Invited Comment

Planet formation and the evolution of the Solar System

M M Woolfson

University of York, Heslington, York YO10 5DD, United Kingdom

E-mail: mmw1@york.ac.uk

Received 13 November 2018, revised 12 April 2019
Accepted for publication 2 May 2019
Published 21 August 2019

Abstract
The Capture Theory gives planet formation through a tidal interaction between a condensed star and a protostar within a dense embedded stellar cluster. Initial extensive and highly eccentric planetary orbits decay and usually round-off in circumstellar material captured from the protostar. Collapsing protoplanets leave behind a circumplanetary disk within which satellites form by an accretion mechanism. Many properties of exoplanets—orbits very close to and very far from stars, highly eccentric orbits, planets around binary stars, the proportion of stars with planets and spin–orbit misalignments are explained in terms of this model. It is proposed that the initial Solar System contained six major planets, the existing four plus Bellona of mass 2.5 $M_J$ and Enyo of mass 1.9 $M_J$, where $M_J$ is Jupiter mass. The products of a collision between the two additional planets explain many features of the Solar System—the larger terrestrial planets, Mars and Mercury and their characteristics, the Earth–Moon relationship and the Moon’s surface features, the formation of asteroids, comets and dwarf planets, the formation of the Kuiper Belt and Oort Cloud, the relationship between Neptune, Pluto and Triton, the characteristics of ice giants and isotopic anomalies in meteorites. All the mechanisms involved in these processes are well understood and occur in other astronomical contexts.

Keywords: star formation, planet formation, exoplanets, terrestrial planets, small solar-system bodies, evolution of the Solar System

(Some figures may appear in colour only in the online journal)

1. Introduction

By mid-2018 nearly 4000 exoplanets—planets around distant stars—had been detected. This indicates that, for any viable theory of planet formation, the process of forming planets must be commonplace and robust—in the sense that it must operate over a wide range of parameters. The Nebula Theory (NT), a monistic theory in which the star and planets form from the same body of material, is the standard model of planet formation. The Capture Theory (CT), a dualistic theory in which the star and planets are derived from separate sources of material, has been developed in parallel with the NT, but is less well known.

The CT has been modified significantly since its first presentation (Woolfson 1964). Its temporal development has been somewhat haphazard, influenced by information from new observations as they occurred. The scenarios within which CT processes take place are supported by those observations, many of them made comparatively recently. These processes, all straightforward and involved in other astronomical contexts, are amenable to computational simulation, mostly using smoothed-particle hydrodynamics (SPH) developed by Gingold and Monaghan (1977) and Lucy (1977).

The complete theory consists of a number of steps, each causally related to previous ones. The initial process, on which all else hinges, is the way that galactic clusters, containing hundreds of stars, form in a star-forming region—e.g. the Trapezium Cluster within the Orion Nebula. However, this review will begin with the most basic ingredient in our galaxy, the interstellar medium (ISM).
2. Star formation

2.1. The formation of dense cool clouds

The ISM, which occupies the space between major bodies in our galaxy, has very low density, \( \sim 10^{-21} \text{ kg m}^{-3} \), and high temperature, \( \sim 8000 \text{ K} \). The shock wave from a supernova compresses the neighbouring ISM and injects dusty material into it. Both effects lead to cooling of the affected ISM, by radiative cooling from dust (Hayashi 1966) and the excitation of molecules, atoms and ions by free-electron collisions (Seaton 1955). The consequent fall in pressure gives an influx of ISM material, which increases the density and hence the cooling rate. Eventually a high-density, \( \sim 10^{-18} \text{ kg m}^{-3} \), low-temperature, \( \sim 20 \text{ K} \), region is produced in pressure equilibrium with the ISM. Golanski and Woolfson (2001) modelled this formation of a dense cool cloud (DCC) using SPH.

2.2. Forming stars in a dense cool cloud

If the DCC mass exceeds the Jeans critical mass (Jeans 1902) then it will begin to collapse. The collapsing cloud becomes turbulent, evidence for which comes from Doppler-shifts of maser emissions from star-forming regions (Cook 1977). Colliding turbulent gas streams generate high-density, high-temperature regions but, since cooling is much faster than re-expansion, a high-density cool region is produced that may, under suitable conditions, collapse to form a star (Woolfson 1979). If the angular momentum in the compressed region is low then a single rotationally-stable star will be produced. For a somewhat greater angular momentum the compressed region will bifurcate to form a binary system where the angular momentum is mostly taken up in the orbits of the stars rather than in stellar spin (Woolfson 2011, pp 348–50).

The Sun, and main-sequence stars of similar and lower mass, spin slowly, the equatorial speed of the Sun being 2 km s\(^{-1}\). However, original equatorial speeds may have been much higher. During the T-Tauri stage of a star’s development a strong ionized stellar wind links to stellar magnetic field lines and is carried out to several stellar radii at constant angular speed before it decouples. The gain of angular momentum by the stellar wind gives reduction of the stellar spin angular momentum, most of which may be removed (Cole and Woolfson 2013).

3. Planet formation

3.1. Interactions in a star-forming cloud

The diffuse bodies initially formed in a star-forming cloud, the collapse of which form main-sequence stars, are protostars. A typical newly-formed protostar, of mass about 0.5 \( M_\odot \), might have radius 2000 au\(^2\), density \( 10^{-14} \text{ kg m}^{-3} \) and temperature 20 K. The free-fall time to high density for such a body is \( t_f = 21000 \text{ years} \); after a time 0.8 \( t_f \) (\( \sim 17000 \text{ years} \)) its radius will fall to 1000 au.

A forming stellar cluster, immersed in gas, is in an embedded state—its contents are embedded in gas. As stars form, the cloud, with contained stars, continues to collapse so that the stellar number density (SND) increases. Stellar radiation slowly expels gas from the cloud but, after a few million years, massive-star supernovae greatly increase the rate of gas expulsion. Released from the gravitational influence of the gas the stellar cluster begins to expand—in 90\% of cases indefinitely to give field (isolated) stars and field binary systems. In the other 10\% of cases a galactic stellar cluster forms, typically containing a few hundred stars.

There is a period of about 5 million years, with gas present, where the SND is very high and the cluster is then in a dense embedded state (DES). For the Trapezium Cluster the core SND has been observed to be \( 4.65 \times 10^4 \text{ pc}^{-3} \) (McCaugrean and Stauffer 1994\(^3\)). A simulation of the evolution of a star-forming cloud by Bonnell \textit{et al} (2003), using high-definition SPH, showed that in the last stages of the collapse, of duration about 5 million years, the cloud fragmented, each fragment containing tens of stars within which the SND was about \( 2 \times 10^5 \text{ pc}^{-3} \), although the whole-cloud average SND was two orders of magnitude less. Subsequently the fragments expanded and combined to form larger fragments that moved closer together; the peak fragment SND decreased but the whole-cloud average increased. Finally there were 400 stars in five very close fragments so that the maximum fragment SND, somewhat greater than \( 2 \times 10^5 \text{ pc}^{-3} \), is only slightly larger than the whole-cloud SND peaking at \( 2 \times 10^4 \text{ pc}^{-3} \).

The average stellar speed in a dense embedded cluster is about 1 km s\(^{-1}\) (Gaidos 1995) so in 17000 years, while still an extended object, a protostar can travel more than 3000 au. For an SND of \( 2 \times 10^4 \text{ pc}^{-3} \) the average interstellar distance is just over 8000 au. From protostar dimensions, the distances they travel and the SNDs that occur it is clear that close approaches of extended protostars and condensed stars will take place. We now consider a tidal interaction between an extended protostar and a condensed star. The frequency of such interactions is considered quantitatively in section 3.3.

3.2. Capture-theory simulations

Jeans (1917) advanced a theory—the standard theory for two decades—that the Solar System formed from the close passage of a massive star past the Sun. A filament of matter drawn out of the Sun was gravitationally unstable and broke up into a string of blobs that eventually condensed to form planets. The blobs were attracted by the retreating massive star and were left in orbit around the Sun.

Similar mechanisms occur in the CT but in a completely different context. Figure 1 shows a CT simulation using SPH with radiation transfer (Oxley and Woolfson 2003, 2004)

The parameters for the simulation were as follows:

\(^3\) A parsec (pc) is 3.26 light years or \( 3.086 \times 10^{13} \text{ km} \).
3.2.1. Star characteristics. Mass, $M = 2 \times 10^{30}$ kg $\approx M_{\odot}$
Luminosity, $L = 4 \times 10^{26}$ W $\approx L_{\odot}$

3.2.2. Protostar characteristics. Mass, $M_P = 7 \times 10^{29}$ kg $\approx 0.35 M_{\odot}$
Initial radius, $R_P = 800$ au
Initial temperature, $T_P = 20$ K
Mean molecular mass of material, $\mu = 4 \times 10^{-27}$ kg.

3.2.3. Characteristics of the protostar orbit. Initial distance of centre of protostar from star, $D = 1600$ au.

![Image 1](1000 au)

Time = 0

![Image 2](1000 au)

Time = 6 000 years

![Image 3](12 000 years)

Time = 12 000 years

![Image 4](18 000 years)

Time = 18 000 years

**Figure 1.** An SPH simulation, with radiation transfer, of the interaction of a star and a protostar (Oxley and Woolfson 2004).

Figure 2. The collapse of a protoplanet at 100 year intervals (Oxley and Woolfson 2004).

Planets can also form if turbulent gas streams collide close to a condensed star. In figure 3 each stream has mass $0.5 M_{\odot}$, density $4 \times 10^{-15}$ kg m$^{-3}$ and speed 1 km s$^{-1}$. The condensations A, B and C, with masses $1.0 M_J$, $1.6 M_J$ and $0.75 M_J$ respectively, are captured. The CT mechanism is very robust and has been shown to give planet formation at scales from one-tenth to ten times the scale used in figures 1 and 3.

While the planet masses are within the range found for exoplanets the initial orbits do not correspond to what is
found for exoplanets and solar-system planets. For the eight captured protoplanets in figures 1 and 3 the semi-major axes, \(a\), and eccentricities, \(e\), are as follows:

\[
(a, e) = (1, 247 \text{ au}, 0.835) \\
(a, e) = (1, 885 \text{ au}, 0.7725) \\
(a, e) = (1, 509 \text{ au}, 0.765) \\
(a, e) = (1, 325 \text{ au}, 0.736) \\
(a, e) = (2, 686 \text{ au}, 0.902) \\
(a, e) = (4, 867 \text{ au}, 0.768) \\
(a, e) = (1, 703 \text{ au}, 0.381) \\
(a, e) = (1, 736 \text{ au}, 0.818).
\]

These orbits are more extensive than any found for exoplanets and their eccentricities are large although some exoplanet orbits do have large eccentricities. This matter is addressed in section 4.4.

The CT process of planet formation is referred to as ‘top-down’ in which major planets are produced directly from an above-Jeans-critical-mass of dusty gas—the condensations seen in figures 1–3. To explain smaller bodies one or more of the initial planets must break up, as described in section 6.2. From figures 1 and 3 the timescale for CT planet formation is of order \(10^4\) years, much less than the duration of a DES.

3.3. The proportion of stars with planets

Observationally-based estimates of the proportion of stars with planets vary with time and tend to increase. Borucki et al. (2011) have given an estimate of 0.34 although there are some higher estimates.

Interactions between protostars and condensed stars would occur frequently in the dense embedded state of a star-forming cloud and Woolfson (2016) developed a model to quantify the proportion of stars with planets. The model was investigated for SNDs of \(n = 5000 \text{ pc}^{-3}\) to \(25 000 \text{ pc}^{-3}\) in steps of \(5000 \text{ pc}^{-3}\) and initial protostar radii \(R_P = 1000, 1500\) and \(2000 \text{ au}\) with free-fall times of \(t_{ff} = 10 200, 18 800\) and \(28 900\) years, respectively, for a protostar of mass \(0.3 \text{ M}_\odot\). The motion of the collapsing protostar among the stars is followed for a maximum period \(0.8 t_{ff}\), at which stage it has about one half of its original radius.

A random configuration of \(N_S\) stars in a cubical cell is surrounded by 26 ‘ghost cells’, each containing a similar configuration, to form a \(3 \times 3 \times 3\) array. A star leaving the central cell re-enters on the opposite face, so keeping the density in the central cell constant. The cell size, \(a\), is set to give the required SND and stars in the ghost cells are only included out to a distance \(ma\) from the centre of the central cell, so forming a spherical cluster containing, on average,
The protostar’s motion is affected by the stars plus the residual gas in the fragment. The gas has two effects—firstly, gravitational and, secondly, by exerting a drag on bodies so reducing their speeds.

Experience with CT calculations show that if the closest approach, $r_c$, of the protostar orbit to a solar-mass star was between 0.5 and 1.5 of the protostar radius then a CT event was almost inevitable. This condition is little affected by the mass of the protostar as long as it is not too extreme. The CT is based on tidal effects, where tidal distortions are produced by differential gravitational forces on different parts of a body. The tidal effect at a distance is proportional to $M/R^3$, where $M$ is the mass of the tide-raising body and $R$ the distance of the body on which the tidal force acts. Thus for a star of mass $M_s$ the distance condition that a CT event will occur is different from that for a solar-mass star by a factor $(M_0/M_s)^{1/3}$, giving the condition

$$0.5 \left( \frac{M}{M_s} \right)^{1/3} R_p \leq r_c \leq 1.5 \left( \frac{M}{M_s} \right)^{1/3} R_p,$$

where $R_p$ is the current radius of the collapsing protostar.

The masses of the stars are chosen randomly from a distribution with mass index $\sim 2.3$. (Kroupa 2001), i.e. the proportion of stars per unit mass range is given by

$$f(M) \propto M^{-2.3},$$

with masses in the range 0.5–3.0 $M_\odot$, all greater than the protostar mass. The average mass of the stars so selected is 1.00 $M_\odot$.

For gravitationally-interacting stars in a bound region in equilibrium, and with no other forces acting, the virial theorem would be valid. This links the translational kinetic energy, $K$, of the stars to their total gravitational potential energy, $\Omega$, by

$$K = -0.5 \Omega.$$  \hspace{1cm} (4)

However, due to gas-dynamical friction, the stars in a dense embedded cluster will have sub-virial energy (Indulekha 2013) so that

$$K = -\beta \Omega$$  \hspace{1cm} (5)

with $\beta < 0.5$. Protskovz et al (2009) have considered $\beta$ in the range 0.04–0.15, corresponding to root-mean-square speeds between 28% and 55% of the virial value. Stars, and the protostar, were given speeds corresponding to the equipartition value in randomly-chosen directions. The total kinetic energy of the $N_T$ stars is given by (5) and hence, applying the equipartition theorem, the $i$th star will have kinetic energy

$$\frac{1}{2} m_i V_i^2 = -\frac{1}{N_T} \beta \Omega,$$

from which we find

$$V_i = \left( -\frac{2\beta \Omega}{N_T M_i} \right)^{1/2}.$$  \hspace{1cm} (6)

The potential energy, $\Omega$, was computed on the basis that the mass of the $N_T$ stars was uniformly distributed within the sphere of radius $ma$.

The speeds assigned to stars of the extreme masses, 0.5 $M_\odot$ and 3.0 $M_\odot$ as a function of the SND are, with $\beta = 0.04$,

For $n = 5, 000$ pc$^{-3}$ and $M_i = 0.5 M_\odot$, $V_i = 483$ m s$^{-1}$

For $n = 5, 000$ pc$^{-3}$ and $M_i = 3.0 M_\odot$, $V_i = 198$ m s$^{-1}$

For $n = 25, 000$ pc$^{-3}$ and $M_i = 0.5 M_\odot$, $V_i = 632$ m s$^{-1}$

For $n = 25, 000$ pc$^{-3}$ and $M_i = 3.0 M_\odot$, $V_i = 258$ m s$^{-1}$.

For comparison, Gaidos (1995) gave stellar speeds in a DEC in the range 500–2000 m s$^{-1}$, corresponding to a larger value of $\beta$ and allowing larger values of $n$.

The equations of motion are numerically solved for the $N_S$ stars plus protostar in the basic cell. The gravitational effects of ghost-cell stars are taken into account but they are not moved during each integration step. The gravitational acceleration due to the gas, of mass $M_G$, within the fragment on a star at vector position $r$ was included and is given by

$$a_G = \frac{G M_G}{(m a)^3} r.$$  \hspace{1cm} (7)

The integration was carried out using the 4-step Runge–Kutta method. If the motion of the protostar gives a periastron distance from a star between 0.5 and 1.5 of its current radius (which changes with time) then planet formation is deemed to have occurred but if it moves closer then it is taken that it is disrupted without planet formation. If the planet formation or protostar disruption stage is reached then the trial is terminated and the next one begun, otherwise the calculation is terminated at time $0.8 t_{ff}$. At the end of each integration step any stars, including the protostar, that have left the basic cell are reintroduced into the central cell as previously described and the ghost cells and the cluster fragment are redefined. For four sets of conditions, with $\beta = 0.04$, 1000 Monte Carlo trials were run. The sets were:

A $N_S = 8$, $m = 1.5$, giving $N_T = 113$

B $N_S = 23$, $m = 1.0$, giving $N_T = 96$

C $N_S = 10$, $m = 1.0$, giving $N_T = 42$.

All these sets had the ratio of mass of gas to stars equal to 1. Set D is as for C but with the mass ratio equal to 3. The percentages of capture events are given in table 1; these results are somewhat conservative because:

(a) Capture-theory events involving the collision of high-density regions (figure 3) have not been included.

(b) The SNDs have been considered up to $2.5 \times 10^4$ pc$^{-3}$. Maximum values up to $10^6$ pc$^{-3}$, have been suggested by some authors (McCaughrean and Stauffer 1994).

(c) The maximum radius of a protostar has been taken as 2000 au; some authors have suggested larger values (McCaughrean and Stauffer 1994).

(d) The value taken for $-\beta = -0.04$ gives root-mean-square speeds less than those indicated by Gaidos (1995). Increasing $\beta$ gives more stars with planets.

\hspace{1cm} 4 The periastron is the closest position of an orbit to a star and apastron the farest position.
The initial protoplanet orbits are very extended so that initial planetary systems can be disrupted by stellar perturbation before orbits have completely evolved. An analysis by Woolfson (2004b) indicated that up to 20% of planetary systems can be lost in this way but, nevertheless, the CT can explain the estimated proportion of stars with planets.

The present analysis does not give the proportion of stars with planets but rather the proportion of protostars giving planets. The following example gives an idea of how to transform from the latter quantity to the former. If 100 protostars give 25 sets of planets and 10 are disrupted then we have added 25 planetary systems (to pre-existing stars) and 65 new stars to the cluster. The ratio of the number of planetary systems produced to the number of stars added is then $0.25/0.65 = 0.38$.

### 3.4. Observation of a forming planet

In July 2018 the image shown in figure 4 was published showing a forming planet around the young star PGS 70, age 4.5 million years. It was interpreted as a forming planet in a gap within a circumstellar disk. However, it could equally represent an annular circumstellar ring produced by the CT (figure 5) with a planet at about 30 au from the star but with unknown motion. The light from the star has been blocked out and the ring is about 120 au in diameter. This image is completely compatible with a CT origin.

There are features of the image that are difficult to interpret—in particular two bridges linking the ring and inner material, the upper bridge possibly having small condensations within it.

### 4. Orbit modification

#### 4.1. Angular momentum in the Solar System

The Sun, with 99.87% of the mass of the Solar System has only 0.5% of its angular momentum in its spin, the remainder being in the planetary orbits. For the CT, a dualistic theory, the slow rotation of the Sun comes from the process described in section 2.2 and the orbital angular momentum from the protostar-Sun orbit. Indeed, the initial orbits from the simulations contain far too much angular momentum and we must explain how orbits evolve to their present states.

#### 4.2. Circumstellar disks

The CT simulations show that some protostar material forms a circumstellar disk, usually of mass 25–50 $M_J$ and extending out to several hundred au. The areal density often falls off

---

**Table 1. Percentage of CT interactions giving planets with various $R_p$ and $n$.**

| $R_p$ (au) | $n$ | Set A capture (%) | Set B capture (%) | Set C capture (%) | Average capture (%) | Set D capture (%) |
|-----------|-----|-------------------|-------------------|-------------------|--------------------|-------------------|
| 1000      | 5000| 0.7               | 0.6               | 0.3               | 0.5                | 0.3               |
| 10 000    | 1.8 | 1.6               | 1.0               | 1.5               | 0.8                |
| 15 000    | 3.4 | 2.1               | 1.8               | 2.4               | 1.9                |
| 20 000    | 3.8 | 3.3               | 2.9               | 3.3               | 3.2                |
| 25 000    | 5.3 | 4.5               | 4.6               | 4.8               | 5.3                |
| 1500      | 5000| 2.3               | 2.3               | 2.7               | 2.4                | 2.4               |
| 10 000    | 6.5 | 5.6               | 5.9               | 6.0               | 5.5                |
| 15 000    | 11.6| 8.9               | 10.2              | 10.2              | 11.0               |
| 20 000    | 11.5| 13.5              | 11.9              | 12.3              | 12.9               |
| 25 000    | 15.9| 16.3              | 17.1              | 16.4              | 19.5               |
| 2000      | 5000| 7.9               | 6.6               | 5.0               | 6.5                | 5.2               |
| 10 000    | 17.7| 16.9              | 16.7              | 17.1              | 17.0               |
| 15 000    | 29.0| 25.9              | 24.7              | 26.5              | 29.1               |
| 20 000    | 36.4| 37.8              | 33.6              | 35.9              | 40.8               |
| 25 000    | 48.0| 46.0              | 44.2              | 46.1              | 52.4               |

---

**Figure 4.** Young planet close to a young star (ESO; Max Planck Institute for Astronomy).

**Figure 5.** A capture-theory simulation showing a strong doughnut-like captured medium (Woolfson 2003).
outwards monotonically but sometimes can be of doughnut form with maximum density a few hundred au from the star. In figure 5, where the disk points have been enhanced for clarity, the peak areal density is about 350 au from the star. The word ‘disk’ conjures up an image of something like a coin, an object of circular cross-section and uniform thickness, but a gaseous disk would not be stable in that configuration. The areal density falls off to zero at the edge of the disk but, for equilibrium, in cross-section it has to flare out in a direction perpendicular to the disk mean-plane.

Woolfson (2003) studied the evolution of planetary orbits, modelling the medium as shown in figure 6. The points were most densely packed close to the star and the masses of individual points were fixed to give the required density distribution. The medium points did not act gravitationally with each other and were in Keplerian orbits so, in the absence of a planet, the medium was stable. Observations of circumstellar disks show that they have typical lifetimes of three million years but up to ten million years. For that reason, as the simulation is run, the mass of each medium points declines as

$$m(t) = m(0)\exp(-\gamma t),$$  \hspace{1cm} (8)$$

where $m(t)$ is the mass of a point at time $t$ and the value of $\gamma$ controls the rate of loss of the medium.

The areal density of the medium, corresponding to a doughnut form, is given by

$$\rho(r) = C \exp\{-\alpha^2(r - r_p)^2\},$$  \hspace{1cm} (9)$$

where $C$ is adjusted to give the total mass of the medium, and this areal mass is spread perpendicular to the disk uniformly between $W$ and $-W$ where

$$W = \frac{\pi c}{2\omega^2}$$  \hspace{1cm} (10)$$

in which $c$ is the sound speed in the medium and $\omega$ is the local Keplerian angular velocity.

### 4.3. Types of Migration Mechanism

The NT community describes the evolution of planetary orbits as migration and there are two main types. Type I migration is when the planet retains contact with the medium within which it moves. It operates on a timescale somewhat less than the normal lifetime of a circumstellar disk and is appropriate for planets of mass less than somewhere between the masses of Saturn and Jupiter (Lubow and Ida 2011).

More massive planets add angular momentum to exterior medium material and subtracts it from interior medium material, thus opening up a gap within which it moves. It is less effective than Type I migration but still quite effective (D’Angelo and Lubow 2010). There is a smooth transition between the Type I and Type II mechanisms.

### 4.4. The Evolution of Planetary Orbits

To investigate orbital evolution a planet is inserted into the model resisting medium with an initial orbit that a CT simulation might give and its motion followed computationally. Figure 7 shows the result of one calculation. The parameters were:

- Planet mass $= 4 M_J$, initial semi-major axis $= 1500$ au; initial eccentricity $= 0.9$; mean molecular mass for medium $= 2 \times 10^{-27}$ kg; medium temperature $= 20$ K; total medium mass $= 50 M_J$, star mass $M_*; r$, the distance from the star in au, $\alpha = 0.007587$ au$^{-1}; r_p = 200$ au (for $\alpha$ and $r_p$, see equation (9)) with time in years $\gamma = 10^{-6}$ yr$^{-1}$; medium represented by 77408 particles.

After 1.5 million years the orbit rounded-off and after eight million years decayed to a semi-major axis of about 2.4 au. Calculations revealed some obvious characteristics. For example, a more massive or a longer-lasting medium makes round-off and decay faster and increases the total orbital decay, which tends to a constant final value as the resisting medium disappears. As the orbit evolves the perihelion, $q$, steadily increases until round-off after which the circular orbit continues to decay. A less intuitive result is that decay increases with the mass of the planet. In table 2 the initial orbital parameters were $(a, e) = (1500$ au, 0.90) and the medium of Gaussian form as given by (9).

For some simulations the final semi-major axis is so small that the conclusion is that the planet would plunge into the star. However, observed exoplanets have semi-major axes down to 0.015 au and there exists a mechanism that can maintain that order of distance even while the medium is still present. It is the same mechanism that causes the Moon to retreat slowly from Earth and depends on the central-body spin period being less than the secondary-body orbital period. The tide raised on the central body by the orbiting body is dragged forward by the central-body spin. The near tidal
bulge on the central body gravitationally pulls the orbiter in the direction of its motion thus increasing its orbital angular momentum, which moves it outward. In the present context, when the energy gain by the tidal interaction equals the energy loss due to the medium resistance then the orbit stabilizes.

With a low-mass, diffuse or short-lived medium the final semi-major axis may be very large. Some exoplanets, which can be directly imaged, are at large distances from their stars, the largest known being a planet of mass 11 \( M_J \) at 650 au from the star HD 106906 (Bailey et al 2014). Speculation concerning a possible solar-system Planet 9 with estimated \((a, e) = (700 \text{ au}, 0.6)\) could correspond to a protoplanet produced in an extremely extended heliocentric orbit that only partially decayed and rounded-off.

Although orbital evolution occurs throughout a planet’s orbit the general pattern of the evolution can be determined by just considering what happens at periastron and apastron. At periastron the planet moves faster than the medium and slows down. This keeps the periastron constant but reduces the apastron, thus reducing the eccentricity. At apastron the planet moves slower than the medium and speeds up. This keeps the apastron constant but increases the periastron thus reducing the eccentricity. At both extremes the eccentricity decreases and if, as commonly occurs, the periastron effect is stronger then the orbit will also decay. However, there are some exoplanet orbits of high eccentricity, up to 0.97, and we must consider how these could occur.

Table 2. The variation of the final \( a \) and \( e \) with protoplanet mass for a given medium.

| Mass of planet  | Semi-major axis (au) | Eccentricity |
|-----------------|----------------------|--------------|
| \( M_J \)       | 2.980                | 0.0067       |
| \( 2M_J \)      | 1.549                | 0.0065       |
| \( 3M_J \)      | 1.056                | 0.0064       |
| \( 4M_J \)      | 0.787                | 0.0064       |
| \( 5M_J \)      | 0.633                | 0.0062       |

Young stars generate strong stellar winds which exert an outward force on the medium so effectively reducing the star’s gravitational influence on the medium. This causes the medium to orbit more slowly as though the star’s mass was reduced. At periastron the planet still moves faster than the medium and so the apastron and eccentricity are reduced. With the slower medium the planet can now move faster than the medium at apastron and is slowed down, the effect being that the periastron is reduced and the eccentricity increased. For a medium with the doughnut structure shown in figure 5 the orbit would initially decay and round-off but once the apastron reached the peak density the effect there, and going inwards from the peak, would dominate and the eccentricity thereafter increase.

Simulations for a planet of mass \( 5M_J \) with initial orbit \((a, e) = (1500 \text{ au}, 0.9)\) and a medium mass \( 50M_J \), with the doughnut form given by equation (9), were run with different diminutions of the effective stellar mass for the medium. For effective stellar mass greater than 0.5 of the true mass the final orbits are circular but for lesser values the final eccentricity steadily increases as the effective stellar mass decreases. Figure 8 shows the variation of semi-major axis and eccentricity for three runs with different effective masses.

### 4.5. Exoplanets around binary stars

Exoplanets have been detected orbiting one or both stars of a binary system. The stars of the system \( \gamma \)-Cephei have masses 1.40 \( M_\odot \) and 0.40 \( M_\odot \) with orbital parameters \((a, e) = (20.2 \text{ au}, 0.41)\). An exoplanet orbits the more massive star with orbital parameters \((a, e) = (2.04 \text{ au}, 0.115)\) (Torres 2007). For the CT, with filament distance from the star usually several hundred au, a close-binary pair would act like a single gravitational centre of slightly varying strength and direction and the CT process would resemble that for a single star. The medium in the vicinity of the binary system would be disturbed by the stellat motions but, nevertheless the planet orbit would decay. As the planet approaches the stars one of at least three things could happen. It could be left in orbit around both stars, acquire enough energy to escape from the binary system or be captured by one of the stars.

### 4.6. Commensurabilities of planetary orbits

Many stars, including the Sun, have several accompanying planets. When orbits round-off and decay, they are influenced
by the star, the medium and each other. The ratios of the orbital periods of pairs of the major solar-system planets are close to the ratio of small integers, e.g.

\[
\frac{\text{Orbital period of Saturn}}{\text{Orbital period of Jupiter}} = \frac{29.46 \text{ years}}{11.86 \text{ years}} = 2.48 \approx \frac{5}{2}
\]

and

\[
\frac{\text{Orbital period of Neptune}}{\text{Orbital period of Uranus}} = \frac{164.8 \text{ years}}{84.02 \text{ years}} = 1.96 \approx \frac{2}{1}.
\]

Melita and Woolfson (1996) explained how these orbital commensurabilities became established. For one computational test Jupiter–mass and Saturn–mass bodies were placed in orbit within a resisting medium with the ratio of their orbital periods 2.5 but further out from the Sun than the actual planets. The initial eccentricities and inclinations were the same for both orbits—0.1 and 0.06 radians (3.4°). Three computational runs were made with different initial relative positions. Both semi-major axes and inclinations fell monotonically but the eccentricities fell at first but then rose again. However, as seen in figure 9, the ratio of the two periods departed from 2.5 and settled down close to 2.0, actually oscillating about 2.02, although both orbits were still decaying.

It is easier to explain these computational results than to predict them in advance. A key phenomenon that influences the behaviour of the system is that when two bodies are in commensurate orbits the outer body removes energy from the inner one and, conversely, the inner body adds energy to the outer one.

Initially the eccentricities fall quickly, which reduces the speeds of the planets relative to the resisting medium. This reduces the rate of energy dissipation and hence the rate of change of orbital parameters. Although Saturn is less massive than Jupiter its orbit decays more rapidly because it is due to Type I migration rather than the Type II that governs the decay of Jupiter’s orbit. When commensurability is established the bodies come closest together repeatedly at the same points of their orbits. This resonance effect amplifies the perturbations and increases the eccentricity which, in its turn, increases the rate of dissipation and hence the rate of decline in the semi-major axis. The increasing dissipation due to the increasing eccentricity of Saturn’s orbit is balanced by a gain of energy from Jupiter due to the 2:1 resonance. Thus the rate of change of Saturn’s semi-major axis is reduced while that of Jupiter is increased because the effect of Saturn as an exterior body is to add to the loss of energy due to dissipation. The resonance state is now lost and once again Saturn decays more rapidly until resonance is re-established. The result of this is that the ratio of the periods, which had fallen from 5:2 to 2.1, now becomes locked at close to 2.1, with a slight variation, although both orbits still decay.

The numerical experiments were repeated for the Jupiter–Saturn system with different initial eccentricities—nine combinations with each eccentricity having the values 0.1, 0.2 and 0.3—but always with an initial orbital period ratio 2.5. The results are shown in figure 10. Five combinations end up with a ratio close to 2.0 and four stayed at 2.5. Other trials gave a Neptune–Uranus system with a ratio close to 2.0 starting with the present observed ratio of 1.96 (figure 11).
The ratio of periods for the neighbouring planets Uranus and Saturn is 2.85. It is possible that it was evolving towards a ratio of 3.0, which would have given commensurabilities linking all four major planets, but that the medium dissipated before resonance was established.

4.7. The inclinations of exoplanet orbits

In 2009 the transiting exoplanet WASP-17b was found to be in a retrograde orbit around its star (Anderson et al 2010). Many more retrograde orbits for transiting planets have been found from Kepler-mission observations using the Rossiter–McLaughlin effect (McLaughlin 1924, Rossiter 1924), which involves observing the change in the peak wavelength of the emitted light as the transit progresses. Another method depends on the measurement of blips on the transit intensity curve due to star spots (Nutzman et al 2011). The figure 12 histogram is derived from spin–orbit misalignment (SOM) — i.e. inclination of the planetary orbit from the star’s equator — observations given by René Heller in 2013. Retrograde orbits (SOM > 90°) are common, although prograde orbits, especially with small inclinations, are predominant, including planets of the Solar System with average SOM 7°.

For the CT there is a random relationship between the initial spin axis of the star and the plane of the star-protostar orbit, which defines the plane of the planetary orbits. Thus all SOMs should be possible and, assuming random inclinations of the initial protostar orbit, the probability of having an SOM is proportional to sin(i), something that figure 12 shows is not true.

The CT mechanism not only produces protoplanets and a resisting medium but also adds protostar material to the star, the angular momentum of which pulls the spin axis of the star towards the normal to the exoplanet orbits so reducing the SOM. Some added material comes directly from the star; other added material — perhaps the majority — comes from the circumstellar disk, the inner part of which will drift inwards (Lynden-Bell and Pringle 1974). Comparatively little absorbed material has a big effect; one-third of a Jupiter mass in orbit at the solar equator has as much angular momentum as the Sun in its spin.

The final SOM depends on the initial value, due to the star-protostar orbit, and the mass of absorbed protostar material. The latter could often be up to a few Jupiter masses so that the likelihood of a small SOM, as shown by figure 12 — could be high. The CT mechanism leads to a complete range of possible spin–orbit misalignments but with a strong bias towards small values (Woolfson 2013a).

5. Satellite formation

5.1. Satellites and angular momentum

When Galileo saw the large Jupiter satellites through his telescope he interpreted it as a small-scale version of the Solar System. For him, and many that followed, it became axiomatic that the mechanism for producing satellites should be a small-scale version of that for producing planets.

The distribution of angular momentum is important in considering solar-system origin and here we compare the planetary system and satellite systems in terms of angular momentum distribution. We find the following ratio with respect to a number of primary and secondary pairs of bodies.
Secondary orbit to that of the spin of the central body at its equator. System planetary system. Does not dominate to the extent that it does for the solar-system planetary system. Particular the orbital angular momentum of satellite systems does not dominate to the extent that it does for the solar system planetary system. This is an indication that is not unreasonable to consider distinction between the two systems even without this effect. Accepted then timescales would be reduced. For Jupiter the total mass of its Galilean satellites is about 4 × 10^21 kg. For the disk we have specified, the total mass beyond 4 × 10^6 km is over 8 × 10^26 kg and if just 0.5% of that is deposited dust then the carpet mass would be 4 × 10^24 kg—an ample source of solid material to form the satellites. If larger dust particles were accepted then timescales would be reduced. 5.2. A mechanism for satellite formation Figure 2 shows that, as a protoplanet collapses, a disk is left behind of mass comparable to that of the collapsing core. Woolfson (2004a) described a process of satellite formation in the circumplanetary disk that exactly parallels that proposed for planet formation for the NT. The steps are: (i) Dust settles towards the mean plane of the disk to form a dense dust carpet. (ii) The dust carpet is gravitationally unstable and fragments to produce solid bodies—satellitesimals. (iii) The satellitesimals aggregate to form satellites. For the NT there is a fourth step for major planets, which is the capture of an atmosphere from the gas in the disk. Some solar-system satellites have atmospheres so this step may sometimes occur. 5.2.1. Dust settling. Although the dust particles most easily detected in molecular clouds are of submicron size, there is a distribution of sizes

\[ n(D) = KD^{-3.5}, \]

where \( n(D) \) is the number density of particles with diameter \( D \) and \( K \) is a constant. Recent work suggests that particle diameters up to 5 \( \mu m \) are present and Wood et al (2001) detected dust particles up to 50 \( \mu m \) in size in the dust disk surrounding a T-Tauri star. More recently, millimetre size grains have been detected in some disks. We accept an upper limit of 5 \( \mu m \) and almost one-half of the mass of dust is then contained in particles between 2 and 5 \( \mu m \). Larger particles settle faster towards the mean plane and they grow by sweeping up smaller particles. The disk will have a flared structure (figure 6(b)) and an areal density decreasing outwards. Particle closer in have less far to fall but do so in a denser medium that offers more resistance to their motion. Settling times at different distances from a typical planet are shown in figure 13, using theory developed by Weidenschilling et al (1989) for the following set of parameters

- Planet mass = 2.0 × 10^{27} kg (approximately the mass of Jupiter)
- Disk mass = 2.0 × 10^{27} kg
- Areal density fall by factor \( e \) every 10^8 km distance from planet
- Mean molecular mass of gas = 4 × 10^{-27} kg
- Temperature = 20 K
- Density of dust grains = 3 × 10^3 kg m^{-3}
- Ratio of principle specific heats of gas = 5/3.Beyond 4.0 × 10^6 km, for a disk lifetime of three million years, settling would be complete. For Jupiter the total mass of its large Galilean satellites is about 4 × 10^{21} kg. For the disk we have specified, the total mass beyond 4 × 10^6 km is over 8 × 10^{26} kg and if just 0.5% of that is deposited dust then the carpet mass would be 4 × 10^{24} kg—an ample source of solid material to form the satellites. If larger dust particles were accepted then timescales would be reduced. 5.2.2. Formation of satellitesimals. The theory for dust-carpet gravitational instability was given by Goldreich and Ward (1973). A necessary condition for a satellitesimal to form at distance \( R \) is that it must be able to withstand disruption due to planetary-produced tidal effects. This is when

\[ \rho_B \geq \frac{3M_p}{2\pi R^3}, \]

in which \( \rho_B \) is the satellitesesimal density and \( M_p \) the planetary mass. Once the dust carpet reaches a thickness, \( h \), such that its density reaches the critical level, gravitational instability will set in and the carpet will break up. If the areal density of the
aggregation of, planetesimals, adapted to satellite formation, is in. For the parameters giving $f_i$ is the dust-carpet thickness when gravitational instability sets in. For the parameters giving figure 13 the masses of satellitesimals as a function of distance from the planet are shown in figure 14.

The Galilean satellites, Io, Europa, Ganymede and Callisto, have masses $8.93 \times 10^{22}$, $4.88 \times 10^{22}$, $1.497 \times 10^{23}$ and $1.068 \times 10^{23}$ kg respectively. At a distance of $4.0 \times 10^6$ km, a satellitesimal mass is about $2 \times 10^{22}$ kg, between about one-half to one-seventh the mass of the Galilean satellites. The next most massive Jovian satellite is Amalthea, closer in than the Galileans and with mass $2 \times 10^{19}$ kg, within the range of masses of satellitesimals in the figure.

5.2.3. From satellitesimals to satellites. The expression found by Safronov (1972) for the time to form a planet by aggregation of, planetesimals, adapted to satellite formation, is

$$\tau_s = \frac{4r_L \rho_m \rho_d}{3\pi \rho_d (1 + \beta)}.$$  (15)

where $r_L$ is the satellite radius, $P$ is the Kepler period in the region of formation, $\beta$ is a constant in the range 4–10, $\rho_m$ is the satellite density and $\rho_d$ is the mean areal density of satellitesimals. For a satellite of mass $10^{23}$ kg and density $2.5 \times 10^3$ kg m$^{-3}$, out to a distance $4.0 \times 10^6$ km all formation times are less than 500,000 years.

For our model disk a dust carpet only formed beyond a distance of $4.0 \times 10^6$ km within $3 \times 10^6$ years but the orbital radius of Callisto, the outermost Galilean satellite, is $1.88 \times 10^6$ km. The answer to this apparent discrepancy is that protosatellite orbits are decaying by the Type-I migration mechanism as they grow and will continue to decay for the lifetime of the disk. Figure 15 shows the decay of a partially formed satellite, with constant mass $10^{22}$ kg, moving in the disk that gave figure 13, but with the disk decaying so that the density falls everywhere by a factor of $e$ every million years. Starting with a radius of $7.5 \times 10^6$ km the orbit decays to the orbit of Callisto, the outermost Galilean satellite, in about $3 \times 10^5$ years and to the orbit of Io, the innermost one, in about $8 \times 10^5$ years. These short decay times are due to the relatively high areal density of the protoplanetary disk. The orbit is close to circular for the whole of the decay period.

The commensurabilities found for the periods of pairs of satellite orbits of Jupiter and Saturn are due to coupling between satellites as described by Melita and Woolfson (1996) for planetary-orbit commensurabilities.

6. The Solar System

6.1. The problem of the terrestrial planets

In the CT, protoplanets form because the mass of a filament condensation, mostly gas, is greater than the Jeans critical mass. Protoplanets initially move towards apastron and by the time they approach the star closely (> $10^5$ years) they are compact objects capable of resisting disruption; gaseous exoplanets, so-called hot Jupiters, are observed at distances down to 0.015 au from their stars. This raises the problem of how, in the Solar System, the terrestrial planets formed.

Dormand and Woolfson (1977) first proposed the idea of a planetary collision to explain terrestrial planets but the knowledge and technology of the time did not enable a realistic model to be created. Part of their consideration, confirmed by modelling, was that when the orbits were evolving, the mass of the circumsolar disk gave a non-central force on the planets, causing orbital precession around the axis of the orbital plane. Differential rates of precession caused slightly inclined pairs of orbits to occasionally intersect in space during the evolutionary period. Dormand and Woolfson postulated an initial system of six planets, the present four
major planets plus two others, and found that the probability that some pair would collide before they all rounded-off was of order 0.1—small but not negligible. A NASA Spitzer Space Telescope observation in August 2009 gave evidence of a planetary collision in the vicinity of the young star HD172555 (age 12 My) within the last few thousand years. In the light of new knowledge and the availability of suitable computational tools the planetary-collision hypothesis has been revisited (Woolfson 2013b).

6.2. Deuterium in the colliding planets

The distribution of deuterium in the colliding planets played a significant role in the postulated collision. The cosmic \( D/H \) ratio is about \( 2 \times 10^{-5} \) but is much higher in many solar-system bodies. In star-forming clouds the overall \( D/H \) ratio is similar to the cosmic ratio but is extremely non-uniform. The \( D/H \) ratios in some molecular species in these clouds, and in low-mass protostars formed within them, are quite high (Roueff et al 2000, Loinard et al 2001, 2002, Parise et al 2002). The average over all the molecular species is estimated as \( D/H > 0.01 \). This concentration of deuterium is due to the phenomenon of grain-surface chemistry. A deuterium atom falling on the surface of an icy grain will exchange places with a hydrogen atom in a molecule because this lowers the energy of the molecule and increases its stability. Over a long period this process concentrates the deuterium in ice molecules. In addition to the common hydrogen-containing molecules—water, ammonia and methane—in cold clouds more complex molecules are present in considerable quantities; the ratio of methanol to water, CH\(_3\)OH/H\(_2\)O, has been found to be in the range 0.1–0.5.

A protoplanet substantially collapses in about \( 10^4 \) years (figure 2) and thereafter evolves to its final state. During the early slow free-fall stage of collapse solid grains migrate towards the centre. Eventually, an iron core with a silicate mantle forms, surrounded by a shell of now-vaporized hydrogen-containing molecules with a high \( D/H \) ratio. Over time, the excess deuterium at the centre of the planet migrates outwards to increase the \( D/H \) ratio in the gaseous envelope.

The concentration of deuterium in ices leads to a diminution of the \( D/H \) ratio in the residual gas. The effect of this in explaining the high \( D/H \) ratios in the ice giants is explained in section 8.1.

6.3. The planetary collision; Earth and Venus

Woolfson (2013b) postulated that in the initial Solar System there were no terrestrial planets but two extra major planets, Enyo of mass 1.9 \( M_\oplus \) and Bellona of mass 2.5 \( M_\oplus \). These planets were modelled in four layers based on incomplete settling of material by density—an iron core with some silicate, a silicate mantle, with some iron, a deuterium-rich gaseous shell with some silicate and a hydrogen–helium atmosphere. The overall composition is given in table 4.

Table 4. The characteristics of the colliding planets.

| Planet | Bellona | Enyo |
|--------|---------|------|
| Mass \((M_\oplus)\)^\text{a} | 798.75 | 598.37 |
| Radius (km) | 9.152 \(10^4\) | 8.647 \(10^4\) |
| Central density (kg m\(^{-3}\)) | 176 500 | 146 500 |
| Central temperature (K) | 85 000 | 74 000 |
| Mass of iron \((M_\infty)\) | 3.00 | 2.50 |
| Mass of silicate \((M_\infty)\) | 12.00 | 10.00 |
| Mass of ice \((M_\infty)\) | 6.00 | 5.00 |

\(^a\) The symbol \( \oplus \) represents the Earth so \( M_\oplus \) is the mass of the Earth.

The four shells—gas, ice, mantle and core are shown alternately as black and white with each planet represented by 4921 SPH points. A Tillotson equation-of-state (Tillotson 1962) was used for the inner three regions and a modified gas law for the atmosphere that accommodated the high pressure regions. Starting with the temperature and density at the centre given by table 4, the equations for gravitational and pressure equilibrium were integrated outwards with the criteria that the density was discontinuous at boundaries between regions but that temperature was continuous. The boundary of the planet was taken when the temperature fell to 100 K. The planets moved on parallel paths with an offset of \( 7 \times 10^4 \) km and had a contact speed of 90 km s\(^{-1}\), corresponding to an approach speed of 42.9 km s\(^{-1}\) when they were far apart. When the shock front reached the Enyo deuterium-rich region the temperature gave a high rate of \( D-D \) nuclear reactions. Holden and Woolfson (1995) had examined in detail the nuclear reactions that would occur with a mixture of deuterium-rich ices and silicates. Once the temperature reached \( \sim 5 \times 10^9 \) K the deuterium was exhausted and heat generation by other thermonuclear reactions was at a lower rate. In the present simulation, when an ice SPH particle reached a temperature of \( 3 \times 10^9 \) K (when the rate of \( D-D \) nuclear reactions was high) the temperature was immediately raised to \( 4 \times 10^9 \) K, lower than the Holden and Woolfson

results indicated. This simplified the incorporation of the nuclear reactions without exaggerating their effect. The locations in which nuclear reactions took place quickly spread to other regions of $D/H$ enhancement.

Figure 17 shows the progress of the simulation. By frame (c) nuclear reactions had occurred, as seen by the rapid outward motion of material. This expansion became greater for successive frames but parts of the cores remained compact and steadily moved apart.

The approach speed of the planets when they were far apart, $42.9 \text{ km s}^{-1}$, suggests a collision in the terrestrial region. It is proposed that the residual cores seen in figure 17 formed the Earth and Venus, from the Bellona and Enyo residual cores, respectively. Their estimated masses from the simulation are $2.4 \, M_\oplus$ and $1.5 \, M_\oplus$, both too large, but the model indicates how the larger terrestrial planets could have originated.

Earth and Venus would have formed by this process while a resisting medium was still in place, although somewhat depleted, and, with a highly centralized medium, their orbits would have rounded off by the Type I migration mechanism within the terrestrial region. It is notable that Venus, while less massive than Earth has a much more massive atmosphere. Since it is closer to the Sun than Earth it would have rounded off in a denser region of the resisting medium, which could explain the difference in the two

Figure 17. The progress of the collision. (a) $t = 0$, (b) $t = 590 \, s$, (c) $t = 1326 \, s$, (d) $t = 2505 \, s$, (e) $t = 3917 \, s$, (f) $t = 5336 \, s$, (g) $t = 6415 \, s$, (h) $t = 7597 \, s$, (i) $t = 8609 \, s$ (Woolfson 2013b).
atmospheres Venus has an atmosphere which is mostly carbon dioxide, the heaviest common gas in the Solar System and hence one that could be retained by a planet that was very hot at birth and remained hot. It also has about 3.5% nitrogen—the only other significant component. These gases almost certainly formed the initial atmosphere of Earth, which was cooler than Venus once it settled down and hence was able to retain more nitrogen. The carbon dioxide was transformed by photosynthesis to give oxygen, initially by cyanobacteria, which formed early in the life of the Earth, and later by more advanced organisms. The earliest oxygen formed was taken up in oxidizing various materials, such as iron and sulphur, the remainder, after all possible oxidization had taken place, forming the present oxygen content of the atmosphere.

A planetary collision, and all that follows from it, could occur in many other scenarios for planet formation, e.g. the NT.

6.4. The Moon

Bellona and Enyo would have had many satellites and, because their masses are much higher than that of Jupiter, some of them could have been more massive than Ganymede. Following the collision, the possible outcomes for a particular satellite are limited to the following:

(I) The satellite could be retained by one or other of the residual cores.
(II) The satellite could end up in a heliocentric orbit.
(III) The satellite could escape from the Solar System.
(IV) The satellite could be disrupted by debris from the collision.

Calculations show that the most likely outcomes are (II) and (III). It is proposed that the Moon was a satellite left in orbit around the Bellona residual core—now Earth. In both mass and density it is intermediate between the Galilean satellites Io and Europa, supporting the view that it was formed by the normal process described in section 6.2.

In 1959 the Soviet Union Luna 3 spacecraft revealed that, although the nearside is dominated by large maria, the far side is predominantly heavily-cratered highlands. Lunar satellite altimeter measurements revealed that the far side has large basins, so the Moon had been bombarded uniformly by large projectiles, but the far-side basins had not filled with magma to give maria. It was suggested that this was due to a difference of crustal thickness on the two sides, which was confirmed when seismometers were left on the Moon by Apollo astronauts. While crustal thickness varied from place to place, the average thickness on the far side is about 12 km greater than that on the nearside.

An early fluid or plastic satellite in synchronous orbit around a planet, should have a thicker low-density crust on the nearside (figure 18). The Moon, formed around either Bellona or Enyo, would have acquired a figure and distribution of material that would have ensured that when it orbited the residual core it would become tidally locked to present the same thicker-crust face to the Earth. Tidal-locking is the process by which satellites present one face to their parent planets and is due to the distribution of material in the satellite, often being elongated in the direction of the planet.

Due to near-surface convection, driven by solidified surface material sinking in the less dense liquid material below, the lunar crust would have solidified to some depth in the perhaps, one or two million years between lunar formation and the planetary collision. The nearside of the Moon, facing the collision, was bombarded by debris travelling at about 100 km s$^{-1}$. Sharing the debris energy with lunar surface material, given that the escape speed from the Moon is only 2.4 km s$^{-1}$, would have led to massive abrasion of the nearside and it has been estimated that up to 50 km thickness of surface material could have been removed in this way (Woolfson 2013b). This requires that the arriving debris removed up to about eight times its own mass that, since the energy of the arriving debris has 1600 times the intrinsic energy of escape from the Moon, is quite feasible.

The current most-supported model for the formation of the Earth–Moon system proposes that, sometime after solar-system formation, a Mars-mass body (Theia) struck the Earth obliquely and that, from the debris of this collision, the Moon accreted in orbit around the Earth. This mechanism has been realistically simulated by Benz et al (1986). Based on the difference of the ages of highland rocks and meteorites, as found by radioactive dating, Benz et al estimated that the Theia event occurred 30–100 million years after solar-system formation. Maria, plus the unfilled basins on the far side, cover about one-third of the Moon’s surface. At their centres the magma depth can be 5–6 km (Head 1976), indicating an original excavation depth of order 7–8 km. Making an extreme assumption that the average depth of maria excavation was only 1 km and that only 10% of the ejecta was retained and spread uniformly over the lunar surface, the average depth of cover in the highlands would be 30 m—probably a gross underestimate. The highland rocks from the Apollo 16 mission probably come from the nearby Mare Nectaris, the solidification ages of which would depend on their excavation depth. Any conclusions about the time of formation of the Moon based on dating highland rocks cannot possibly be valid.

Figure 18. The initial structure of a satellite formed in synchronous orbit around a planet showing core (black), mantle (dark grey) and crust (light grey). The satellite distortion and thickness of the crust are exaggerated.
Mars and Mercury, with about four times and twice the mass of Ganymede respectively, are proposed as escaped satellites. Their relatively-high orbital eccentricities, 0.093 and 0.206 respectively, may indicate that the orbits of these small bodies evolved slowly and that their orbital evolution was terminated before round-off, by loss of the resisting medium.

Like the Moon, Mars has hemispherical asymmetry with magma-covered northern plains and heavily-cratered southern highlands containing one large deep depression, the Hellas Basin (figure 19). A 2 km high scarp separates the two hemispheres. Heavy abrasion of one face could have removed a high proportion of the solid crustal material on the exposed side and volcanism would then have produced a magma plain. The Hellas Basin represents the effect of an exceptionally energetic projectile that penetrated the thicker crust of the southern highlands.

Due to tidal coupling to the Earth, the Moon’s spin axis is contained within its plane of asymmetry. For Mars, not linked to another body, the plane of asymmetry makes an angle of 35° with the spin axis. Mars would have had a molten mantle over which the lithosphere could move—something akin to continental drift. A theorem by Lamy and Burns (1972) states that a spinning body with internal energy dissipation eventually settles down with its spin axis along the principal axis of maximum moment of inertia, a process known as polar wander. McConnell and Woolfson (1983) modelled the surface features of Mars either as positive features—raised regions in the highlands such as the Tharsis Uplift, Argyre Plain, Elysium Plain and Olympus Mons—or as negative features such as the northern plains and the Hellas Basin. They calculated that the principal axis of maximum moment of inertia was 11.9° from the spin axis. The probability of being this close just by chance is about 0.02 and the discrepancy might be due to the crudeness in modelling surface features, or, perhaps, that polar wander was incomplete before the underlying mantle became too rigid to maintain it.

The uncompressed density of Mercury is higher than that of any other planet. Its iron core is about the size of that of Mars and it has been suggested that it was once similar to Mars but had much of its mantle stripped away by a collision with another body.

The most prominent feature on Mercury’s surface is the Caloris Basin, a circular structure about 1550 km in diameter with circular ripples in the surface like the ripples on the surface of a pond when a stone has been thrown into it (figure 20). It is thought to be due to a giant impact. The antipodal region to the Caloris Basin is an area of very disturbed surface called the Chaotic Terrain. The spin period and orbital period of Mercury are in the ratio 2:3, the result of which leads to the Caloris Basin and the Chaotic Terrain being sub-solar at alternate perihelion passages. The possibility that these special features on Mercury’s surface are linked to the spin–orbit characteristics just by chance is rather small.
If Mercury had been a close satellite, and/or in the plane of the collision interface where the debris density was very high (figure 17), then a large proportion of its mantle that faced the collision could have been stripped away. Its closeness to its parent planet would also give had the effect of distorting its internal structure due to spin–orbit synchronization, which presented the bombarded face towards the planet. This extensive damage to Mercury would have required massive reorganization of the residual body to restore it to hydrostatic equilibrium. If the reorganized body still retained some of the original internal distortion then this would explain the present relationship of its surface features with the Sun.

A representation of the way that Mercury was first abraded and then evolved to its present state is shown in figure 21 (Woolfson 2011, pp 229–32). Frame (a) represents the original body with approaching debris. The fully abraded Mercury is shown in frame (b) followed by the flow of the residual mantle material to restore hydrostatic equilibrium. This gives an overshoot of mantle material that flowing in from all directions collides in the centre of the hemisphere facing the collision. Collapse of the overshoot plume gives the Caloris Basin and a reverse flow, as seen in figure 17. A smaller overshoot in the antipodal region gives the Chaotic Terrain.

The interpretation of the Caloris basin as being due to a giant impact would need to explain the relationship of the feature to the orbital characteristics—unless it assumed that it is a chance relationship.

7. Smaller bodies of the Solar System

7.1. The Neptune–Pluto–Triton system

Pluto, a dwarf planet, has an orbit of eccentricity 0.249 and inclination 17° that, at perihelion, passes just within the orbit of Neptune, although, due to a 3:2 commensurability of their orbital periods, these bodies never approach closely. Triton, the seventh largest satellite in the Solar System, is in a retrograde orbit around Neptune, which rules it out as a regular satellite. Woolfson (1999) explained the relationship between these bodies as another outcome of the planetary collision.

The scenario that explains this arrangement is that Triton was a satellite of a colliding planet released into an extended heliocentric orbit taking it beyond Neptune. Pluto was the largest regular satellite of Neptune with mass about two-thirds that of Triton. A computer simulation was made of a collision involving Triton and Pluto. The starting point is illustrated in figure 22. Triton was in a direct heliocentric orbit with perihelion 2.6 au and aphelion 55.6 au and Pluto was in a direct circular orbit, of radius 545 000 km, around Neptune. The before-and-after collision situations are shown in figures 23 and 24. Triton, moving inwards, strikes Pluto a glancing blow that ejects it into a heliocentric orbit with \((a, e) = (39.5\;\text{au}, 0.253)\), very similar to its present orbit. The glancing collision sets it into retrograde spin and a large part of it is sheared off to form its large satellite Charon in a retrograde orbit—a process similar to the Benz et al moon-formation process. Collision fragments form Pluto’s smaller satellites.

Triton loses energy and is captured by Neptune into a retrograde orbit with \((a, e) = (436\;000\;\text{km}, 0.88)\). Tidal effects give rapid round-off and decay of retrograde satellite orbits (McCord 1996), so giving Triton’s present orbit with \((a, e) = (355\;000\;\text{au}, 0.000)\).

7.2. Meteorites, asteroids and comets

Only a small part of the inventory of iron and silicate is in the residual cores. Figure 17 shows that debris is thrown out in all directions; if its intrinsic energy is positive then it will leave the Solar System. Figure 24 shows separately for iron, silicate and ice debris the mass in Earth units per unit intrinsic energy, expressed in GJ kg\(^{-1}\).

More than one-half of the core material is retained and a somewhat smaller fraction of mantle material but most of the ice is expelled from the system. The further out from the planetary centre that material originated, the greater its volatile content and the further out it would tend to end up in the Solar System. Meteorites are fragments from colliding asteroids and these asteroid samples are important sources of information about the early Solar System. Observations, based on comparing meteorite and asteroid reflection spectra confirm this relationship between volatility and distance from the Sun.

Meteorites are of three main types, iron, stony and stony-iron, although irons contain some silicate and stones contain some metal; the stony-iron contain a considerable proportion of each component. This is what would be expected for debris coming from a planet in which the material was at high temperature, was very fluid and where, although the gravitational field was strong, differentiation by

![Figure 22. The initial orbits of Triton and Pluto before the collision.](image-url)
density was not quite complete; stony-irons represent material around the interface of the two main components.

The eccentricities and inclinations of the retained material show some interesting trends, as seen in figure 25. Nearly all the iron and mantle debris orbits are prograde but a considerable proportion of the ice orbits are retrograde.

Asteroid bodies, considered as debris, would have interacted with each other and with planets. Some would have attained safe orbits, such as those in the Asteroid Belt. The total mass of the surviving asteroids is about 4% that of the Moon; most debris, not initially expelled from the Solar System, would have either been swept up by the major planets or thrown outwards by interactions with them.

Comets are associated with two regions of the Solar System—the Kuiper Belt (KB), beginning just beyond Neptune’s orbit, and the Oort Cloud (OC), tens of thousands of au from the Sun. It has been suggested that OC comets must have an external origin because they have $D/H$ ratios 20 times the cosmic value but the high ice $D/H$ ratios of the colliding planets weakens that argument. Bailey (1983) proposed that there are comets, between the OC and the KB, which are drawn outwards to replenish the OC when it is depleted by a close stellar passage or when the Solar System has passed through a disruptive Giant Molecular Cloud. The distributions for silicate and ice in figure 24, which stretch from the inner Solar System continuously to the region of the OC (theoretically to infinity), are consistent with the Bailey model. The KB bodies, perturbed by Neptune to give
short-period comets, form the inner boundary of this distribution. So-called new comets are from the outer part of the distribution—the OC, bodies perturbed inwards by external sources. If there are no major perturbation sources for comets between the KB and OC their presence will not be detected.

Initially the debris orbits pass through the inner Solar System and unless an orbit evolves to give a perihelion outside Neptune’s orbit, in due course it will almost inevitably either be swept up by a major planet or be ejected from the Solar System.

Some volatile debris, with extensive orbits, could have interacted with major planets in evolving orbits at hundreds of au from the Sun and be swung into orbits completely outside the orbit of Neptune. With aphelion = \( Q \) and perihelion = \( q \), figure 26 shows an interaction of a comet with \((Q, q) = (110\text{ au}, 0.5\text{ au})\) with a Jupiter-mass planet with \((Q, q) = (100\text{ au}, 10\text{ au})\) with closest approach \(1.84 \times 10^6 \text{ km}\). After the interaction, for the comet \((Q, q) = (109.9\text{ au}, 42.1\text{ au})\), which would place the body well within the KB. However, it is unlikely that many fragments were affected in this way.

Small bodies such as comets do not have sufficient mass to have an appreciable gravitational effect on the resisting medium and the resistance is primarily due to the ram pressure they experience. For a spherical comet the force experienced will be

\[
F = \pi \rho a^2 V V, \tag{16}
\]

where \( \rho \) is the local medium density, \( a \) the comet radius and \( V \) the velocity of the medium relative to the comet. The effect of such a force has been found for a comet of mass \( 7 \times 10^{11} \text{ kg} \), of density \( 500 \text{ kg m}^{-3} \) (published estimates are between 100 and 1000 kg m\(^{-3}\)), with original perihelion 0.5 au and a range of original semi-major axes. The medium had a total mass of \( 40 M_J \) with an annular distribution of density, similar to that seen in figure 6, given by

\[
\rho(r, z) = C \exp \left( -\frac{(r - d)^2}{2\sigma_r^2} \right) \exp \left( -\frac{hr^2 z^2}{s^2 \sigma_z^2} \right), \tag{17}
\]

where \( d = 100\text{ au} \), \( \sigma_r = 30\text{ au} \), \( h = 10 \), \( s = 20\text{ au} \), \( \sigma_z = 30\text{ au} \) and \( r \) and \( z \) (distance from the mean plane) are expressed in au. The constant \( C \) is determined from the total mass of the disk. The results are shown in figure 27.

It will be seen from the figure that all orbits with original semi-major axes greater than about 70 au end up with perihelia beyond Neptune’s orbit and within the KB region. Their aphelia stretch out to several hundred and even thousands of au and represent Bailey’s inner cloud of comets that form a replenishment reservoir for the OC.

7.3. Dwarf planets

The Moon, Mars, Mercury and Triton have been identified as onetime satellites of the colliding planets. A class of bodies known as dwarf planets, seven in number—Ceres, Pluto, Eris, Makemake, Haumea, V774104 and 2015 TG387 all have masses within the range of solar-system satellites and we also identify them as ex-satellites.

The orbits of bodies of satellite mass are slowly modified by the Type 1 migration process (section 4.4), so slowly that it is unlikely that orbital evolution will progress to the round-off stage within the lifetime of the circumstellar disk. Figure 7 shows that the perihelion distance increases up to the time of round-off. An escaped satellite with an orbit that evolved so that the final perihelion was within the KB would survive. Others would eventually be swept up by a major planet or expelled from the Solar System—except for Ceres that orbits within the Asteroid Belt.

There would have been many other ex-satellites, some possibly more massive than the present dwarf planets. They could have escaped from the Solar System directly as a result of the collision, could have been absorbed by major planets or be in outer regions of the KB and as yet be undetected. One or more of them may have reached the OC. There is a tendency for a group of new comets to come from similar directions with similar orbital parameters, which may be due to the presence of major perturbing bodies within the OC.

If dwarf planets were redefined as ex-satellites large enough for their self-gravity to mould them into hydrostatic
equilibrium and in heliocentric orbits then, according to this model, at present there would now be nine—Ceres, Pluto, Eris, Makemake, Haumea, V774104, 2015 TG387, Mars and Mercury.

8. More deuterium-related events

8.1. The ice giants

The ice giants, Uranus and Neptune, are characterized by three properties:

(i) Low masses compared to the gas giants, due to much smaller gaseous atmospheres.
(ii) Large spin-axis tilts—98.7° for Uranus although that for Neptune, 28.3°, is similar to that of Saturn, 26.7°.
(iii) Atmospheric $D/H$ ratios higher than both the cosmic average and those of the gas-giants.

There are several possible explanations for each of these properties individually.

It has often been suggested that the large axial tilt of Uranus was due to a tangential collision by a large body; if the axial tilt were originally zero then it would require the colliding body to have several Earth masses and an impact speed of about 100 km s$^{-1}$, giving the impacting body the characteristics of a large planet.

A clue as to the history of Uranus and Neptune is given by the $D/H$ values in their atmospheres. The values for the two pairs of giant planets are:

Uranus, $4.4 \pm 0.4 \times 10^{-5}$ and Neptune, $4.1 \pm 0.4 \times 10^{-5}$ (Feuchtgruber et al 2013).

We take the star-forming cloud to have an overall $D/H$ ratio of $2 \times 10^{-6}$, that 75% by mass of the gas was hydrogen and that hydrogen in ice, with a $D/H$ ratio of 0.015, formed 0.1% of the mass of the cloud. On this basis the $D/H$ ratio in the gas was $6.7 \times 10^{-6}$. The ratio of hydrogen in gas form to that contained in ices is 750:1 and any planet with a lesser ratio will have a final $D/H$ ratio greater than the universal value once the deuteration had diffused from the ice into the atmosphere.

For the CT the original spin-axis tilts should not have been large. Here we describe a scenario, consistent with the CT, which explains all three of the characteristics of the ice giants in terms of a single event for each of them. It proposes that the ice giants originally had more extensive atmospheres that were partially stripped off in a close oblique collision with a massive planet—taken as Bellona—soon after their formation so that their ices still retained the original large $D/H$ ratio. This collision also gave a loss of mass and large final spin-axis tilts. SPH models of a proto-ice-giant (PIG) and Bellona were formulated, as described by Woolfson (2007), based on a planet in four distinct layers. For recently-formed planets segregation by density would not be complete so the four layers had compositions as specified below.

Core, consisting of 40% iron with the remainder silicate, mantle, consisting of 85% silicate with the remainder iron, silicate + ice layer with 10% of the mass (~25% by volume) as ice, atmosphere, consisting of hydrogen and helium.

The Tillotson equations of state were used for iron, silicate and ices and previously-published models of Jupiter and Saturn (Stevenson and Salpeter 1976) were used to choose a best model for an atmosphere of the form

$$p = \frac{\rho k T}{\mu} (1 + c \rho),$$

with $c = 0.08$. The PIG had mass $54.58 M_\oplus$ and radius $5.049 \times 10^4$ km. The compositions, by mass, of the inner three layers are given in table 5. The density of points, 3319 in total for each body, is highest at the centre of each body, the region of greatest interest, and falls off towards the boundary. The number of points representing each body is small by modern standards but it was sufficient to give a faithful representation of the collision event and enabled a precise pairs-representation of gravity. An initial configuration of the two planets is shown in figure 28; the motion of the PIG, along the $x$-direction, is displaced relative to Bellona along the $y$ direction by a distance $10^5$ km. Both bodies initially have spin axes in the $z$ direction with a spin period 10 h, similar to the periods of Jupiter and Saturn.

The progress of the oblique collision can be followed in figure 29, which shows the arrangement of the two bodies at four times relative to zero time for figure 28. The PIG is considerably disrupted and distorted, as is seen in frames (a) and (b), but quickly reassembles itself into a spherical form.

The mass and radius of the residual ice giant (RIG), as seen in figure 29(d), are given in table 6. The RIG retains all the core, mantle and ice material of the PIG so the final mass of the atmosphere is $10.18 M_\oplus$, a fraction 0.222 that of the original atmospheric mass of $45.88 M_\oplus$. The mass of hydrogen in

![Figure 28. The initial configuration for the grazing collision.](image-url)
the final atmosphere is $0.75 \times 10.18 M_\oplus = 7.635 M_\oplus$. We now have to estimate how much hydrogen is contained in the $2.5 M_\oplus$ of material described as ice, which is really ice-impregnated silicate material. Taking 0.1% as hydrogen, the estimated mass of hydrogen in the ice layer is $0.025 M_\oplus$. Thus for the RIG the ratio of hydrogen in the atmosphere to that in ice is 305 giving the value of $D/H$ shown in the table.

The characteristics of the RIG are clearly those that we associate with an ice giant. The large tilt is brought about by the direct transfer of angular momentum by the impact plus the huge tidal effect Bellona exerts on the PIG, especially during the close approach of the two bodies. By varying the parameters of the model other outcomes are possible, with more or less mass and radius and with a greater or lesser tilt of the spin axis.

Although from table 6 it seems that the scenario presented here can explain the three cited characteristics of the ice giants it must be treated with some reservation. It is not possible to deduce the form of evolution of the solar-system planetary orbits, including those of the addirional gas giants, and hence how likely it is that a PIG will interact with a gas giant. Since two such interactions are required to explain both Uranus and Neptune the combined probability may be small.

Table 6. Characteristics of Uranus, Neptune and the RIG.

|                  | Uranus | Neptune | RIG   |
|------------------|--------|---------|-------|
| Mass ($M_\oplus$)| 14.54  | 17.15   | 18.93 |
| Radius ($10^4$ km)| 2.556  | 2.476   | 2.543 |
| Mean density ($10^3$ kg m$^{-3}$)| 1.27 | 1.64 | 1.64 |
| Spin-axis tilt (°)| 97.8  | 28.3    | 68.7  |
| $D/H (10^{-5})$   | 4.4    | 4.1     | 4.91  |
| Moment-of-inertia factor | 0.225 | unknown | 0.220 |
| Spin period (h)   | 17.24  | 16.11   | 17    |

Nevertheless, the scenario is consistent with the CT origin of planets.

8.2. Isotopic anomalies in meteorites

Many meteorites show features suggesting the onetime presence of silicate vapour, which would have come naturally from the planetary collision. Chondritic meteorites contain solidifed silicate droplets, chondrules, that condensed from a vapour and minerals condensed in a sequence controlled by their condensation temperatures as the vapour cooled (Larimer 1967).
The isotopic compositions of some meteorites are very different from those of terrestrial material; the differences are denoted as isotopic anomalies. These are described here for carbon, nitrogen, oxygen and neon, but there are many others.

In terrestrial carbon the ratio of the two stable isotopes $^{12}\text{C}$ and $^{13}\text{C}$ is 89.9:1. In mineral grains of silicon carbide, SiC, contained in some chondritic meteorites, the ratio is much less, down to 20:1, giving what is known as heavy carbon. The usual explanation given by meteoriticists is that grains containing heavy carbon drifted into the Solar System from six or more distant carbon stars, each with carbon of different heaviness.

Silicon carbide also contains nitrogen trapped in grain interstices. The ratio of the two stable isotopes $^{14}\text{N}$ and $^{15}\text{N}$ on Earth is 89.9:1. In mineral grains of silicon carbide, SiC, contained in some chondritic meteorites, the ratio is much less, down to 20:1, giving what is known as heavy carbon. The usual explanation given by meteoriticists is that grains containing heavy carbon drifted into the Solar System from six or more distant carbon stars, each with carbon of different heaviness.

Silicon carbide also contains nitrogen trapped in grain interstices. The ratio of the two stable nitrogen isotopes on Earth, $^{14}\text{N}$ to $^{15}\text{N}$, is 270:1. Most SiC-derived nitrogen is light nitrogen with ratios up to 2000:1 but, rarely, heavy nitrogen occurs with ratios down to 50:1.

The three stable isotopes of oxygen have ratios in terrestrial samples given by
\[
^{16}\text{O} : ^{17}\text{O} : ^{18}\text{O} = 0.9527 : 0.0071 : 0.0401.
\]
This mixture, known as SMOW (Standard Mean Ocean Water), also occurs in moon rocks. Two kinds of meteorite contain oxygen isotopic ratios that cannot be explained by processing terrestrial oxygen. These are ordinary chondrites and carbonaceous chondrites, the latter being stony meteorites that contain volatile material. The oxygen anomalies can be explained as the addition of different amounts of pure, or nearly pure, $^{18}\text{O}$ to SMOW. The usual explanation again involves grains drifting across interstellar space and entering the Solar System carrying excess $^{18}\text{O}$.

The final example is neon. The three stable isotopes have terrestrial ratios
\[
^{20}\text{Ne} : ^{21}\text{Ne} : ^{22}\text{Ne} = 0.095 : 0.003 : 0.092.
\]
Neon atoms are trapped in atomic-size cavities, but can be released by heating the meteorite. If neon is found in meteorites then they could not have been substantially heated after the neon was incorporated. The same is true for other gases trapped in the interstices of meteorite grains.

Isotopic compositions of neon in meteorites are very variable and it has been deduced that they come from admixtures of three neon sources with different compositions. However, some meteorites contain pure, or almost pure, $^{22}\text{Ne}$ that is a 9.2% component of terrestrial neon. This anomalous neon is called neon-E. Since $^{22}\text{Ne}$ cannot be separated from a mixture of isotopes some other explanation is required. Sodium has one stable isotope, $^{23}\text{Na}$, but there is a radioactive isotope, $^{22}\text{Na}$, which decays into $^{22}\text{Ne}$. One suggested scenario is that just before solar-system formation a nearby supernova produced $^{22}\text{Na}$ that was incorporated with stable sodium, in minerals. The $^{22}\text{Na}$ decayed and the resultant $^{22}\text{Ne}$ was trapped within the mineral grains. A problem with that scenario is that $^{22}\text{Na}$ has a short half-life, 2.6 years, so that, after production in a supernova, it has to be incorporated into a cool solid body within a period of 10–20 years. This puts a very tight constraint on the timing of the supernova.

Holden and Woolfson (1995) examined the effect of subjecting a mixture of iron, silicates and ices, with a $D/H$ ratio of 0.016, to a triggering temperature of $3 \times 10^8$ K. The rates given by Fowler et al. (1967, 1975) for 548 nuclear reactions were used and 40 decay processes were incorporated in their calculation. All possible cooling factors were included; iron, which took no part in reactions was a coolant and ionization of material, implemented by a solution of the Saha equations (Zel’dovich and Raiser 1966), greatly increased the number of particles present to share the generated energy. The outcome was a nuclear explosion, the products of which explained a number of important light-atom isotopic anomalies. The final temperature was well in excess of $5 \times 10^8$ K.

Another calculation by Woolfson (2011, pp 260–83), using a lesser $D/H$ ratio of 0.01, gave similar results, explaining all anomalies for carbon, nitrogen, oxygen, neon, magnesium, aluminium and silicon. This single explanation replaced a number of ad hoc explanations for individual anomalies.

A large quantity of $^{13}\text{C}$, and radioactive $^{13}\text{N}$ that decays to $^{13}\text{C}$ with a half-life of 9.97 min, was produced, which
explained the full range of heavy-carbon observations. Figure 30 shows the concentrations of $^{12}\text{C}$ and $^{13}\text{C}$ (including $^{15}\text{N}$ contribution). When the temperature exceeds $3 \times 10^8$ K there is a sharp increase in the amount of $^{13}\text{C}$.

There is a small reduction in the amount of $^{14}\text{N}$, including the contribution of $^{14}\text{C}$ that quickly decays to $^{14}\text{N}$, as the explosion progresses, although it picks up at very high temperatures (figure 30). The concentration of $^{15}\text{N}$ is augmented by $^{14}\text{C}$ decay with a half-life of 5739 years. Starting with $^{14}\text{N}$ and $^{15}\text{N}$ trapped inside the grains once they were cool enough to retain the gas, $^{14}\text{C}$, which had been part of the silicon-carbide host, decayed, boosting the amount of $^{15}\text{N}$, so giving light nitrogen. If the original amount of the two stable nitrogen isotopes was not very great then the $^{15}\text{N}$ contribution can make a large proportional change to the $^{15}\text{N}$ concentration.

Towards the end of the explosion there is an almost 100-fold increase in the amount of $^{15}\text{N}$ present, produced by reactions involving heavier elements. These elements may not be uniformly distributed so that $^{15}\text{N}$-rich pockets can form. Even after the decay of $^{14}\text{C}$, the result can be the occasional occurrence of heavy nitrogen.

Figure 31 shows the variation of stable oxygen isotopes during the explosion. The $^{16}\text{O}$ and $^{18}\text{O}$ concentrations include
contributions from the fluorine radioactive isotopes, $^{18}_{}$F and $^{19}_{}$F, which quickly decay to $^{17}_{}$O and $^{19}_{}$O, respectively. Above about $5.8 \times 10^8$ K, the concentrations of $^{17}_{}$O and $^{19}_{}$O greatly diminish leaving virtually pure $^{16}_{}$O that, mixed with SMOW in various proportions, gives the oxygen isotopic anomaly.

A sufficient quantity of $^{22}_{}$Ne was produced in the explosion to explain the production of neon-E. The scale of a planetary collision is small by astronomical standards and the formation of cool grains, which would retain $^{22}_{}$Ne, would take place within hours or days (Woolfson 2011, pp 440–3).

9. Conclusions

This review began with the formation of a star-forming cloud from the ISM and ended with explanations of all important features of the Solar System. Each step is causally related to what preceded it and all the mechanisms involved—tidal disruption, gravitational instability, migration, orbital precession, collisions and nuclear reactions are well understood and occur in other astronomical contexts.

During the rather haphazard development of the CT the consequence of various assumptions were investigated and later the assumptions were confirmed by observations. For example, the CT requires an interaction between a protostar and the protostars they contain. Finally, from its very nature the CT was bound to give some retrograde orbits in planetary systems and later the fact that they existed was confirmed by observations.

Every type of product of the proposed planetary collision can be linked to the formation of some important feature of the Solar System and every important feature of the Solar System can be explained by some product of the collision.

An overview of the contents of this review is given in figure 32. There are two parts to the review linked by the planetary collision. Any theory that led to planets in near-coplanar orbits of high eccentricity evolving in a resisting medium could give a collision and what follows in figure 32.

If the observations of dense embedded clusters and the analysis given in section 3.4 are valid then it appears that the CT can explain the present estimates of the proportion of stars with planets. However, that does not exclude the possibility that other planet-forming processes can occur.

References

Anderson D R et al 2010 Astrophys. J. 709 159–67
Bailey M E 1983 Mon. Not. R. Astr. Soc. 204 603–33
Bailey V et al 2014 Astrophys. J. Lett. 780 L4
Benz W, Slattery W L and Cameron A G W 1986 Icarus 66 515–35
Bonnell I A, Bate M R and Vine S G 2003 Mon. Not. R. Astr. Soc. 349 413–8
Borucki N J et al 2011 Astrophys. J. 736 19–40
Cole G A H and Woolfson M M 2013 Planetary Science: The Science of Planets Around Stars 2nd edn (Boca Raton: CRC Press) p 501
Cook A H 1977 Celestial Masers (Cambridge: Cambridge University Press)
D’Angelo G and Lubow S H 2010 Astrophys. J. 724 730–47
Dormand J R and Woolfson M M 1977 Mon. Not. R. Astr. Soc. 180 243–79
Feuchtgruber H et al 2013 Astron. Astrophys. 551 A126
Fowler W A, Caughlan G R and Zimmerman B 1967 Ann. Rev. Astron. Astrophys. 5 525–76
Fowler W A, Caughlan G R and Zimmerman B 1975 Ann. Rev. Astron. Astrophys. 13 60–112
Gaidos E J 1995 Icarus 114 258–68
Gingold R A and Monahan J J 1977 Mon. Not. R. Astr. Soc. 181 375–89
Golanski Y and Woolfson M M 2001 Mon. Not. R. Astr. Soc. 320 1–11
Goldreich P and Ward W R 1973 Astrophys. J. 183 24–6
Hayashi C 1966 Ann. Rev. Astron. Astrophys. 4 171–92
Head J W 1976 Rev. Geophys. Space Phys. 14 265–300
Heller R 2013 (http://ooo.aip.de/People/heller/content/main_spinorbit.html)
Holden P and Woolfson M M 1995 Earth, Moon Planets 69 201–36
Indulekha K 2013 J. Astrophys. Astron. 34 207–21
Jeans J H 1902 Phil. Trans. R. Soc. A 199 1–53
Jeans J H 1917 Mon. Not. R. Astr. Soc. 77 186–99
Kroupa P 2001 Mon. Not. R. Astr. Soc. 323 231–46
Lamy P L and Burns J A 1972 A. J. Phys. 40 441–4
Larimer J W 1967 Geochim. Cosmochim. Acta 31 1215–38
Lellouch E et al 2001 Astron. Astrophys. 670 610–22
Loinard L et al 2001 Astrophys. J. 552 L163–6
Loinard L et al 2002 Planet. Space Sci. 50 1205–13
Lubow S H and Ida S 2011 Planet Migration ed S Seager (Tucson, AZ: University of Arizona Press) pp 347–71
Lucas P and Roche P 2000 Mon. Not. R. Astr. Soc. 314 858–64
Lucy L B 1977 Astron. J. 82 1013–24
Lynden-Bell D and Pringle J E 1974 Mon. Not. R. Astr. Soc. 168 603–37
McConnell A J and Woolfson M M 1983 Mon. Not. R. Astr. Soc. 204 1221–40
McCord T B 1966 Astron. J. 71 585–90
McLaughlin D B 1924 Astrophys. J. 60 22–31
McCaughrean M J and Stauffer J 1994 AJ 108 1382–97
Melita M D and Woolfson M M 1996 Mon. Not. R. Astr. Soc. 280 854–62
Nutzman P A, Fabrycky D C and Fortney J J 2011 Astrophys. J. Lett. 740 L10
Oxley S and Woolfson M M 2003 Mon. Not. R. Astr. Soc. 343 900–12
Oxley S and Woolfson M M 2004 Mon. Not. R. Astr. Soc. 348 1135–49
Parise B et al 2002 Astron. Astrophys. 393 L49–53
Protskov E-M, Adama F C, Hartmann L W and Tobin J J 2009 Astrophys. J. 697 1020–32
Rossiter R A 1924 Astrophys. J. 60 15–21
Roueff E et al 2000 Astron. Astrophys. 354 L63–6
Safronov V S 1972 *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets* (Jerusalem: Israel Program for Scientific Translation)

Seaton M 1955 *Ann. Astrophys.* **18** 188–205

Stevenson D J and Salpeter E E 1976 *Jupiter* ed T Gehrels (Tucson, AZ: University of Arizona Press) pp 85–112

Sumi T *et al* 2011 *Nature* **473** 349–52

Tillotson J H 1962 *Tec. Rep. General Atomic Report G-3216* General Atomics, San Diego, CA

Torres G 2007 *Astrophys. J.* **654** 1095–109

Weidenschilling S J, Dunn B and Meakin P 1989 *The Formation and Evolution of Planetary Systems* ed H A Weaver and L Danley (Cambridge: Cambridge University Press) pp 131–50

Wood K, Wolff M J, Bjorkman J E and Whitney B 2001 *Astrophys. J.* **564** 889–92

Woolfson M M 1964 *Proc. R. Soc. A* **282** 485–507

Woolfson M M 1979 *Phil. Trans. R. Soc. A* **291** 219–52

Woolfson M M 1999 *Mon. Not. R. Astron. Soc.* **304** 195–8

Woolfson M M 2003 *Mon. Not. R. Astron. Soc.* **340** 43–51

Woolfson M M 2004a *Mon. Not. R. Astron. Soc.* **354** 419–26

Woolfson M M 2004b *Mon. Not. R. Astron. Soc.* **354** 1150–6

Woolfson M M 2007 *Mon. Not. R. Astron. Soc.* **376** 1173–81

Woolfson M M 2011 *On the Origin of Planets: By Means of Natural Simple Processes* (London: Imperial College Press)

Woolfson M M 2013a *Mon. Not. R. Astron. Soc.* **436** 1492–6

Woolfson M M 2013b *Earth, Moon Planets* **111** 1–14

Woolfson M M 2016 *Earth, Moon Planets* **117** 77–91

Zel’dovich Y B and Raiser Y P 1966 *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (New York: Academic)