**The Birth of the Sun**

The solar system formed 4.5–4.6 billion years ago (Patterson 1956) by collapse of a portion of a molecular cloud of gas and dust rather like the Eagle or Orion nebulae. Some of the star dust from the solar nebula has now been isolated from primitive meteorites. The isotopic compositions of many elements in these grains are vastly different from those of our solar system and provide fingerprints of the stars that preceded our Sun (Nittler 2003). These include red giants, supernovae and novae.

Why this cloud of stellar debris collapsed to form our Sun has been uncertain, but recent discoveries provide support for the theory that a shock wave from a supernova explosion provided the trigger. The former presence of short-lived radionuclides at the start of the solar system has been detected from the isotopic compositions of their corresponding radiogenic daughter elements in meteorites (Table 1). Some of these radionuclides, such as calcium-41, aluminum-26 and iron-60, have such short half-lives that they must have been formed at, or shortly before, the start of the solar system, or else they would no longer have been present. The relative proportions of aluminum-26 and iron-60 are consistent with production in a supernova.

**Disks and Planet Formation**

Immanuel Kant (1724–1804) and Pierre-Simon de Laplace (1749–1827) both argued that the solar system originated from a swirling disk of gas and dust (Fig. 1). In 1755, Kant proposed the idea that the planets accreted from circumstellar dust under gravity. Laplace presented a similar model in 1796 and explained mathematically the rotation of the disk and planets. This resulted from the increase in angular momentum associated with the contraction of the solar nebula as it collapsed to form the Sun. Although this model has been debated over the years, the Hubble Space Telescope has now provided fascinating images of young stars embedded in such opaque, dusty disks around stars in the Orion Nebula. These young (<1 Myr old) objects are thought to develop into very energetic T-tauri stars characterised by excess infrared radiation emanating from circumstellar dust. The rates of accretion of gas and dust onto a star have been studied by examining the inner edge of the disk spectroscopically (Hartmann 2000). These rates suggest that the disk’s dust and, presumably, gas are consumed within a few million years.
years. Dynamic simulations applied to our solar system show that dust and small objects would have been quickly swept into the Sun, unless incorporated into planetary objects. Accretion of the terrestrial planets is usually modeled (Chambers 2004) in terms of four mechanistically distinct stages:
1. Settling of circumstellar dust to the mid-plane of the disk
2. Growth of planetesimals of the order of 1 km in size
3. Runaway growth of planetary embryos of the order of 1000 km in diameter
4. Growth of larger objects through late-stage collisions

Stage 1 would have taken no more than thousands of years and provided a dense plane of material from which planetesimals grew. Stage 2 is poorly understood but must have happened if sufficient mass were to accumulate for gravity to play a major role, as required by stage 3. Scientists have succeeded in growing in the laboratory centimeter-sized fluffy aggregates of dust held together with electrostatic and Van der Waals forces but do not know how to make something that is the size of a house, say. One suggestion is that some kind of glue was involved, but ices would not have condensed in the inner solar system. Another idea is that there was local sorting and clumping of gas and dust in the swirling disk, leading to regions of localized higher gravity.

However planetesimals first formed, runaway growth then transformed these ~1 km sized bodies into ~1000 km sized planetary embryos. The term “runaway” is used because the bigger the object the larger it becomes until all of the material available within a given heliocentric distance (‘feeding zone’) is incorporated. Models indicate that planetary accretion of this type took no more than a few hundred thousand years. The ultimate size depended on the amount of material available – even Mars-sized (but not Earth-sized) objects could have originated in this fashion.

How much nebular gas was present during these early growth stages is unclear. The presence of at least a small amount of nebular gas would have dampened eccentricities and would therefore explain the near-circular orbits of the planets. At one time it was thought that the planets were dominated by ‘planetary’ noble gases, probably derived from dust and different from the composition of the Sun or solar wind. However, noble gases with solar relative proportions and solar isotopic compositions have now been detected in the mantles of both Earth and Mars. Whether these solar noble gases were acquired by being dissolved into a magma ocean in the presence of a heavy nebular atmosphere, or were contributed to the Earth and Mars by the later accretion of planets and planetesimals is a matter of current debate.

George Wetherill (1986) ran Monte Carlo simulations of terrestrial planetary growth in the absence of nebular gas and succeeded in generating planets of the same size and distribution as Mercury, Venus, Earth and Mars. Most of the mass was accreted in the first 10 Myr, but significant accretion continued for up to 100 Myr. Wetherill tracked the provenance of material that built the terrestrial planets and showed that, in contrast to the runaway growth epoch, there was no localized feeding zone. His simulations suggested that the bits and pieces that built the Earth came from distances that extended over more than 2 AU. These planetary-scale collisions were catastrophic (Fig. 2) and would have raised the temperature of the Earth by thousands of degrees. The most widely held theory for the formation of the Moon is that a Mars-sized planet collided with the proto-Earth when the latter was approximately 90% of its current mass (Canup and Asphaug 2001). The putative impactor planet, sometimes named Theia (the mother of Selene, who was the goddess of the Moon), struck the proto-Earth with a glancing blow, accounting for the high angular momentum of the Earth–Moon system.
UNRAVELING CIRCUMSTELLAR DISKS AND EARLY PLANETARY GROWTH IN THE ISOTOPE LABORATORY

The same short-lived nuclides with half-lives of ≤100 Myr that yield so much information on the birth of the Sun also provide powerful tools for reconstructing the earliest history of the solar system (Fig. 3). The advantage of short-lived nuclides is that the changes in daughter isotope abundances took place over a restricted and definable early time window. Hence, for samples formed while the parent nuclide was still extant, high-precision analyses of the isotopic composition of the daughter can, in principle, result in high temporal precision. A disadvantage is that the parent isotope can no longer be detected, so its abundance at the start of the solar system must be determined by comparing the isotopic composition of the daughter element in rocks and minerals of independently known age. Only then can the isotopic compositions be calibrated with an absolute timescale. Determining the initial abundance is difficult, and some former nuclides still have uncertainties of more than an order of magnitude as a result of conflicting results or poor constraints. Time differences can be determined precisely without this information, however, as long as it can be safely assumed that the short-lived isotope was homogeneously distributed over the region of interest in the early solar system.

The events at the very start of the solar system provide dramatic examples of the information afforded by short-lived nuclides. The clues to the very earliest events are found in chondrites, a class of meteorite with compositions similar to that determined spectroscopically for the Sun’s photosphere. Chondrites are composed of a range of early debris, much of which had not been processed by melting in planetary objects. They come from asteroids that accumulated as rubble piles of this circumstellar debris. Many chondrites contain calcium–aluminum-rich refractory inclusions (CAIs), which are usually a few millimeters in size and composed of silicates and oxides. CAIs are enriched in the major and trace elements calculated to condense at high temperatures from a cooling gas of solar composition. Unlike presolar grains they have relative abundances of stable isotopes that are almost identical to those found in the terrestrial planets and meteorites. Therefore, they appear to have formed within our solar system but in a very hot environment, possibly very close to the Sun, and probably at a very early stage. They were then scattered across the swirling circumstellar disk of gas and debris that became the birthplace of the planets. Using $^{235}/^{238}\text{U}-^{207}/^{206}\text{Pb}$ the absolute ages of some CAIs have been determined to be 4.5672 ± 0.0006 billion years (Amelin et al. 2002). This is impressive precision, yet using $^{26}\text{Al}-^{26}\text{Mg}$ (Table 1) it has recently been shown that some CAIs formed within <100,000 years of each other (Bizzarro et al. 2004). A recent attempt to combine precise $^{235}/^{238}\text{U}-^{207}/^{206}\text{Pb}$ and precise $^{26}\text{Al}-^{26}\text{Mg}$ chronologies on differentiated meteorites has raised the question of whether the solar system may have started earlier than previously thought, at >4.5695 ± 0.0002 billion years (Baker et al. 2005). Reconciling these apparently conflicting results requires further work.

Most chondrites also contain chondrules – drop-shaped Mg-rich silicate objects with textures, thought by many to reflect rapid heating, melting and quenching of pre-existing dust in the circumstellar disk. Using $^{26}\text{Al}-^{26}\text{Mg}$ techniques it can be shown that many chondrules formed within 1 to 2 Myr of CAIs (Kita et al. 2005) (Fig. 3). Therefore, the disk must have persisted for at least 2 million years in the region where chondrules formed, unless they formed from secondary dust produced by destruction of pre-existing planetesimals. It also follows that chondrites – the primitive, undifferentiated meteorites that contain both CAIs and chondrules – must have formed at least a few million years after the start of the solar system. Though primitive, they are not the equivalents of the first planetesimals and planetary embryos predicted from theory.

TUNGSTEN ISOTOPE TESTS FOR TERRESTRIAL PLANET FORMATION

Hafnium-182 (Table 1) has been a particularly useful short-lived nuclide for studying planetary growth and differentiation for four reasons:

1. Its half-life (8.9 Myr) is ideal for evaluating the first 50 Myr of solar system history.

2. The initial abundance of $^{182}\text{Hf}$ was relatively high ($^{182}\text{Hf}/^{180}\text{Hf} = 10^{-4}$), producing isotopic effects that are easily resolved.

3. Hafnium and tungsten are fractionated by core formation, which is considered to be an early process in planetary development.

4. Hafnium and tungsten are both refractory, so the parent/daughter ratios and W isotope compositions of bulk planets are usually chondritic, and hence well defined.

Two recent reviews of $^{182}\text{Hf}/^{182}\text{W}$ isotope systematics, but with different emphases, are provided by Jacobsen (2005) and Halliday and Kleine (2006). Hafnium/tungsten ratios are changed by core formation because W tends to partition into metal whereas Hf remains in silicates. The silicate Earth has a $^{182}\text{W}$ atomic abundance that is high relative to chondrites, which represent average solar system (Kleine et al. 2002; Schönberg et al. 2002; Yin et al. 2002). Therefore, a portion of the (high Hf/W) silicate Earth formed during the lifetime of $^{182}\text{Hf}$. The isotopic effect on the $^{182}\text{W}/^{184}\text{W}$ ratio is only ~200 ppm, and the growth of the Earth must therefore have been protracted. If the Earth and its core formed within a million years of the start of the solar system, the isotopic effect would be close to 2000 ppm.

The more exact interpretation of this isotopic difference in terms of timescales is model dependent. In the simplest case, the atomic abundance of $^{182}\text{W}$ could date an instantaneous core formation at 30 Myr in a fully formed Earth (Fig. 4A). However, such a model is inconsistent with other
evidence for how the Earth formed (including Wetherill’s modeling and the Giant Impact hypothesis for Moon formation). With a half-life of 8.9 Myr, W isotopes will not record Hf/W changes >60 Myr after the start of the solar system. In principle roughly half the Earth’s core could have formed at the start of the solar system and the other half during the Archean or any other period of Earth history, and the W isotopic result would not be different from that measured. Geochemical evidence against such a protracted development of the core comes from the constancy of the relative concentration of certain key trace elements in basalts over time. Those elements that like to partition into core-forming metal and sulfide liquids should have decreased in the mantle in response to core growth. In fact they appear to have stayed approximately constant since the early Archean. Furthermore, the average Pb isotope compositions of the crust and mantle (or the combined bulk silicate Earth) plot close to the geochron determined by Patterson (1956) for the solar system. The geochron represents the line of equal age, corresponding to Pb isotope compositions of objects formed at the start of the solar system. The Pb isotope composition indicates that the bulk silicate Earth separated from the core in the first ~100 Myr of solar system history. Therefore, it is clear that the W isotope composition has to be interpreted in terms of a relatively simple early growth history of the Earth and its core.

A more realistic model may be that of a planet and its core growing over tens of millions of years as implied by the Monte Carlo dynamic simulations of Wetherill (1986) (Fig. 4a). The W isotope composition then is limited by the overall longevity of planetary growth. The simulations approximate to exponentially decreasing rates of growth, and on this basis Yin et al. (2002) calculated a mean-life of 11 Myr for the accretion of the Earth. The mean-life is the inverse of a time constant for exponential change. For example, in nuclear physics it is the inverse of the decay constant. In the context used here it corresponds to the time taken to achieve 63% planetary growth, and this is in excellent agreement with the timescales inferred directly from the simulations of Wetherill (1986).

FASTER ACCESSION OF SMALLER OBJECTS

Although the Earth accreted over a long period of time, the information from studying smaller planetary objects is in fact different. The most recent data for Martian meteorites (Foley et al. 2005) confirm earlier evidence that accretion and core formation on Mars were fast. Some recent models (Halliday and Kleine 2006) place the time required for for- mation of the core of Mars at <1 Myr (Fig. 3). If this is correct, Mars probably formed by a mechanism such as runaway growth, rather than by protracted collision-dominated growth. Mars may be an example of a large planetary embryo with a totally different accretion history from that of the Earth.

A similar story is being recovered from iron meteorites. Very precise W isotope measurements provide evidence that accretion and core formation were short-lived (Markowski et al. 2006; Scherstén et al. 2006). It can be demonstrated that some magmatic iron meteorites, thought to represent planetesimal cores, formed within 500,000 years of the start of the solar system, assuming this corresponds to the time calibrated by using the W isotope compositions of CAIs (Fig. 3) (Markowski et al. 2006). Therefore, they too appear to be examples of early planetary embryos, as predicted from dynamic theory.

THE ORIGIN OF THE MOON

Although the details vary, most Giant Impact models predict that the Moon formed after the Earth had achieved more than 80% of its current mass. The W isotope compositions of lunar samples provide support for the Giant Impact theory. The first attempts at these analyses revealed high $^{182}W$ atomic abundances in many lunar samples. Subsequently it was shown that in some cases, a portion of the $^{182}W$ was cosmogenic, formed from cosmic rays converting $^{181}Ta$ to $^{182}Ta$, which decays to $^{182}W$. This can be corrected using exposure ages or internal mineral systematics (Lee et al. 2002). Kleine et al. (2005) used native iron with low Ta/W from lunar basalts. The amount of excess $^{182}W$ that is not cosmogenic can be compared with the parent/daughter (Hf/W) ratios modeled for the lunar interior to determine a time at which the Moon formed. All of these approaches are consistent with an age for the Moon of 30–55 Myr after the start of the solar system. Such a late origin is not predicted.
Strictly speaking what is being dated in these studies is the differentiation of the lunar mantle as a high Hf/W reservoir in the lunar magma ocean. The earliest age of the Moon is provided by the time it is calculated to have last equilibrated with a chondritic object. This time is 30 Myr after the beginning of the solar system (Kleine et al. 2002; Yin et al. 2002). But the Moon was formed from the silicate reservoir of a differentiated planet (Canup and Asphaug 2001). The lunar initial W isotope composition provides an alternative constraint of ~340 Myr after the start of the solar system (Halliday 2003). This is more consistent with the 207/206 Pb age of Pb loss from the material that formed the Moon (Hanau and Tilton 1987).

CORE FORMATION, ACCRETION AND THE EARLY EARTH – THINKING BEYOND THE STANDARD MODELS

Theories of core formation were originally based on the concept of metal segregation from a fully formed Earth. The models range variously from intergranular percolation (Rushmer et al. 2000) to descending diapirs (Stevenson 1990). Similarly, the models upon which we base our many ideas of partitioning of trace elements in a magma ocean assume a fully formed Earth undergoing metal segregation. These models form the backbone of thinking about the physical and chemical processes by which the Earth’s core was formed. The challenge is now to develop them into more sophisticated theories that take account of the concurrent increase in size of the Earth and mass transfer processes as accretion and core formation proceeded over tens of millions of years (Figs. 4b and c). This is partly a question of tracking the phase changes, reactions and changes in physical properties as the Earth increased in size (Wade and Wood 2005).

What is harder to quantify, however, is the direct effect of planetary collisions. Impacts may have produced magma oceans that facilitated core formation, but they may also have blown off blanketing proto-atmospheres, leading to cooling and crystallization (Ahrens 1990). Furthermore, collisions between differentiated planetary bodies may have resulted in the metal in each planetary core simply merging (Yoshino et al. 2003). With mounting evidence of very early asteroidal melting and core formation, this could have been the norm. The growth of the Earth’s core would then have been via core–core mixing (Fig. 4c), rather than by metal segregation from silicate.

The core–core mixing process would affect $^{182}$Hf–$^{182}$W chronometry. In calculating accretion timescales, it is assumed that accreted W equilibrated isotopically with W in the silicate Earth (Fig. 4b). But if a fraction of the incoming W was in metal that mixed directly with the Earth’s core (Fig. 4c), then the $^{182}$Hf–$^{182}$W “age” of the Earth or its core would appear older than it really is (Halliday 2004). It is essential to understand the details of these processes. As the planet and its core fragment during impact, Rayleigh–Taylor instabilities will form. These are disruptions caused by layers of lower density and viscosity rising and impregnating overlying material. We need to know the size distributions of the resultant droplets of metal in silicate and silicate in metal before a realistic assessment can be made of the degree and processes of equilibration. We still have a primitive understanding of what happens when planets tear each other apart.

An alternative approach is to use isotope geochemistry to infer the physics. If we know the age of the Moon, we also know the timing of the last major accretion stage in the history of the Earth. From this we can determine the extent to which the W isotope composition of the silicate Earth reflects planetary-scale disequilibrium of material added during accretion. Preliminary attempts at this (Halliday 2004) show that on average the degree of equilibration between incoming metal and the silicate Earth must be high (70–90%) and would be even higher if the Moon formed before 45 ± 5 Myr after the start of the solar system. Therefore, full equilibration with incoming material has been the norm. However, a very large event, such as the Giant Impact, could have resulted in considerable disequilibrium, with between 40 and 60% of the incoming W in metal not equilibrating with the silicate Earth, if at other times metal and silicate were fully equilibrated.

DIFFERENT TIMESCALES FOR DIFFERENT CHRONOMETERS

Following the classic study of Patterson, several scientists tried to estimate the average Pb isotope composition of the bulk silicate Earth and found that it lies below and to the right of the slope defined by the geochron. In other words the classic piece of evidence that the Earth and solar system all formed at the same time has revealed evidence of a relatively small time difference as more data have been acquired. The standard explanation for this is that the Earth or its core formed somewhat later than the first objects in the solar system. As such, Pb and W are in qualitative agreement. In fact, since both U/Pb and Hf/W are fractionated by core formation, the W and Pb isotope compositions of the silicate Earth should yield the same protracted timescales if modeled in the same manner. Instead, the timing based on W appears faster than that based on any of the estimates of the Pb isotope composition of the silicate Earth (Halliday 2003). Given the difficulty with defining a meaningful average, it is very possible that all the Pb isotope estimates are wrong. The silicate Earth must then have a higher $^{207}$Pb/$^{206}$Pb than currently estimated.

It is also very likely, however, that the rates of transfer of W and Pb to the core were different (Wood and Halliday 2005). Tungsten is partitioned into metal, whereas Pb is thought to be more readily partitioned into sulfide liquids, which would not have become stable until the Earth cooled following the Giant Impact. Removal of this sulfide to the core may have been responsible for a late-stage increase in U/Pb in the silicate Earth that over time generated the observed Pb isotope compositions. Very roughly speaking, the Earth’s upper mantle appears to have cooled from temperatures of about 7000K at the time of the Giant Impact to about 3000K, when sulfides would have become stable, in tens of millions of years (Wood and Halliday 2005).

LINKS TO HADEAN MANTLE DEPLETION, CRUST FORMATION AND HABITABILITY

There must be overlap between the events discussed above and the mounting evidence of mantle differentiation events recorded in other studies. Those links are unclear because the models are under-constrained and some of the data are open to alternative interpretation. This will change as more data are acquired. The Hadean zircon archive is particularly exciting. For example, Wilde et al. (2001) reported a $^{235/238}$U–$^{207/206}$Pb age of a terrestrial zircon that formed in the first 200 Myr of the solar system and argued that it provided evidence of continental crust and a hydrosphere. Watson and Harrison (2005) used Ti geothermometry to show that Hadean zircons crystallized at temperatures consistent with water-saturated crustal melting, which implies...
that a regulated mechanism of melt production was in place. Harrison et al. (2005) have argued that mantle depletion by partial melting probably linked to crust formation occurred within the first 70 Myr, based on $^{176}\text{Lu} - ^{176}\text{Hf}$ isotope data from these early zircons.

These new Hf data display more variability than the previously reported relatively precise measurements of later Hadean zircons (Amelin et al. 1999) but in broad terms are consistent with the modeling of Caro et al. (2003), who used the $^{146}\text{Sm} - ^{142}\text{Nd}$ chronometer (Table 1) to estimate that the mantle’s earliest melt depletion event occurred in the first 220 Myr of the solar system. The main system is the proposal of Boyet and Carlson (2005), who used $^{146}\text{Sm} - ^{142}\text{Nd}$ systematics to argue that a hidden reservoir in the Earth’s mantle formed within the first 30 Myr. However, these data could reflect a slight difference in Sm/Nd between the Earth and chondrites or even minor nucleosynthetic effects.

REFERENCES


Hartmann L (2000) Observational constraints on transport (and mixing) in pre-main sequence disks. Space Science Reviews 92: 55-68


