978-1-108-42050-1 — Astrophysics of Planet Formation
Philip J. Armitage
Excerpt
More Information

1

Observations of Planetary Systems

Planets can be defined informally as large bodies, in orbit around a star, that are not massive enough to have ever derived a substantial fraction of their luminosity from nuclear fusion. This definition fixes the maximum mass of a planet to be at the deuterium burning threshold, which is approximately 13 Jupiter masses for solar composition objects ($1\,M_{\rm J} = 1.899 \times 10^{30}$ g). More massive objects are called brown dwarfs. The lower mass cut-off for what we call a planet is not as easily defined. For a predominantly icy body self-gravity overwhelms material strength when the diameter exceeds a few hundred km, leading to a hydrostatic shape that is near spherical in the absence of rapid rotation (the critical diameter is larger for rocky bodies). Planets (including dwarf planets) are defined as exceeding this threshold size. As planets get larger they typically become more interesting as individual objects; larger bodies retain more internal heat to power geological processes and can hold on to more significant atmospheres. As members of a planetary system the dynamical influence of massive bodies also acts to destabilize and clear out most neighboring orbits. These physical and dynamical characteristics can be used to sub-divide the class of planets, but we will not have cause to make such distinctions in this book. It is likely that some objects of planetary mass exist that are *not* bound to a central star, having formed either in isolation or following ejection from a planetary system. Such objects are normally called "planetary-mass objects" or "free-floating planets."

Complementary constraints on theories of planet formation come from observations of the Solar System and of extrasolar planetary systems. Space missions have yielded exquisitely detailed information on the surfaces (and in some cases interior structures) of the Solar System's planets and satellites, and an increasing number of its minor bodies. Some of the most fundamental facts about the Solar System are reviewed in this chapter, while other relevant observations are discussed subsequently in connection with related theoretical topics. By comparison with the Solar System our knowledge of individual extrasolar planetary systems is meager – in many cases it can be reduced to a handful of imperfectly known numbers characterizing the orbital properties of the planets – but this is compensated in part by the large and rapidly growing number of known systems. It is only by studying extrasolar planetary systems that we can make statistical studies of the range of outcomes of the planet formation process, and avoid bias introduced by the fact that the Solar System must necessarily be one of the subset of planetary systems that admit the existence of a habitable world.

1.1 Solar System Planets

The Solar System has eight planets. Jupiter and Saturn are gas giants composed primarily of hydrogen and helium, although their composition is substantially enhanced in heavier elements when compared to that of the Sun. Uranus and Neptune are ice giants, composed of water, ammonia, methane, silicates, and metals, atop which sit relatively low mass hydrogen and helium atmospheres. There are also four terrestrial planets, two of which (Earth and Venus) have quite similar masses. Mars is almost an order of magnitude less massive and Mercury is smaller still, though its density is anomalously high and similar to that of the Earth. There is more than an order of magnitude gap between the masses of the most massive terrestrial planets and the ice giants, and these two classes of planets have entirely distinct radii and structures. In addition there are a number of dwarf planets, including the trans-Neptunian objects Pluto, Eris, Haumea, and Makemake, and the asteroid Ceres. Many more dwarf planets of comparable size, and possibly even larger objects, remain to be discovered in the outer Solar System.

The orbital elements, masses and equatorial radii of the Solar System's planets are summarized in Table 1.1. With the exception of Mercury, the planets have almost circular, almost coplanar orbits. There is a small but significant misalignment of about 7° between the mean orbital plane of the planets and the solar equator. Architecturally, the most intriguing feature of the Solar System is that the giant and

Table 1.1 *The orbital elements (semi-major axis a, eccentricity e and inclination i), masses and equatorial radii of Solar System planets. The orbital elements are quoted for the J2000 epoch and are with respect to the mean ecliptic. Data from JPL.*

	a / AU	e	i / deg	M_p/g	R_p/cm
Mercury	0.3871	0.2056	7.00	3.302×10^{26}	2.440×10^{8}
Venus	0.7233	0.0068	3.39	4.869×10^{27}	6.052×10^{8}
Earth	1.000	0.0167	0.00	5.974×10^{27}	6.378×10^{8}
Mars	1.524	0.0934	1.85	6.419×10^{26}	3.396×10^{8}
Jupiter	5.203	0.0484	1.30	1.899×10^{30}	7.149×10^{9}
Saturn	9.537	0.0539	2.49	5.685×10^{29}	6.027×10^{9}
Uranus	19.19	0.0473	0.77	8.681×10^{28}	2.556×10^{9}
Neptune	30.07	0.0086	1.77	1.024×10^{29}	2.476×10^{9}

terrestrial planets are clearly segregated in orbital radius, with the giants only being found at large radii where the Solar Nebula (the disk of gas and dust from which the planets formed) would have been cool and icy.

The planets make a negligible contribution ($\simeq 0.13\%$) to the mass of the Solar System, which overwhelmingly resides in the Sun. The mass of the Sun, $M_\odot = 1.989 \times 10^{33}$ g, is made up of hydrogen (fraction by mass in the envelope $X = 0.73$), helium ($Y = 0.25$), and heavier elements (described in astronomical parlance as "metals," with $Z = 0.02$). One notes that even most of the condensible elements in the Solar System are in the Sun. This means that if a significant fraction of the current mass of the Sun passed through a disk during the formation epoch the process of planet formation need not be 100% efficient in converting solid material in the disk into planets. In contrast to the mass, most of the angular momentum of the Solar System is locked up in the orbital angular momentum of the planets. Assuming rigid rotation at angular velocity Ω, the solar angular momentum can be written as

$$J_\odot = k^2 M_\odot R_\odot^2 \Omega, \tag{1.1}$$

where $R_\odot = 6.96 \times 10^{10}$ cm is the solar radius. Taking $\Omega = 2.9 \times 10^{-6}\ \mathrm{s}^{-1}$ (the solar rotation period is 25 dy), and adopting $k^2 \approx 0.1$ (roughly appropriate for a star with a radiative core), we obtain as an estimate for the solar angular momentum $J_\odot \sim 3 \times 10^{48}\ \mathrm{g\ cm^2\ s^{-1}}$. For comparison, the orbital angular momentum associated with Jupiter's orbit at semi-major axis a is

$$J_\mathrm{J} = M_\mathrm{J}\sqrt{G M_\odot a} \simeq 2 \times 10^{50}\ \mathrm{g\ cm^2\ s^{-1}}. \tag{1.2}$$

Even this value is small compared to the typical angular momentum contained in molecular cloud cores that collapse to form low mass stars. We infer that substantial segregation of angular momentum and mass must have occurred during the star formation process.

The orbital radii of the planets do not exhibit any relationships that yield immediate clues as to their formation or early evolution. (We briefly mention the Titius–Bode law in Section 7.4.3, but this empirical relation is not thought to have any fundamental basis.) From a dynamical standpoint the most relevant fact is that although the planets orbit close enough to perturb each other's orbits, the perturbations are all nonresonant. Resonances occur when characteristic frequencies of two or more bodies display a near-exact commensurability. They adopt disproportionate importance in planetary dynamics because, in systems where the planets do not make close encounters, gravitational forces between the planets are generally much smaller (typically by a factor of 10^3 or more) than the dominant force from the star. These small perturbations are largely negligible unless special circumstances (i.e. a resonance) cause them to add up coherently over time. The simplest type of resonance, known as a *mean-motion resonance* (MMR), occurs when the periods P_1 and P_2 of two planets satisfy

$$\frac{P_1}{P_2} \simeq \frac{i}{j}, \tag{1.3}$$

where i and j are integers and use of the approximate equality sign denotes the fact that such resonances have a finite width. One can, of course, always find a pair of integers such that this equation is satisfied for arbitrary P_1 and P_2, so a more precise statement is that there are no dynamically important resonances among the major planets.[1] Nearest to resonance in the Solar System are Jupiter and Saturn, whose motion is affected by their proximity to a 5:2 mean-motion resonance known as the "great inequality" (the existence of this near resonance, though not its dynamical significance, was known even to Kepler). Among lower mass objects Pluto is one of a large class of Kuiper Belt Objects (KBOs) in 3:2 resonance with Neptune, and there are many examples of important resonances among satellites and in the asteroid belt.

1.2 The Minimum Mass Solar Nebula

The mass of the disk of gas and dust that formed the Solar System is unknown. However, it is possible to use the observed masses, orbital radii and compositions of the planets to derive a *lower limit* for the amount of material that must have been present, together with a crude idea as to how that material was distributed with distance from the Sun. This is called the "minimum mass Solar Nebula" (Weidenschilling,1977a). The procedure is simple:

(1) Starting from the observed (or inferred) masses of heavy elements such as iron in the planets, augment the mass of each planet with enough hydrogen and helium to bring the augmented mixture to solar composition.
(2) Divide the Solar System up into annuli, such that each annulus is centered on the current semi-major axis of a planet and extends halfway to the orbit of the neighboring planets.
(3) Imagine spreading the augmented mass for each planet across the area of its annulus. This yields a characteristic gas surface density Σ (units g cm^{-2}) at the location of each planet.

Following this scheme, out to the orbital radius of Neptune the derived surface density scales roughly as $\Sigma(r) \propto r^{-3/2}$. Since the procedure for constructing the

[1] Roughly speaking, a resonance is typically dynamically important if the integers i and j (or their difference) are small. Care is needed, however, since although the 121:118 mean-motion resonance between Saturn's moons Prometheus and Pandora formally satisfies this condition (since the *difference* is small) one would not immediately suspect that such an obscure commensurability would be significant.

distribution is somewhat arbitrary it is possible to obtain a number of different normalizations, but the most common value used is that quoted by Hayashi (1981):

$$\Sigma(r) = 1.7 \times 10^3 \left(\frac{r}{1\ \mathrm{AU}}\right)^{-3/2}\ \mathrm{g\ cm^{-2}}. \tag{1.4}$$

Integrating this expression out to 30 AU the enclosed mass works out to be 0.01 $M_\odot$, which is comparable to the estimated masses of protoplanetary disks around other stars (though these have a wide spread). Hayashi (1981) also provided an estimate for the surface density of solid material as a function of radius in the disk:

$$\Sigma_s(\mathrm{rock}) = 7.1 \left(\frac{r}{1\ \mathrm{AU}}\right)^{-3/2}\ \mathrm{g\ cm^{-2}}\ \text{for}\ r < 2.7\ \mathrm{AU}, \tag{1.5}$$

$$\Sigma_s(\mathrm{rock/ice}) = 30 \left(\frac{r}{1\ \mathrm{AU}}\right)^{-3/2}\ \mathrm{g\ cm^{-2}}\ \text{for}\ r > 2.7\ \mathrm{AU}. \tag{1.6}$$

These distributions are shown in Fig. 1.1. The discontinuity in the solid surface density at 2.7 AU is due to the presence of icy material in the outer disk that would be destroyed in the hotter inner regions.

Although useful as an order of magnitude guide, the minimum mass Solar Nebula (as its name suggests) provides only an approximate lower limit to the amount of mass that must have been present in the Solar Nebula. As we will discuss later, it is very likely that both the gas and solid disks evolved substantially over time. There is no reason to believe that the minimum mass Solar Nebula reflects either the initial inventory of mass in the Solar Nebula, or the steady-state profile of the protoplanetary disk around the young Sun.

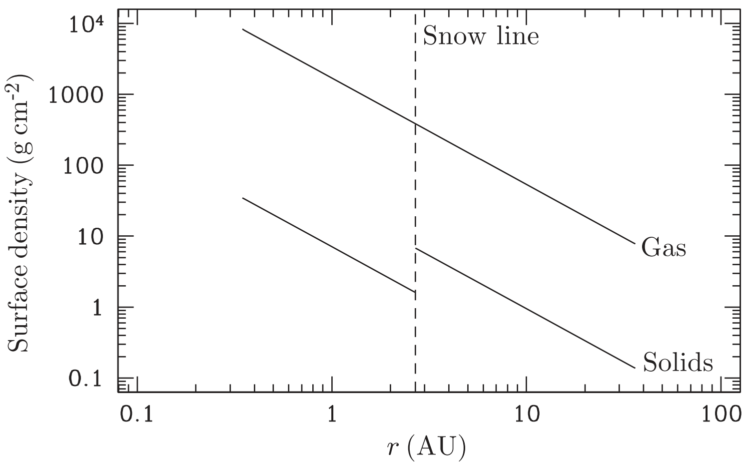

Figure 1.1 The surface density in gas (upper line) and solids (lower broken line) as a function of radius in Hayashi's minimum mass Solar Nebula. The dashed vertical line denotes the location of the snow line.

1.3 Minor Bodies in the Solar System

In addition to the planets, the Solar System contains a wealth of minor bodies: asteroids, Trans-Neptunian Objects (TNOs, including those in the Kuiper Belt), comets, and planetary satellites. The total mass in these reservoirs is now small.[2] The main asteroid belt has a mass of about $5 \times 10^{-4}\ M_\oplus$ (Petit *et al.*, 2001), while the more uncertain estimates for the Kuiper Belt are of the order of $0.1\ M_\oplus$ (Chiang *et al.*, 2007). Although dynamically unimportant, the distribution of minor bodies is extremely important for the clues it provides to the early history of the Solar System. As a very rough generalization the Solar System is dynamically full, in the sense that most locations where small bodies could stably orbit for billions of years are, in fact, populated. In the inner Solar System, the main reservoir is the main asteroid belt between Mars and Jupiter, while in the outer Solar System the Kuiper Belt is found beyond the orbit of Neptune.

Figure 1.2 shows the distribution of a sample of numbered asteroids in the inner Solar System, taken from the *Jet Propulsion Laboratory*'s small-body database. Most of the bodies in the main asteroid belt have semi-major axes a in the range between 2.1 and 3.3 AU. The distribution of a is by no means smooth, reflecting the crucial role of resonant dynamics in shaping the asteroid belt. The prominent regions, known as the Kirkwood (1867) gaps, where relatively few asteroids are

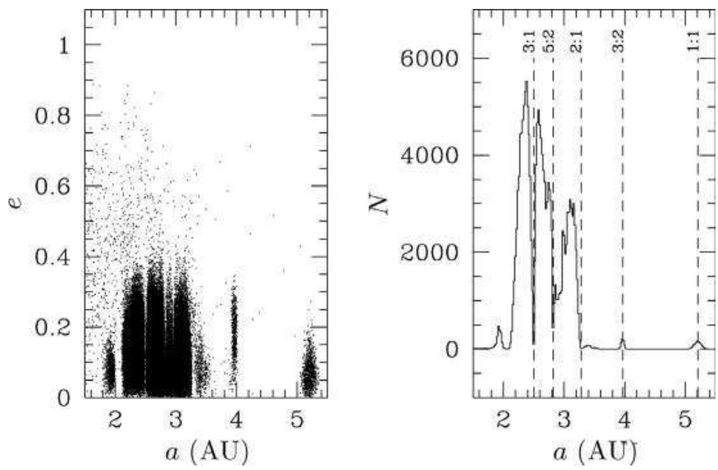

Figure 1.2 The orbital elements of a sample of numbered asteroids in the inner Solar System. The left-hand panel shows the semi-major axes a and eccentricity e of asteroids in the region between the orbits of Mars and Jupiter. The right-hand panel shows a histogram of the distribution of asteroids in semi-major axis. The locations of a handful of mean-motion resonances with Jupiter are marked by the dashed vertical lines.

[2] Indirect evidence suggests that the primordial asteroid and Kuiper belts were much more massive. A combination of dynamical ejection, and/or collisional grinding of bodies to dust that is then rapidly lost as a result of radiation pressure forces is likely to be responsible for their depletion.

found coincide with the locations of mean-motion resonances with Jupiter, most notably the 3:1 and 5:2 resonances. In addition to these locations – at which resonances with Jupiter are evidently depleting the population of minor bodies – there are *concentrations* of asteroids at both the co-orbital 1:1 resonance (the Trojan asteroids), and at the interior 3:2 resonance (the Hilda asteroids). This is a graphic demonstration of the fact that different resonances can either destabilize or protect asteroid orbits (for a thorough analysis of the dynamics involved the reader should consult Murray & Dermott, 1999). Also notable is that the asteroids, unlike the major planets, have a distribution of eccentricity e that extends to moderately large values. Between 2.1 and 3.3 AU the mean eccentricity of the numbered asteroids is $\langle e \rangle \simeq 0.14$. As a result, collisions in the asteroid belt today typically involve relative velocities that are large enough to be disruptive. Indeed, a number of asteroid families (Hirayama, 1918) are known, whose members share similar orbital elements (a, e, i). These asteroids are interpreted as debris from disruptive collisions taking place within the asteroid belt, in some cases relatively recently (within the last few Myr, e.g. Nesvorný *et al.*, 2002).

Figure 1.3 shows the distribution of a sample of outer Solar System bodies, maintained by the IAU's *Minor Planet Center*. Among the known planets outer Solar System bodies interact most strongly with Neptune, and to leading order they are classified based upon the nature of that interaction.

- *Resonant Kuiper Belt Objects (KBOs)* currently occupy mean-motion resonances with Neptune. The most common resonance is the 3:2 that is occupied by Pluto, and such objects are also called Plutinos. The eccentricity of some Plutinos – including Pluto itself – is large enough that their perihelion lies within the orbit of Neptune, and these objects depend upon their resonant configuration to avoid close encounters. The existence of this large population of moderately eccentric resonant bodies provided the original evidence for models in which Neptune migrated outward early in Solar System history.
- *Classical KBOs* orbit in a relatively narrow belt between Neptune's 3:2 and 2:1 MMRs (39.5AU $< a <$ 47.8 AU), and their number drops sharply toward the upper end of this range of semi-major axes (Trujillo & Brown, 2001). These objects are nonresonant and they have low enough eccentricity to avoid scattering encounters with Neptune. They can be divided into two sub-populations. The *cold* classical belt objects have lower inclinations $i < 2^\circ$ (and generally also lower eccentricities) than the *hot* objects, which have $i > 6^\circ$ (Dawson & Murray-Clay, 2012). (Objects with intermediate inclinations cannot be classified reliably using only orbital information.) The dynamical classification matches up to apparent physical differences that are inferred from measurements of the color and size distribution, which suggests that the cold and hot populations derive from distinct source populations.

Cambridge University Press
978-1-108-42050-1 — Astrophysics of Planet Formation
Philip J. Armitage
Excerpt
More Information

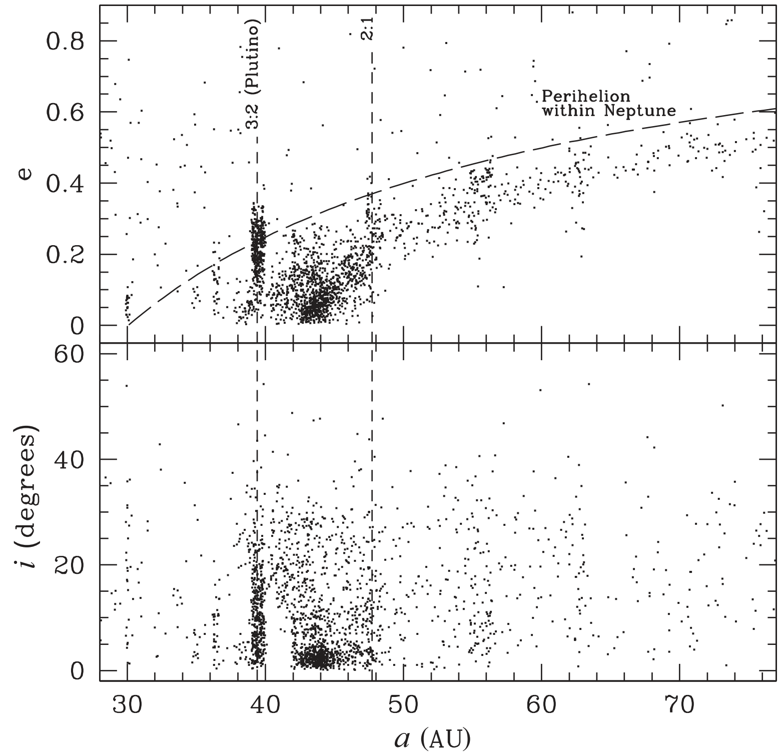

Figure 1.3 The distribution of eccentricity and inclination for a sample of minor bodies in the outer Solar System beyond the orbit of Neptune. The dashed vertical lines indicate the locations of mean-motion resonances with Neptune. Objects with eccentricity above the long-dashed line have perihelia that lie within the orbit of Neptune.

- The *scattering* population have perihelia $a(1 - e) \approx a_{\rm Nep}$, where $a_{\rm Nep}$ is the semi-major axis of Neptune. These objects are in close dynamical contact with Neptune, and their orbits evolve as a result of the planet's perturbations.
- The *detached* population makes up the rest – objects on typically quite eccentric orbits that are not currently in dynamical contact with Neptune. Some of these objects are so detached that their orbits must have been established under different dynamical conditions earlier in Solar System history. A notable example is the large object Sedna, whose perihelion distance of 76 AU lies way beyond the orbit of Neptune.

Comets that approach the Sun are more easily accessible messengers from the outer Solar System. The Jupiter Family comets have low inclinations and are thought to originate from within the TNO reservoirs discussed above. Other comets, however, have a clearly distinct origin. In particular, among comets that are identified for the first time there is a population that has a broad inclination distribution and semi-major axes that cluster at $a \sim 2 \times 10^4$ AU, far beyond the Kuiper Belt. It was this evidence that led Jan Oort to postulate that the Sun is surrounded by a quasi-spherical reservoir of comets, now called the Oort cloud (Oort, 1950). The Oort cloud was established at an early epoch and delivers comets toward the inner Solar System over time as a consequence of Galactic tidal forces and perturbations from passing stars.

Planetary satellites in the Solar System also fall into several classes. The regular satellites of Jupiter, Saturn, Uranus, and Neptune have relatively tight prograde orbits that lie close to the equatorial plane of their respective planets. This suggests that these satellites formed from disks, analogous to the Solar Nebula itself, that surrounded the planets shortly after their formation. The total masses of the regular satellite systems are a relatively constant fraction (about 10^{-4}) of the mass of the host planet, with the largest satellite, Jupiter's moon Ganymede, having a mass of $0.025\ M_\oplus$. The presence of resonances between different satellite orbits – most notably the *Laplace resonance* that involves Io, Europa, and Ganymede (Io lies in 2:1 resonance with Europa, which in turn is in 2:1 resonance with Ganymede) – is striking. As in the case of Pluto's resonance with Neptune, the existence of these nontrivial configurations among the satellites provides evidence for past orbital evolution that was followed by resonant capture. Orbital migration within a primordial disk, or tidal interaction with the planet, are candidates for explaining these resonances.

The giant planets also possess extensive systems of irregular satellites, which are typically more distant and which do not share the common disk plane of the regular satellites. These satellites were probably captured by the giant planets from heliocentric orbits.

The sole example of a natural satellite of a terrestrial planet – the Earth's Moon – is distinctly different from any giant planet satellite. Relative to its planet it is much more massive (the Moon is more than 1% of the mass of the Earth), and its orbital angular momentum makes up most of the angular momentum of the Earth–Moon system. The Moon's composition is not the same as that of the Earth; there is less iron (resulting in a lower density than the uncompressed density of the Earth) and evidence for depletion of some volatile elements. Some aspects of the composition, in particular the ratios of stable isotopes of oxygen, are however essentially indistinguishable from those measured from terrestrial mantle samples. Qualitatively these properties are interpreted within models in which the Moon formed from the cooling of a heavy-element rich disk generated following a giant impact early in the Earth's history (Hartmann & Davis, 1975; Cameron & Ward, 1976), though some of the quantitative constraints remain challenging to reproduce. Pluto's large moon Charon may have formed in the aftermath of a similar impact.

1.4 Radioactive Dating of the Solar System

Determining the ages of individual stars from astronomical observations is a difficult exercise, and good constraints are normally only possible if the frequencies of stellar oscillations can be identified via photometric or spectroscopic data.

Much more accurate age determinations are possible for the Solar System, via radioactive dating of apparently pristine samples from meteorites.

It is worth clarifying at the outset how radioactive dating works, because it is not as simple as one might initially think. Consider a notional radioactive decay A $\rightarrow$ B that occurs with mean lifetime τ. After time t the abundance n_A of "A" is reduced from its initial value n_{A0} according to

$$n_A = n_{A0} e^{-t/\tau}, \tag{1.7}$$

while that of "B" increases,

$$n_B = n_{B0} + n_{A0}(1 - e^{-t/\tau}). \tag{1.8}$$

We can assume that τ is known precisely from laboratory measurements. However, it is clear that we cannot in general determine the age because we have three unknowns (t and the initial abundances of the two species) but only two observables (the current abundances of each species). Getting around this roadblock requires considering more complex decays and imposing assumptions about how the samples under consideration formed in the first place.

For a simple example that works we can look at a rock containing radioactive potassium (^{40}K) that solidifies from the vapor or liquid phases during the epoch of planet formation. One of the decay channels of ^{40}K is

$$^{40}\mathrm{K} \rightarrow {}^{40}\mathrm{Ar}. \tag{1.9}$$

This decay has a half-life of 1.25 Gyr and a branching ratio $\xi \approx 0.1$. (The branching ratio describes the probability that the radioactive isotope decays via a specific channel. In this case ξ is small because ^{40}K decays more often into ^{40}Ca.) If we assume that the rock, once it has solidified, traps the argon and that *there was no argon in the rock to start with*, then we have eliminated one of the generally unknown quantities and measuring the relative abundance of ^{40}Ar and ^{40}K suffices to determine the age. Quantitatively, if the parent isotope ^{40}K has an initial abundance $n_\mathrm{p}(0)$ when the rock solidifies at time $t = 0$, then at later times the abundances of the parent isotope n_p and daughter isotope n_d are given by the usual exponential formulae that characterize radioactive decay:

$$\begin{aligned} n_\mathrm{p} &= n_\mathrm{p}(0) e^{-t/\tau}, \\ n_\mathrm{d} &= \xi n_\mathrm{p}(0) \left[1 - e^{-t/\tau}\right], \end{aligned} \tag{1.10}$$

where τ, the mean lifetime, is related to the half-life via $\tau = t_{1/2}/\ln 2$. The ratio of the daughter to parent abundance is

$$\frac{n_\mathrm{d}}{n_\mathrm{p}} = \xi \left(e^{t/\tau} - 1\right). \tag{1.11}$$

A laboratory measurement of the left-hand-side then fixes the age provided that the nuclear physics of the decay (the mean lifetime and the branching ratio) is