

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 ($1 M_J = 1.899 \times 10^{30}$ g). More massive objects are called brown dwarfs. The lower mass cut-off for what we call a planet is not as well defined. Currently, the International Astronomical Union (IAU) requires a Solar System planet to be massive enough that it is able to clear the neighborhood around its orbit of other large bodies. Smaller objects that are massive enough to have a roughly spherical shape but which do not have a major dynamical influence on nearby bodies are called “dwarf planets.” It is likely that some objects of planetary mass exist that are *not* bound to a central star, either having formed 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 to all of the planets have yielded exquisitely detailed information on the surfaces (and in some cases interior structures) of the Solar System’s planets, satellites, and minor bodies. A handful 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 indeed – 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 any bias introduced

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

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. Two are gas giants (Jupiter and Saturn) composed primarily of hydrogen and helium, although even their composition is substantially enhanced in heavier elements as compared to that of the Sun. Two are ice giants (Uranus and Neptune), composed of water, ammonia, methane, silicates, and metals, atop which sit relatively low mass hydrogen and helium atmospheres. Finally there are four terrestrial planets, two of which (Earth and Venus) are substantially more massive than the other two (Mars and Mercury). In addition there are a number of dwarf planets, including the trans-Neptunian objects Pluto, Eris, Haumea, and Makemake, and the asteroid Ceres. It is very likely that many more dwarf planets of comparable size 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 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.

1.1 Solar System planets

3

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

$$J_{\odot} = k^2 M_{\odot} R_{\odot}^2 \Omega, \quad (1.1)$$

where $R_{\odot} = 6.96 \times 10^{10}$ cm is the Solar radius. Taking $\Omega = 2.9 \times 10^{-6}$ 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 $J_{\odot} \sim 3 \times 10^{48}$ g cm 2 s $^{-1}$. For comparison, the orbital angular momentum associated with Jupiter’s orbit at semi-major axis a is

$$J_J = M_J \sqrt{GM_{\odot} a} \simeq 2 \times 10^{50} \text{ g cm}^2 \text{ s}^{-1}. \quad (1.2)$$

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.¹ Although the planets orbit close enough to perturb each other’s orbits, the perturbations between the main planets 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*, occurs when the periods P_1 and

¹ The Titius–Bode law, a well-known empirical relation between the orbital radii of the planets, is not thought to have any fundamental basis.

P_2 of two planets satisfy

$$\frac{P_1}{P_2} \simeq \frac{i}{j}, \quad (1.3)$$

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.² 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.

1.1.1 *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 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, one finds that 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 distribution is somewhat arbitrary it is possible to obtain a number

² 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.

1.1 Solar System planets

5

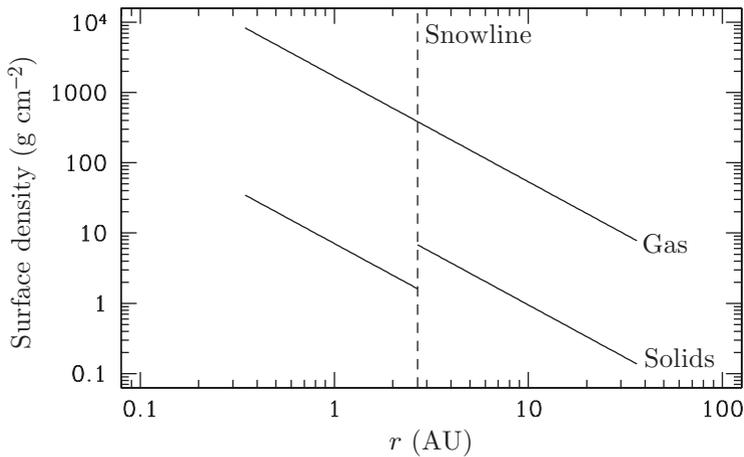


Fig. 1.1. The surface density in gas (upper line) and solids (lower broken line) as a function of radius in Hayashi's minimum mass Solar Nebula. The dashed vertical line denotes the location of the snowline.

of different normalizations, but the most common value used is that quoted by Hayashi (1981),

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

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,

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

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

These distributions are shown in Figure 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, there are myriad reasons to suspect 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.

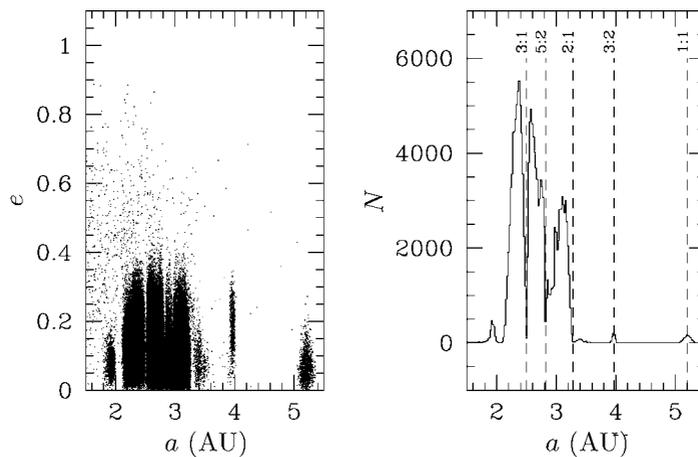


Fig. 1.2. The orbital elements of a sample of numbered asteroids in the inner Solar System. The left-hand panel shows the semi-major axes a and eccentricity e of asteroids in the region between the orbits of Mars and Jupiter. The right-hand panel shows a histogram of the distribution of asteroids in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

1.2 Minor bodies in the Solar System

In addition to the planets, the Solar System contains a wealth of minor bodies: asteroids, KBOs, comets and planetary satellites. Although the total mass in these reservoirs is now small³ – estimates for the Kuiper Belt, for example, are of the order of $0.1 M_{\oplus}$ (Chiang *et al.*, 2007) – the distribution of minor bodies is as important as study of the planets for the clues it provides to the early history of the Solar System. The first significant fact to note is that 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. However, the distribution of a is by no means smooth, and the crucial role of resonant dynamics in shaping the asteroid belt is obvious. There

³ 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.

1.2 Minor bodies in the Solar System

7

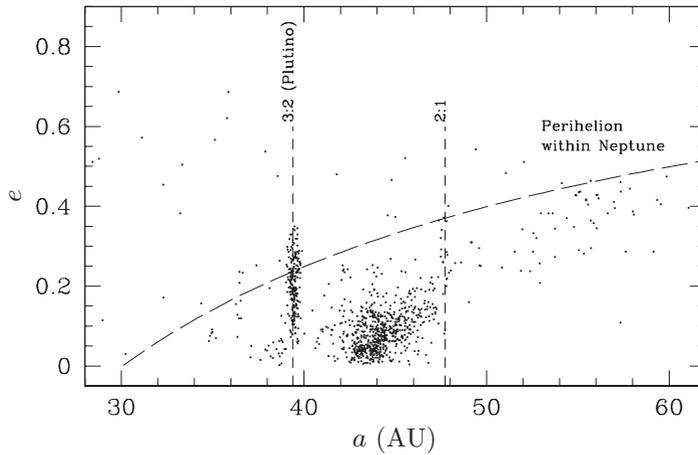


Fig. 1.3. The orbital elements of a sample of minor bodies in the outer Solar System beyond the orbit of Neptune. The dashed vertical lines indicate the locations of mean-motion resonances with Neptune. Objects with eccentricity above the long-dashed line have perihelia that lie within the orbit of Neptune.

are prominent regions, known as the Kirkwood (1867) gaps, where relatively few asteroids are found. These 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). Evidently 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*. Outer Solar System bodies are divided into a number of dynamical classes. Resonant Kuiper Belt Objects orbit within one of Neptune's mean-motion resonances, most commonly the 3:2 resonance occupied by Pluto (such objects are called Plutinos). Some of these KBOs, like Pluto itself,

cross Neptune's orbit and depend upon their resonant configuration to avert close encounters. The existence of this large resonant population of KBOs is believed to result from the outward migration of Neptune early in the Solar System's history. Classical KBOs comprise low-eccentricity bodies that are not in resonance with Neptune. Their orbits, when treated as test particles in the restricted three-body problem with Neptune as a perturber, are such that they will never cross the orbit of Neptune. The number of known classical KBOs drops rapidly for semi-major axes $a \gtrsim 47$ AU, reflecting either a physical edge to the population or, perhaps, a discontinuity in the physical properties of classical KBOs at this radius (Trujillo & Brown, 2001). Finally the scattered KBOs have typically highly eccentric and inclined orbits that do not cross the orbit of Neptune. A notable example is the large object Sedna, whose perihelion distance of 76 AU lies way beyond the orbit of Neptune.

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. Finally the properties of the Moon seem most consistent with yet a third formation scenario – a giant impact early in the Earth's history which resulted in a heavy-element rich disk that condensed into the Moon (Hartmann & Davis, 1975; Cameron & Ward, 1976). It is possible that Pluto's large moon Charon formed in the aftermath of a similar impact.

1.3 Radioactive dating of the Solar System

Determining the ages of individual stars from astronomical observations is a difficult and usually imprecise exercise. For the Solar System, uniquely, the availability

1.3 Radioactive dating of the Solar System

9

of apparently pristine meteorites allows accurate determination of its age and good constraints on the timing of some phases of the planet formation process.

The principle of radioactive dating of rock samples can be illustrated with a simple example. Consider 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



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 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 $n_p(0)$ 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,

$$\begin{aligned} n_p &= n_p(0)e^{-t/\tau} \\ n_d &= \xi n_p(0) [1 - e^{-t/\tau}], \end{aligned} \quad (1.8)$$

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

$$\frac{n_d}{n_p} = \xi (e^{t/\tau} - 1). \quad (1.9)$$

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 accurately known. Notice that this method works to date the age of the rock (rather than the epoch when the radioactive potassium was formed) because minerals have distinct chemical compositions that differ – often dramatically so – from the average composition of the protoplanetary disk. In the example above, it is reasonable to assume that any ^{40}Ar atoms formed prior to the rock solidifying will not be incorporated into the rock, first because the argon will be diluted throughout the disk and second because it is an unreactive element that will not be part of the same minerals as potassium.

Radioactive dating is rarely as simple as the above illustration would suggest. A somewhat more representative example is the decay of rubidium 87 into strontium 87,



which occurs with a half-life of 48.8 Gyr. Unlike argon, strontium is not a noble gas, and we cannot assume that the rock is initially devoid of strontium. If we denote the initial abundance of the daughter isotope as $n_d(0)$, then measurement of the ratio (n_d/n_p) yields a single constraint on two unknowns (the initial daughter abundance and the age) and dating appears impossible. Again, the varied chemical properties of rocks allow progress. Suppose we measure samples from two different minerals within the same rock, and compare the abundances of ^{87}Rb and ^{87}Sr not to each other, but to the abundance of a separate stable isotope of strontium ^{86}Sr . Since ^{86}Sr is chemically identical to the daughter isotope ^{87}Sr that we are interested in, it is reasonable to assume that the ratio $^{87}\text{Sr}/^{86}\text{Sr}$ was initially constant across samples. The ratio $^{87}\text{Rb}/^{86}\text{Sr}$, on the other hand, can differ between samples. As the rock ages, the abundance of the parent isotope drops and that of the daughter increases. Quantitatively,

$$\begin{aligned} n_p &= n_p(0)e^{-t/\tau} \\ n_d &= n_d(0) + \xi n_p(0) [1 - e^{-t/\tau}]. \end{aligned} \quad (1.11)$$

Eliminating $n_p(0)$ between these equations and dividing by the abundance n_{ds} of the second stable isotope of the daughter species (^{86}Sr in our example) we obtain

$$\left(\frac{n_d}{n_{ds}}\right) = \left(\frac{n_d(0)}{n_{ds}}\right) + \xi \left(\frac{n_p}{n_{ds}}\right) [e^{t/\tau} - 1]. \quad (1.12)$$

The first term on the right-hand-side is a constant. We can then plot the relative abundances of the parent isotope (n_p/n_{ds}) and the daughter isotope (n_d/n_{ds}) from different samples on a ratio–ratio plot called an isochron diagram, such as the one shown schematically in Fig. 1.4. Inspection of Eq. (1.12) shows that we should expect the points from different samples to lie on a straight line whose slope (together with independent knowledge of the mean lifetime) fixes the age. Two samples are in principle sufficient to yield an age determination, but additional data provide a check against possible systematic errors – if the points fail to lie on a straight line something is wrong.

Radioactive dating of primitive meteorites known as chondrites using these techniques dates the formation of the Solar System to an epoch 4.57 Gyr ago. Knowing this age accurately is useful for calibrating Solar evolution models, but is otherwise of little interest for planet formation. More valuable are constraints on the time scale of critical phases of the planet formation process, where the questions we would like to answer are more subtle. For example, it would be valuable to be able to know whether the formation of km-sized bodies called planetesimals was sudden or spread out over many Myr. Addressing such questions is challenging using absolute chronometers based on long-lived isotopes (those with half-lives of the order of Gyr), so a complementary approach that derives *relative* ages from