A Strategy for Understanding Planet Formation

Thematic Areas: ✅ Star and Planet Formation

Principal Author:
Name: Alycia J. Weinberger
Institution: Department of Terrestrial Magnetism, Carnegie Institution for Science
Email: aweinberger@carnegiescience.edu
Phone: 202-478-8852

Co-authors: Neal Turner (JPL), Yasuhiro Hasegawa (JPL)

Co-signers: John Debes (STScI), Thomas Haworth (Imperial College London), Andrea Isella (Rice University), Bruce Macintosh (Stanford), Dimitri Mawet (Caltech), Karin Oberg (Harvard), Ryan Loomis (NRAO), Nicholas Ballering (University of Arizona)

Abstract: An understanding of the physical processes that lead to the panoply of observed planets is an undertaking worthy of the next generation of facilities and astrophysicists. The diversity of exoplanets has reinvigorated the field of planet formation with the scientific goal to connect the raw materials for planets to the final outcomes. A pan-chromatic approach including observations from X-rays through radio is required to understand the diverse processes that determine where planets form in circumstellar disks, how material is transported through the disk and into planets, and how disks evolve in concert with their stars and planets.
Understanding planet formation is an endeavor at the heart of the astrophysics because it is the science of determining the physical processes that occur to create the Solar System, the population of exoplanets, and exoplanetary system architectures. While the census of exoplanets is not yet complete, it is clear that planets come in an astounding variety and that the solar system may not be representative. Exoplanets as the outcomes of planet formation provide a zoo of endpoints, and while characterizing that menagerie is a thrilling subject in its own right, the true scientific goal will be to connect outcomes with progenitors through the underlying physics. While the 2018 National Academies’ Exoplanet Science Strategy provided an excellent road-map for studying the zoo, it did not provide a comprehensive strategy for studying formation. Here, the goal is to outline a strategy for planet formation per se.

Planet formation begins as the leftover gas and dust from star formation settles into a disk rotating about the central star. Fantastic efforts to study disks have been made over the last 40 years, since the imaging of the first disks (see Figure 1). Our general picture is that, after the central star has formed from its natal cloud via gravitational fragmentation and collapse, the disk forms as angular momentum and magnetic fields support material against accretion onto the central star. Ongoing accretion onto the star necessitates removal of angular momentum, which probably occurs via winds and jets. Until the cloud fully accretes or is blasted away by photoevaporation, clumps of material continue to accrete onto the disk and from the disk onto the star, causing bursts of activity such as FU Orionis events. Accretion must be stochastic because the accretion rates, measured by the energy released by the gas as it falls onto the star, are too slow to build up the stellar mass in the time when the disk persists. Extensive studies of disk lifetime are based on studying clusters of stars and measuring the fraction of stars that display disk emission and the quantity of that emission. Disks appear to be ubiquitous around young stars, with >80% of stars at ages <1 Myr showing disk emission. This emission disappears, at both short and long wavelengths, so that half of stars have left their disks by ∼3 Myr. Accretion disappears on the same timescale, suggesting that disk dissipation is in both gas and dust.

At very high rates of infall, the disk may start to form, and solids will condense as soon as the gas becomes cool enough, ∼1500 K. In the Solar System, the oldest solids are the calcium aluminum rich inclusions found in meteorites, which set the age of the Solar System. How exactly this age aligns with the age of the Sun, or more generally how the timing of the first condensed solids aligns with stellar ages is not precisely known, and may occur during the protostar phase. During disk evolution, in what we call the protoplanetary phase, the dust must somehow grow into planetesimals that eventually form terrestrial and super-earth sized planets. Dust may also grow to form the massive (>10 M⊕) cores of giant planets. Planets will inherit their compositions during formation, and planetary architectures will be established by dynamical interactions.

Here, we present major questions in planet formation and the pan-chromatic approach that is required to answer them. All the questions would benefit from being studied in a spatially resolved sense to distinguish the ice-rich outer disks (>5 AU) from warm inner disks. At the distances to the closest sites of ongoing planet formation at ∼140 pc, this requires a spatial resolution better than ∼35 mas. While resolving lines with high spectral resolution observations can to some extent make up for poorer intrinsic spatial resolution, our goal should be observations at the level needed to separate outer and inner disk processes.

Here, we can consider four intertwined processes that must be addressed to understand planet formation.

1 What factors determine the locations in disks where planets form?

Solar system planetesimals formed within 1 Myr of the first mm-sized solids, so these formation processes must be efficient. There is consensus within the planet-formation community that pair-wise dust collisions cannot build planetesimals, as collisions cause mm-size grains to shatter rather than grow. Therefore, turbulent concentration and gravitational coagulation are attractive for speeding along the growth process for planetesimal and planet formation (e.g. Johansen et al., 2007).

**Disk Mass:** Initial disk mass, spatial distribution, and dissipation timescale will control the mass of planets formed. Traditional means of estimating disk mass from submillimeter emission depend on knowing the extent of grain growth and thus the efficiency of emission of grains of unknown size. Estimates tend to give low masses, below that of the minimum mass solar nebula for most stars (e.g. Cieza et al., 2015; Eisner et al., 2018). If the masses are to be believed, this implies that disk masses must be built up early and planets must form quickly, before the disks are revealed from their infall cocoons to have their masses measured. Most of the mass is in gas rather than dust, but molecular hydrogen is notoriously difficult to observe due to its lack of a ro-vibrational dipole moment. Abundant CO can be observed, but freezes out onto grains.
or is dissociated, so that extensive chemical models must be employed to turn CO and its isotopologues into mass estimates. A simpler route to disk mass is via far-infrared measurements of HD (the deuterated form of molecular hydrogen) and models of disk temperature structure; however, the Herschel Space Telescope was sensitive enough to make only three such measurements. Clearly we need a route to estimate the masses of early disks. Because envelopes are optically thick in the infall phase, and the midplanes of disks are optically thick throughout planet formation, disk masses must be inferred with long wavelength observations in optically thin tracers. HD is the most direct, but others are being explored with ALMA. It seems likely that we need more capabilities than ALMA presently provides. Because HD emits at 112 µm, only instruments above the Earth’s atmosphere can observe it efficiently. SOFIA’s new HIRMES instrument will make progress on the youngest most massive systems, but a cold space telescope such as the Origins Space Telescope will be necessary for the bulk of disks. Long wavelength gas tracers will require new ALMA receivers and the ngVLA.

■ Ice-lines: Condensation of ices can lead to increased surface density, and the location of the water ice line was originally hypothesized to be associated with the location of Jupiter in the Solar System. However neither temperature nor density structures in disks are likely to be static, and the role of ice lines in generating stable surface density or pressure structures (see the next section) is unclear. The wavelength region of 100-500 µm traces the most important water lines, covering gas temperature of 10-1000 K, and can only be done in a large sample with a cold space telescope such as the Origins. Otherwise, the location of ice lines must be deduced from observations of trace molecules in chemical networks that operate once the ice has condensed (e.g. Qi et al., 2013), which can be done by ALMA and ngVLA.

■ Turbulence: Fast grain growth will require high midplane disk masses with a certain amount of turbulence. Turbulence plays a big role in disk evolution, not only in planetesimal formation but also in controlling mass flow through the disk, and the amount of vertical mixing from midplane to surface of the disk. Gas turbulence may be measured through the broadening of otherwise narrow spectral lines, but most attempts have only arrived at upper limits. There is a difficult observational competition between having enough flux for a good measurement but choosing a line that isn’t thermally broadened (Flaherty et al., 2018). ALMA has succeeded in placing upper limits on outer disk turbulence. It is likely that future ALMA and ELT observations at high spectral resolution make some progress, but new facilities such as ngVLA are also needed.

2 How is material transported through the disk and into planets?

■ Transport and Drift: Determining the mass flow through disks, i.e., migration or drift, has been a persistent problem for planet formation theory, which says that grains larger than mm-sized (i.e., pebbles) will face a gas headwind that causes them to lose angular momentum and spiral into the central star on timescales much faster than the lifetime of the disk (Weidenschilling, 1977). Planetesimals face turbulence-induced velocities that can lead to their collisional destruction (Nelson & Gressel, 2010), and planets face gas-induced migration (see text on migration in Section 3). There is now a vast theoretical literature on migration, with few observational constraints. Migration rates depend on the gas:dust ratio and radial pressure structure of the disk. ALMA images of disks with dust ring structures suggest that radial pressure bumps can confine dust of certain sizes (e.g. Pinilla et al., 2012; van der Marel et al., 2015). However, how
these pressure maxima are created and whether they are sustained is an open question. To understand radial drift, we must measure the gas pressure and turbulence in a radially resolved manner, and the gas:dust ratio as close to the midplane as can be done. Both of these will require a panchromatic approach covering warm gas at the near to mid infrared (JWST and ELTs), small dust grains in scattered light (high performance coronagraphs on ELTs or in space), larger dust grains and cooler gas (long-wavelength arrays) and all with spatial resolution of a few AU.

Initial and evolving disk structure probably dictates the migration physics of pebbles, planetesimals and planets, and how much material of different chemical abundance is accreted, dispersed, and made into planets (e.g. Lambrechts et al., 2014; Sato et al., 2016). The sequence of formation of ever-larger bodies within the disk can change the supply of material to still-forming bodies. We have yet to determine whether the Solar System sequence of planetesimals, to cores of giant planets, to terrestrial planets is the most common pathway. If giant planets form rapidly, they can choke off the supply of large grains and, at large enough mass, gas, to the inner disk. A key question is whether condensation fronts generate the radial pressure gradients that allow giant planets to form, whether this happens quickly so that the inner disk is always depleted in volatiles from the outer disk. So far, the rings observed in ALMA images do not seem to correspond to condensation locations (Huang et al., 2018), but the evolution of radial structures may have been significant by the times at which they are being observed. It is imperative to study disks as they form, when the condensation fronts are being established.

■ Chemistry: Whether a rocky core accretes enough gas and/or ice to become a mini-Neptune or giant planet depends on the distribution of volatiles in disks. Whether a terrestrial planet acquires water and carbon similarly depends on disk processing. Measurements of gas motion can constrain the amount of radial and vertical mixing. Because different wavelengths probe different temperature areas of the disk, we must combine observations across a wide range of wavelengths. For example, large ground-based telescopes will measure the gas motions in the inner disk, JWST will measure the warm molecular layer, Origins the molecular layers below the surface, ALMA the cool outer disk and molecules in and below the outer disk surface, and ngVLA the deepest molecular layers. Theoretical work on the chemical networks between these layers will be needed to connect the observations. In all cases, we will need sensitivity to optically thin lines of new molecular tracers and high spatial resolution.

Finally, the compositions of planets will depend on how chemistry is modified through planetesimal parent body processing. Some handle on this stage can be achieved by studying the material released from planetesimals into debris disks.

3 How do disks evolve?

Material not incorporated into planets must either accrete onto the central star or be removed through outflows (winds). The central stars may drive photoevaporation, natal planets may sculpt their disks, and stellar-disk-planet interactions may determine the ultimate fate of the original disk dust and gas.

■ Migration: The radial drift problem that afflicts our understanding of growth to planetary embryo sizes becomes the migration problem for planetary cores, when Type 1 migration should cause rapid inward movement, and for planets massive enough to open gaps, when Type 2 migration causes the planet to move inward with gas accretion. Physical constraints on the in-
interaction can come in the form of imaging of disk gaps, warps, and spirals created by planets and by studying the non-Keplerian motion of gas. Hydrodynamic simulations predict that gaps and spiral arms can be caused by planets embedded in and interacting with the disk (e.g. Kley & Nelson, 2012; Dong et al., 2015b). The appearance of gaps and spiral arms can also reveal planet properties (e.g. Debes et al., 2013; Dong et al., 2015a; Fung et al., 2014) or indicate gravitational instability (e.g. Pérez et al., 2016; Dong et al., 2018). A range of wavelengths is again important. Optical to near-IR imaging highlights structure and also the scale height by resolving shadowing. Thermal emission over the full range of inner to outer disk wavelengths, from 10 \( \mu \text{m} \)– mm wave, can reveal the surface density of material. If planets can be directly imaged within disks, their masses (as inferred from their luminosities) can be compared to the disk structure (e.g. \( \beta \) Pic b, Lagrange et al. (2012)) to diagnose the disk physics including density and turbulence.

The solar system planets may have undergone migration during the gas phase (Walsh et al., 2011) as well as late migration driven by planetesimals (Gomes et al., 2005). Observations of debris disks are necessary to determine the late stages of setting planetary system architectures.

- **Winds and Outflows:** Photoevaporation models (e.g. Clarke et al., 2001) have had some success in reproducing disk lifetimes and, combined with viscous accretion models, explaining why there are few disks with inner holes, i.e., transitional disks, observed. Observations of disk fractions in star forming regions are needed to evaluate the relative importance of external photoevaporation, from nearby massive young stars, compared to each star’s own high energy radiation. Furthermore, the predictions of photoevaporation models need to be tested against better estimates of disk mass (see previous sections) over time and better observations of the amount of escaping material over time. While disk winds have been observed in forbidden emission lines (Rigliaco et al., 2013) and submm molecular lines (Güdel et al., 2018), it is not yet clear if they can carry away enough mass to account for short disk lifetimes. Nor is it clear how the winds are tied to turbulence and accretion (Simon et al., 2018). In most observations, kinematic resolution has substituted for spatial resolution in determining the launching points of jets and winds, and there is a need to add high spatial resolution observations in both the visual and submm. These observations of outflowing gas need to be correlated with stellar activity and accretion rates as traced in the X-ray and UV. Studies of the high energy emission of stars over time may help explain why high mass and low mass stars both appear to retain their protoplanetary disks longer than Sun-like stars. Efficient planet formation may play an important role in disk dissipation by changing the structure of the disk and locking mass into bodies that can quickly spiral inward. So, disk dissipation is strongly tied to the problems outlined in previous sections of this white paper.

4 Conclusions

A pan-chromatic approach including observations from X-rays through radio is required to understand planet formation. Nevertheless, the very high high mid-plane densities and need to study the molecular (e.g. HD and water) and ice content of disks at high spatial resolution imply that the the most useful future facilities will be those that observe at long wavelengths (beyond JWST). Spatial resolutions <35 mas to separate inner (ice free) from outer (icy) regions will be essential for deducing the mixing and chemical complexity that results in a wide variety of planets.
References