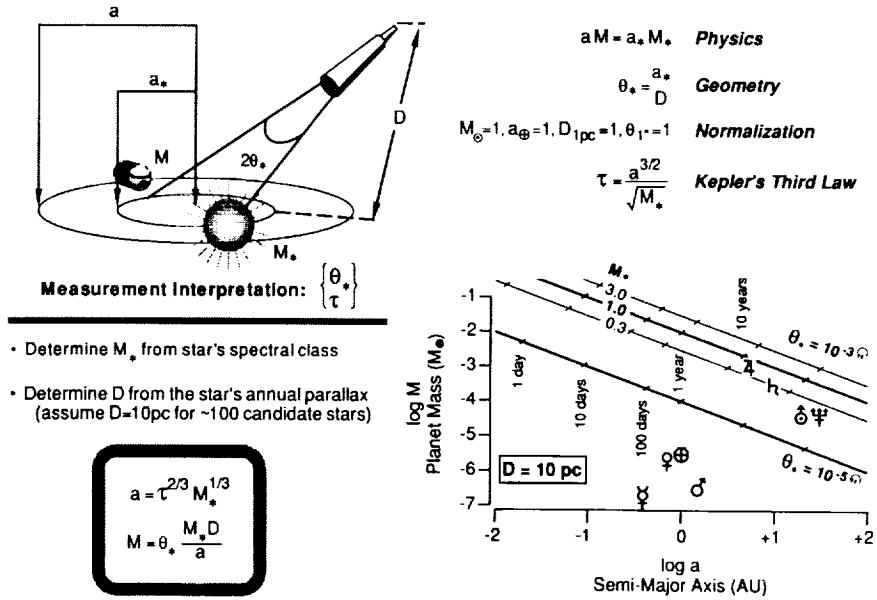
FIGURE 5 A tutorial on the astrometric technique for indirectly detecting extra-solar planets.

minimum detectable signal in reflex motion searches. The systematic errors may be real variabilities of the star (Gilliland and Baliunas 1987) or limitations of the instruments. As a class, systematic errors demand the planet searcher's assiduous attention.

Once detected, the amplitude of the reflex signal, V_* or θ_*, and its period, τ, provide specific information about the planetary mass and semi-major axis. The graduated lines in the graphs in Figures 4 and 5 signify the interpretation. The star's mass is required, and for main sequence stars, M_* can be determined adequately from the stellar-spectral type. For evolved stars, the mass assignment is more uncertain.

The radial velocity amplitude, V_*, is independent of Earth-star distance. The true orbital speed is $V_*/\sin(i)$, where the orbit's inclination angle, i, with respect to the line of sight is unknown unless it is determined separately. The average value of $\sin(i)$ in a random sample is 0.79.

The astrometric amplitude, θ_*, is a two-dimensional vector with components of right ascension and declination. When viewed from an inclined angle, a circular orbit is an apparent ellipse on the celestial sphere. In principle, the secular motion of the star along this elliptical path uniquely determines the true orbit, including the inclination angle.
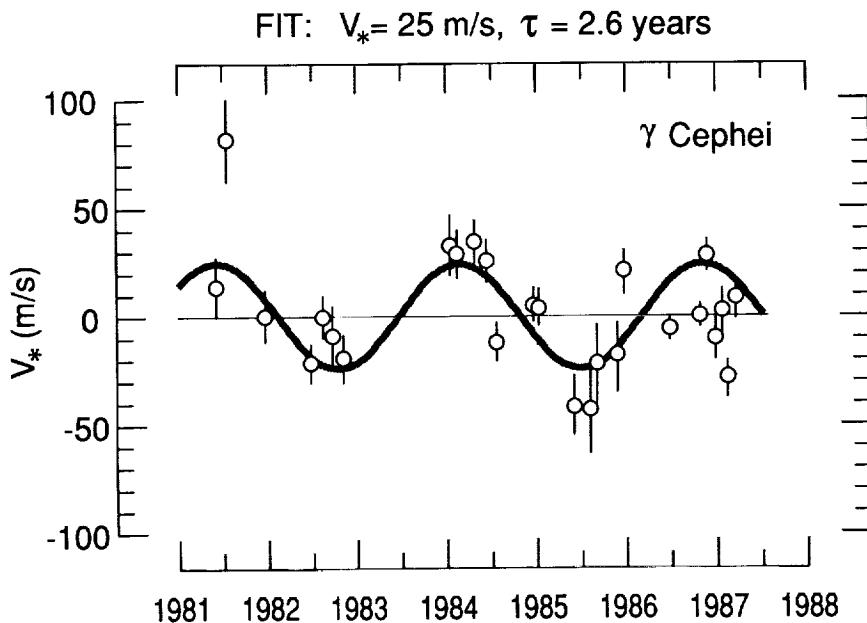
FIGURE 6 The data for the "probable" detection claimed by Campbell *et al.* of a planetary companion to γ Cephei.

Campbell *et al.* (1988) have conducted a search program that measured the radial velocities of 15 stars for six years with a precision of about 10 meters per second. The authors report statistically significant, long-term accelerations for seven stars, and in one case, they claim the "probable" detection of a period. The λ Cephei observations, with a large quadratic drift subtracted, are shown in Figure 6. The investigators have fitted a sine wave with amplitude $V_* = 25$ meters per second and period $\tau = 2.6$ years to the data.

γ Cephei is classed as spectral type K1 III-IV, indicating it has evolved far from the main sequence on the HR diagram. The rather uncertain mass estimate is $M_* = 1.15 \pm 0.1M_\odot$ Therefore, the period implies an orbital semi-major axis a = 2AU, which subtends 0.13 arc seconds at the 15pc distance of γ Cephei. Assuming the orbit is viewed edge-on, the implied mass for γ Cephei b is M = 1.3 × $10^{-3}M_\odot$.

Figure 7 shows a radial velocity detection by Latham *et al.* (1989) which has been confirmed independently by the CORAVEL program. For the low-metal, but otherwise solar-type star HD114762, the Center for Astrophysics team obtained 208 measurements with a typical precision of

FIT: V$_*$= 0.551±0.42 km/s, τ = 84.0±0.1 days
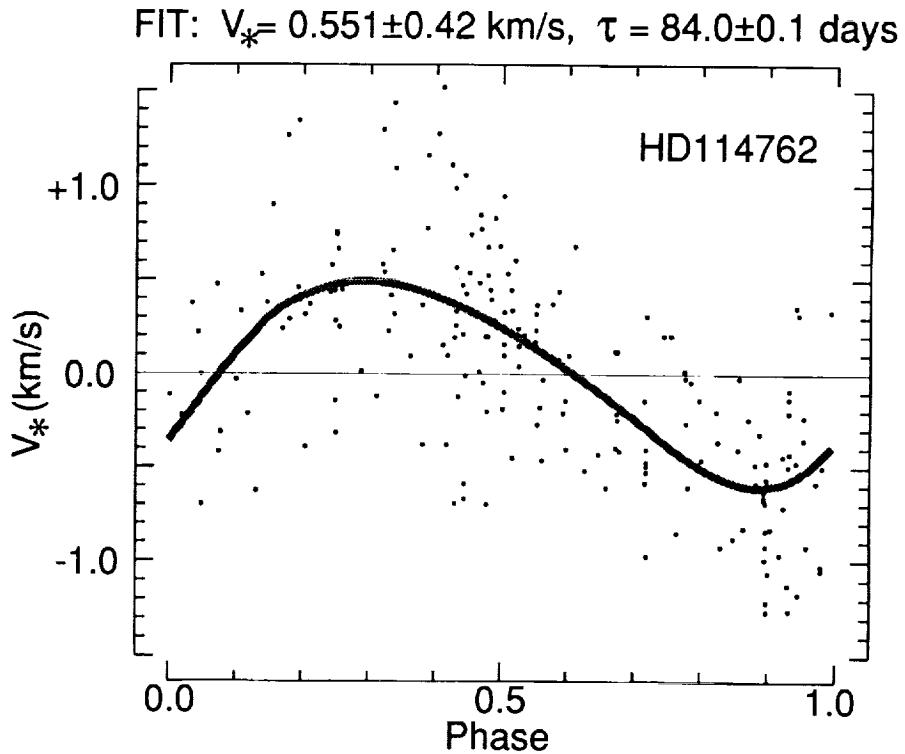
FIGURE 7   The data for the detection claimed by Latham *et al.* for a planetary companion HD114762.

0.4 kilometers per second over a 12-year interval. A periodogram analysis indicates a highly significant signal with period τ = 84 ± 0.1 days and amplitude V$_*$ = 0.551 ± 0.042 kilometers per second. The best fit is not a pure sine wave, which may indicate either an elliptical orbit or another orbiting body.

Estimating the mass of HD114762 at M$_*$ = 1M$_\odot$, the short period indicates an orbit like Mercury's: a = 0.38AU. At the 28pc distance of this system, the estimated semi-major axis subtends 0.14 arc seconds. Assuming a *single* orbit is viewed edge-on, the implied mass for HD114762 b is M = 1.1 × 10^{-2}M$_\odot$.

Van de Kamp (1986) claims the detection of two planets in his astrometric record of Barnard's star, which is shown in Figure 8. Barnard's star is late-type M dwarf, which is faint (m$_v$ = 9.5), although close (1.8pc). Van de Kamp fits his data with two amplitude-period combinations: (0.0070 arc

$$\text{FIT: } \begin{cases} \theta_*(1)=0.0070^{\circ}, \tau(1)=12\text{ yrs} \\ \theta_*(2)=0.0064^{\circ}, \tau(2)=20\text{ yrs} \end{cases}$$
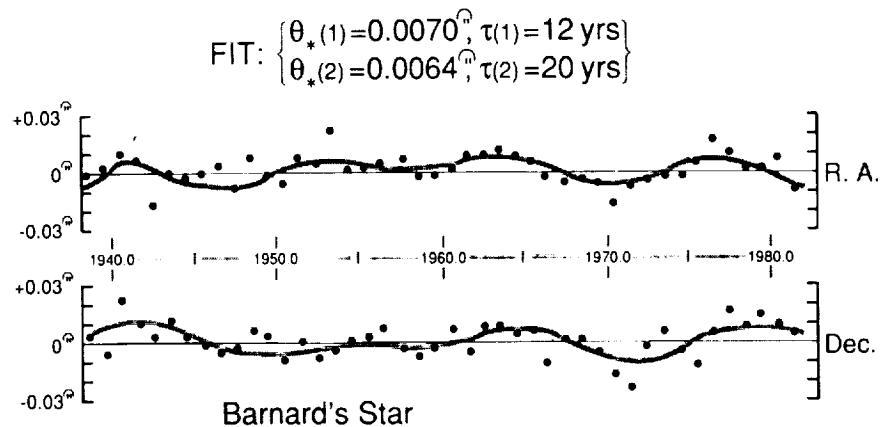

FIGURE 8 The data for the detection claimed by van de Kamp for two planetary companions to Barnard's Star.

seconds, 12 years) and (0.0064 arc seconds, 20 years). Estimating the mass of Barnard's star at $0.14M_{\odot}$, the semi-major axes are $a_b = 2.7\text{AU}$ and $a_c = 3.8\text{AU}$. (For the astrometric technique, the semi-major axis is independent of the system's inclination angle.) The masses derived for the companions are $M_b = 6.6 \times 10^{-4}M_{\odot}$ and $M_c = 4.2 \times 10^{-4}M_{\odot}$.

The van de Kamp observations extend over more than 40 years, and they have been widely discussed and disputed. The independent observations of Barnard's star shown by Harrington and Harrington (1987) are not consistent with the orbit solution by van de Kamp.

SUMMARY

Three investigator groups have reported detecting objects that are candidates for extra-solar planets according to the broad definition of the term based on mass. The findings are summarized in Figure 9.

All three claimed planet detections are based on stellar reflex motions, an indirect method. Searches based on direct imaging are currently limited to brown dwarfs because smaller objects are not sufficiently luminous to overcome scattered starlight.

Latham *et al.*'s (1989) detection of HD114762 b is solid. If the orbit is viewed edge-on, this object has a mass about 10 times that of Jupiter and is a planet by the definition adopted in this review. However, the inclination angle of the orbit is uncertain, and if sin(i) is small, the edge-on assumption would cause the actual mass of HD114762 b to be significantly underestimated. On an *a priori* basis, though, this is improbable. More radial velocity observations will clarify whether the departure of HD114762

● *HD114762 b*   ▼ *Barnard b*
■ *γ Cephei b*   ▲ *Barnard c*
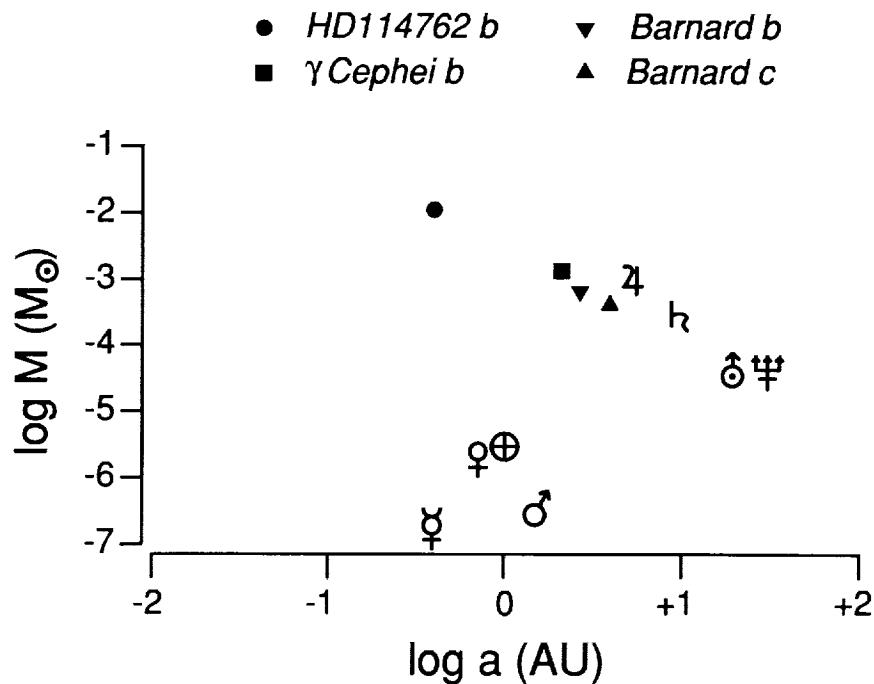
FIGURE 9 Summary of currently claimed extra-solar planet detections plotted versus orbital semi-major axis and mass.

b's radial velocity variations from a sine curve implies a non-circular orbit or a second orbiting body.

γ Cephei b and the companions to Barnard's Star are uncertain. More measurements over a longer time base are needed to confirm or deny their existence.

PROSPECTIVE CONCLUSIONS

The technological frontier for extra-solar planet detection lies in space-based systems. While the radial velocity approach is operating near the limits set by stellar atmospheric effects, the high-image quality potentially available in space will greatly benefit other search techniques (Borucki *et al.* 1988; Terrile 1988; Levy *et al.* 1988).

Regarding the future for extra-solar planet observations, Marcy and Moore (1989) offer a glimpse that is deceptively simple in a subtle way. They synthesize radial velocity, astrometric, and photometric studies of the low-mass ($0.067M_\odot \leq M \leq 0.087M_\odot$) companion of Gliese 623. These

data sets are independent and complementary. Analyzed together, the measurements reveal a rich picture of the object, with regard to the special capabilities of the individual observing techniques. As a scientific bonus, the result challenges theory: the luminosity of Gliese 623 B is significantly greater than is predicted by current models for stars of its mass and age.

Liebert and Probst (1987) have reviewed the scientific issues for low-mass stars and brown dwarfs. For those objects, the key scientific questions are about their total numbers and their intrinsic properties. The issues are not systemic, and current observing approaches address the critical questions, as Marcy and Moore show.

The scientific issues for extra-solar planets are qualitatively different from those for low-mass stars and brown dwarfs. To understand the origin and evolution of the solar system in the context of the astronomical record, *systems* of extra-solar planets must be studied as such. That means finding and understanding multiple planets per star, because only then can systemic aspects can be measured.

This review has really discussed the observational progress toward "the existence theorem:" discovering *just the first* extra-solar planets. Measuring the joint properties of multiple-planetary systems is a qualitatively more difficult challenge that will demand major technological advances. Nevertheless, this goal has durable importance for planetary science.

Someday, with much investment, work, and care, incisive observations of extra-solar planetary systems will challenge our theories and ideas about the solar system's formation and evolution.

ACKNOWLEDGEMENTS

REFERENCES

Bahcall, J.N. 1986. Brown dwarfs: conference summary. Pages 233-237. In: Kafatos, M.C., R.S. Harrington, and S.P. Maran (eds.). Astrophysics of Brown Dwarfs. Cambridge University Press, Cambridge.

Becklin, E.E., and B. Zuckerman. 1989. A low-temperature companion to a white dwarf star. Nature 336:656.

Berriman, G., and N. Reid. 1987. Observations of M dwarfs beyond 2.2 μm. Mon. Not. R. Astr. Soc. 227:315-329.

Black, D.C. 1980. In search of other planetary systems. Sp. Sci. Rev. 25:35-81.

Black, D.C. 1980. On the detection of other planetary systems: detection of intrinsic thermal background. Icarus 43:293-301

Black, D.C. 1986. Significance of brown dwarfs. Pages 139-147. In: Kafatos, M.C., R.S. Harrington, and S.P. Maran (eds.). Astrophysics of Brown Dwarfs. Cambridge University Press, Cambridge.

Black, D.C., and J.D. Scargle. 1982. On the detection of other planetary systems by astrometric techniques. Ap. J. 263:854-869.

Borucki, W.J., L.E. Allen, W.S. Taylor, A.T. Young, and A.R. Schaefer. 1988. A photometric approach to detecting earth-sized planets. Pages 107-116. In: Marx, G. (ed.). Bioastronomy—The Next Steps. Kluwer Academic Publishers, Dordrecht.

Boss, A.P. 1986. Theoretical determination of the minimum protostellar mass. Pages 206-211. In: Kafatos, M.C., R.S. Harrington, and S.P. Maran (eds.). Astrophysics of Brown Dwarfs. Cambridge University Press, Cambridge.

Brown, R.A. 1988. Systematic aspects of direct extrasolar planet detection. Pages 117-123. In: Marx, G. (ed.). Bioastronomy—The Next Steps. Kluwer Academic Publishers, Dordrecht.

Brown, R.A., and C.J. Burrows. 1990. On the feasibility of direct planet detection using Hubble space telescope. Icarus, in press.

Campbell, B., G.A.H. Walker, and S. Yang. 1988. A search for substellar companions to solar-type stars. Ap. J. 331:902-921.

D'Antona, F., and I. Mazzitelli. 1985. Evolution of very low mass stars and brown dwarfs. I. The minimum main-sequence mass and luminosity. Ap. J. 296:502-513.

Forrest, W.J., M.F. Skrutskie, and M. Shure. 1988. A possible brown dwarf companion to Gliese 569. Ap. J. 330:L119-L123.

Gilliland, R.L., and S.L. Baliunas. 1987. Objective characterization of stellar activity cycles. I. Methods and solar cycle analyses. Ap. J. 314:766-781.

Harrington, R.S., and B.J. Harrington. 1987. Barnard's star: a status report on an intriguing neighbor. Mercury, May-June 1987, 77-79, 87.

Horne, J.H., and S.L Baliunas. 1986. A prescription for period analysis of unevenly sampled time series. Ap. J. 302:757-763.

Hubbard, W.B. 1984. Planetary Interiors. 258. Van Nostrand, New York.

Latham, D.W., T. Mazeh, R.P. Stefanik, M. Mayor, and G. Burki. 1989. The unseen companion of HD114762: a probable brown dwarf? To be submitted to Nature, 339:38-40.

Levy, E.H., R.S. McMillan, G.D. Gatewood, J.W. Stein, M.W. Castelaz, A. Buffington, N. Nishioka, and J.D. Scargle. 1988. Discovery and study of planetary systems using astrometry from space. Pages 131-136. In: Marx, G. (ed.). Bioastronomy—The Next Steps. Kluwer Academic Publishers, Dordrecht.

Liebert, J., and R.G. Probst. 1987. Very low mass stars. Ann. Rev. Astron. Astrophys. 25:473-519.

Marcy, G.W., and D. Moore. 1989. The extremely low mass companion to Gliese 623. In press in Ap. J. 341:961.

McCarthy, D.W., Jr., R.G. Probst, and F.J. Low. 1985. Infrared detection of a close cool companion to Van Biesbroeck 8. Ap. J. 290:L9-L13.

Nelson, L.A., S.A. Rappaport, and P.C. Joss. 1986. The evolution of very low-mass stars. Ap. J. 311:226-240.

Nelson, L.A., S.A. Rappaport, and P.C. Joss. 1986. The evolution of very low-mass stars. Pages 177-189. In: Kafatos, M.C., R.S. Harrington, and S.P. Maran (eds.). Astrophysics of Brown Dwarfs. Cambridge University Press, Cambridge.

Perrier, C., and J.-M. Mariotti. 1987. On the binary nature of Van Biesbroeck 8. Ap. J. 312:L27-L30.

Scargle, J.D. 1982. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. Ap. J. 263:835-853.

Stringfellow, G.S. 1986. Evolution of substellar "brown" dwarfs and the evolutionary status of BV8B. Pages 190-197. In: Kafatos, M.C., R.S. Harrington, and S.P. Maran (eds.). Astrophysics of Brown Dwarfs. Cambridge University Press, Cambridge.

Strom, S.E., K.M. Strom, and S. Edwards. 1988. Energetic winds and circumstellar disks associated with low-mass young stellar objects. Pages 53-88. In: Pudritz, R.E. and M. Fich (eds.). Galactic and Extragalactic Star Formation. Kluwer Academic Publishers, Dordrecht.

Terrile, R.J. 1988. Direct imaging of extra-solar planetary systems with a low-scattered light telescope. Pages 125-130. In: Marx, G. (ed.). Bioastronomy—The Next Steps. Kluwer Academic Publishers, Dordrecht. van de Kamp, P. 1986. Dark companions of stars. Sp. Sci. Rev. 43:211-327.

van de Kamp, P. 1986. Dark companions of stars. Sp. Sci. Rev. 28:211-327.

Wetherill, G.W., and G.R. Stewart. 1989. Accumulation of a swarm of small planetesimals. In press in Icarus 77:330-357.

Zuckerman, B., and E.E. Becklin. 1987. Excess infrared radiation from a white dwarf—an orbiting brown dwarf? Nature 330:138-140.

Appendix I
List of Participants

THOMAS M. DONAHUE, Co-CHAIRMAN
University of Michigan

ROALD Z. SAGDEEV, Co-CHAIRMAN
Institute of Space Research

ALEKSANDR T. BASILEVSKIY
Vernadskiy Institute of Geochemistry and Analytical Chemistry

PETER H. BODENHEIMER
University of California, Santa Cruz

ALAN P. BOSS
Carnegie Institution of Washington

ROBERT A. BROWN
Space Telescope Science Institute

ROBERT A. GEHRZ
University of Minnesota

MIKHAIL V. GERASIMOV
Institute of Space Research

GEORGIY S. GOLITSYN
Atmospheric Physics Institute

DONALD M. HUNTEN
University of Arizona

JAMES F. KASTING
The Pennsylvania State University

VLADIMIR I. KEILIS-BOROK
Schmidt Institute of the Physics of the Earth

AVGUSTA P. LAVRUKHINA
Vernadskiy Institute of Geochemistry and Analytical Chemistry

EUGENE H. LEVY
University of Arizona

VLADISLAV M. LINKIN
Institute of Space Research

LEONID S. MAROCHNIK
Institute of Space Research

ANDREY MONIN
Shirshov Institute of Oceanology

VASILIY MOROZ
Institute of Space Research

LEV MUKHIN
Institute of Space Research

RONALD G. PRINN
Massachusetts Institute of Technology

TAMARA V. RUZMAIKINA
Schmidt Institute of the Physics of the Earth

VIKTOR S. SAFRONOV
Schmidt Institute of the Physics of the Earth

DAVID J. STEVENSON
California Institute of Technology

ANDREY V. VITIYAZEV
Schmidt Institute of the Physics of the Earth

VLADISLAV P. VOLKOV
Vernadskiy Institute of Geochemistry and Analytical Chemistry

JAMES C.G. WALKER
University of Michigan

STUART J. WEIDENSCHILLING
Planetary Sciences Institute

GEORGE W. WETHERILL
Carnegie Institution of Washington

GEORGIY M. ZASLAVSKIY
Vernadskiy Institute of Geochemistry and Analytical Chemistry

VILEN A. ZHARIKOV
Institute of Experimental Mineralogy

VLADIMIR N. ZHARKOV
Schmidt Institute of the Physics of the Earth

291

Appendix II
List of Presentations

AMERICAN PRESENTATIONS

P. Bodenheimer (University of California, Santa Cruz) "Numerical Two-Dimensional Calculations of the Formation of the Solar Nebula"

A. Boss (Carnegie Institution of Washington) "Three-Dimensional Evolution of the Early Solar Nebula"

R. Brown (Space Telescope Science Institute) "Progress in Extra-Solar Planet Detection"

R. Gehrz (University of Minnesota) "Astrophysical Dust Grains in Stars, the Interstellar Medium, and the Solar System"

J. Kasting (Pennsylvania State University) "Runaway Greenhouse Atmospheres: Applications to Earth and Venus"

E. Levy (University of Arizona) "Magnetohydrodynamic Puzzles in the Protoplanetary Nebula"

R. Pepin (University of Minnesota) "Mass Fractionation and Hydrodynamic Escape" (Presented by D. Hunten)

D. Stevenson (California Institute of Technology) "Giant Planets and Their Satellites: What are the Relationships Between Their Properties and How They Formed?"

S. Strom (University of Massachusetts, Amherst) "Astrophysical Constraints on the Time Scale for Planet Building" (Presented by T. Donahue and R. Brown)

J. Walker (University of Michigan) "Degassing"

S. Weidenschilling (Planetary Sciences Institute) "Formation of Planetesimals"

292

G. Wetherill (Carnegie Institution of Washington) "Formation of the
    Terrestrial Planets from Planetesimals"

SOVIET PRESENTATIONS

A. Basilyevskiy (Vernadskiy Institute of Geochemistry and Analytical
    Chemistry) "Venus Geology"

A. Lavrukhina (Vernadskiy Institute of Geochemistry and Analytical
    Chemistry) "Physical and Chemical Processes at Early Stages of
    Protoplanetary Nebula Evolution"

V. Linkin (Institute of Space Research) "Dynamics and Thermal Structure
    of the Venusian Atmosphere"

L. Marochnik (Institute of Space Research) "The Oort Cloud"

V. Moroz (Institute of Space Research) "Mars: Evolution of Climate and
    Atmosphere"

L. Mukhin, M. Gerasimov (Institute of Space Research) "The Role of
    Impact Processes in the Chemical Evolution of the Atmosphere of
    Primordial Earth"

T. Ruzmaikina, A. Makalkin (Schmidt Institute of the Physics of the
    Earth) "Origin and Evolution of Protoplanetary Discs"

V. Safronov (Schmidt Institute of the Physics of the Earth) "Small Bodies
    of the Solar System and the Rate of Planet Formation"

A. Vityazev, G. Pechernikova (Schmidt Institute of the Physics of the
    Earth) "The Latest Stage of Accumulation and the Initial Evolution
    of Planetary Resources"

V. Volkov (Vernadskiy Institute of Geochemistry and Analytical
    Chemistry) "Interaction Between Lithosphere and Atmosphere on
    Venus"

G. Zaslavskiy (Vernadskiy Institute of Geochemistry and Analytical
    Chemistry) "Dynamic Stochasticity of Comets Coming from the Oort
    Cloud and Numeric Simulation"

V. Zharkov, V. Solomatov (Schmidt Institute of the Physics of the Earth)
    "Thermal Regime on Venus"