Planet formation in evolving protoplanetary discs

Richard Alexander
Department of Physics & Astronomy, University of Leicester, Leicester, LE1 7RH, UK
email: richard.alexander@leicester.ac.uk

Abstract. I attempt to summarize our knowledge of planet formation in evolving protoplanetary discs. I first review the physics of disc evolution and dispersal. For most of the disc lifetime evolution is driven by accretion and photoevaporation, and I discuss how the interplay between these processes shapes protoplanetary discs. I also discuss the observations that we use to test these models, and the major uncertainties that remain. I will then move on to consider planet formation and migration in evolving discs, and discuss how observations of both discs and planets can be used to inform our understanding of protoplanetary disc evolution.

Keywords. protoplanetary discs, planet–disc interactions, planetary systems

1. Introduction

We now know that many, if not most, stars host planetary systems, but their origins remain a mystery. Understanding the evolution of protoplanetary discs is a critical ingredient of any predictive theory of planet formation. Protoplanetary discs are the sites of planet formation, but planet formation is an inefficient process (only a small fraction of the initial disc mass ends up in planets) and consequently discs dominate the dynamics of forming planetary systems. Here I review our understanding of the physics of protoplanetary disc evolution and dispersal, and its consequences for the architecture of planetary systems.

1.1. Disc evolution

The existence of gas-rich protoplanetary discs was first suggested as long ago as the work of Kant & Laplace, but such discs were not directly observed until the late 1980s (e.g., Sargent & Beckwith 1987). Since then observations of discs have progressed at a startling rate, and we now have a census of thousands of these objects. Discs are observed across a wide range of wavelengths, using both gas and dust tracers, and this has allowed us to build up a detailed picture of their properties. Observations of young clusters find that the fraction of stars with an infrared excess (which traces warm dust at ~AU radii) declines from ~100% at ages \(\lesssim 1\) Myr to \(\lesssim 10\%\) at ages of 5–10Myr (e.g., Haisch et al. 2001; Mamajek 2009). A similar trend is seen in signatures of accretion on to the stellar surface (e.g., Fedele et al. 2010), which trace gas in the inner disc, and in (sub-)mm continuum emission, which traces cold dust at large (~100AU) radii (e.g., Andrews & Williams 2005). Although age estimates for young stars, which are typically derived from comparing observations to pre-main-sequence evolutionary tracks, are subject to significant uncertainties (Hillenbrand 2009; Bell et al. 2013), from these data we can conclude that typical disc lifetimes are a few Myr, with order-of-magnitude scatter.

In addition to statistical studies of discs in clusters, these observations also allow us to determine global properties of individual discs. By making standard assumptions about the dust opacity and dust-to-gas ratio we can estimate disc masses from the (sub-)mm
luminosity: typical disc masses range from $M_d \sim 0.1M_\odot$ to $\lesssim 0.001M_\odot$ (e.g., Andrews & Williams 2005). We can also derive instantaneous accretion rates, by measuring the excess short-wavelength emission produced by the accretion shock on the stellar surface; if we know the stellar mass and radius, we can convert an observed accretion luminosity into a mass accretion rate (e.g., Calvet & Gullbring 1998). Accretion rates measured in this manner (for $\sim$ solar-mass stars) range from $\dot{M}_{\text{acc}} \gtrsim 10^{-7}M_\odot\text{yr}^{-1}$ to $\lesssim 10^{-10}M_\odot\text{yr}^{-1}$ (e.g., Gullbring et al. 1998). We can therefore estimate an empirical accretion timescale $t_{\text{acc}} \sim M_d / \dot{M}_{\text{acc}}$, and find that for most sources $t_{\text{acc}}$ is of order Myr, again with large scatter (e.g., Jones et al. 2012). The accretion time-scale is approximately the time it will take all of the observed disc mass to be accreted (at the observed accretion rate), so we expect protoplanetary discs to evolve substantially due to accretion during their $\sim$Myr lifetimes. This in turn implies that there is no such thing as “typical conditions” for planet formation; planets form in evolving protoplanetary discs, and the ubiquity of planets suggests that they form under a wide range of different conditions.

These observations show that protoplanetary discs evolve, at least in part due to disc accretion, but shed little light on the physical processes driving this evolution. The decline in observed accretion rates with stellar age is broadly consistent with classical viscous accretion disc evolution (e.g., Lynden-Bell & Pringle 1974; Hartmann et al. 1998), but the origin of the required angular momentum transport remains unclear (see, e.g., the review by Armitage 2011). At early times, when the disc is still embedded in its parent envelope, disc evolution is probably dominated by gravitational instabilities (e.g., Lodato 2007). Magnetically-driven jets and winds can also remove angular momentum and drive accretion in the disc, particularly at small radii, and may remain important throughout the disc’s evolution (e.g., Königl & Salmeron 2011). However, for most of the disc lifetime the dominant source of angular momentum transport is thought to be magnetohydrodynamic (MHD) turbulence, driven by the magnetorotational instability (MRI: Balbus & Hawley 1991; Balbus 2011). It remains unclear how much of the disc is sufficiently ionized to sustain the MRI, and it seems likely that substantial “dead zones”, with little or no turbulence, exist in the majority of discs (e.g., Gammie 1996; Bai 2011). Recent work also suggests that MRI-driven winds may play a role in disc evolution, potentially driving significant mass- and angular momentum-loss over a wide radial extent (e.g., Fromang et al. 2013; Bai & Stone 2013). Thus, although simple viscous disc models are sufficient to explain most observations of protoplanetary disc accretion, our ignorance of how discs accrete remains a major obstacle to a complete theory of planet formation.

1.2. Disc dispersal

Observations of young stars also allow us to infer a great deal about disc dispersal. Different disc tracers are strongly correlated, despite originating in very different regions of the disc, and tend to vanish together; few objects are seen with “partial” discs (e.g., Andrews & Williams 2005). This suggests that disc dispersal is almost simultaneous across a wide range of radii, from $\lesssim 0.1AU$ to $>100AU$. Moreover, the relative lack of objects with properties between those of disc-bearing and disc-less stars suggests that dispersal is rapid: statistical estimates suggest that the dispersal time-scale is at least an order of magnitude shorter than the disc lifetime (e.g., Simon & Prato 1995; Koepferl et al. 2013). Finally, non-detections of warm or hot gas around disc-less weak-lined T Tauri stars set very strict upper limits on the gas surface density (orders of magnitude below canonical protoplanetary disc values), implying that disc dispersal is very efficient (Pascucci et al. 2006; Ingleby et al. 2011). Disc evolution therefore has two distinct time-scales (the $\sim$Myr disc lifetime and the much shorter disc dispersal time), and the
process(es) which drive final disc clearing must operate very efficiently over the entire radial extent of the disc.

These observational constraints rule out several obvious candidates as mechanisms for disc dispersal. Resolved observations of disc structure suggest that most of the disc mass resides at large radii ($\gtrsim 50$AU; Andrews et al. 2009), but the characteristic viscous time-scale is $\gtrsim \text{Myr}$ at these radii. Discs therefore live for no more than a few viscous time-scales in their outer regions, and consequently accretion cannot play a major role in outer disc clearing. Similarly, winds or jets driven by the stellar magnetic field cannot be the dominant disc clearing mechanism at large radii, though MRI-driven winds may play a role. However, in most discs the primary driver of final disc dispersal is thought to be disc photoevaporation (see, e.g., Clarke 2011; Alexander et al. 2014).

2. Disc evolution models

2.1. Disc Photoevaporation

Photoevaporation occurs when high-energy (UV and/or X-ray) radiation is incident on protoplanetary discs. This irradiation creates a hot ($\sim 10^3$–$10^4$K) layer on the disc surface, and beyond a certain radius the heated gas is unbound. The result is a pressure-driven flow, which is referred to as a photoevaporative wind. The characteristic critical radius depends on the temperature of the hot gas, and is approximately

$$R_c \simeq \frac{0.2GM_*}{c_{\text{hot}}^2} \simeq 1.8 \left( \frac{M_*}{1\,\text{M}_\odot} \right) \left( \frac{T_{\text{hot}}}{10^4\text{K}} \right)^{-1} \text{AU},$$

(2.1)

where $M_*$ is the stellar mass, and $c_{\text{hot}}$ and $T_{\text{hot}}$ are the sound speed and temperature in the heated surface layer, respectively (Hollenbach et al. 1994; Liffman 2003; Font et al. 2004). The irradiation can be both “internal” (i.e., from the central star) or “external” (from nearby massive stars); external heating dominates in some cases (e.g., the proplyds in the centre of the Orion Nebula Cluster; Johnstone et al. 1998), but for the majority of discs we expect photoevaporation to be dominated by irradiation from the central star.

The temperature of the surface layers depends primarily on the irradiation, and three regimes are important for protoplanetary discs: ionizing, “extreme” UV (EUV; 13.6–100eV); non-ionizing far-UV (FUV; 6–13.6eV) and X-rays (0.1–10keV):

**EUV:** In the EUV case heating and cooling is dominated by ionization and recombination of atomic hydrogen, and for an optically thick disc ionizing photons from recombinations in the disc atmosphere dominate the heating at the ionization front (Hollenbach et al. 1994). The resulting disc atmosphere is close to isothermal at $T_{\text{hot}} \simeq 10^4$K (akin to an H II region on the disc surface), and most of the mass-loss originates at 1–2AU. As the flow is recombination-limited the integrated mass-loss rate $\dot{M}_w$ scales as the square-root of the ionizing luminosity, and for fiducial parameters we find $\dot{M}_w \sim 10^{-10}M_\odot\text{yr}^{-1}$ (Hollenbach et al. 1994; Font et al. 2004). However, if the inner disc is optically thin to ionizing photons (as expected during disc dispersal) then direct irradiation of the inner disc edge dominates. In this case most of the mass-loss comes from close to the inner edge, and $\dot{M}_w$ increases by a factor of $\sim 10$ (Alexander et al. 2006a).

**FUV:** FUV irradiation is analogous to the well-studied problem of photodissociation regions (PDRs): photons are predominantly absorbed by either dust grains or polycyclic aromatic hydrocarbons (PAHs), and the gas is heated by a combination of collisions with grains and FUV-pumping of H$_2$. The resulting gas temperatures vary substantially with both radius and height above the disc midplane, from $\sim 100$K at large radii to $\gtrsim 1000$K close to the star (Adams et al. 2004; Gorti & Hollenbach 2008, 2009).
Figure 1. TW Hya [Ne i] 12.81 μm line profile: the black line shows the observed line profile from Pascucci & Sterzik (2009), while the red and blue curves show the predictions from photoevaporative wind models driven by stellar EUV (Alexander 2008b) or X-ray (Ercolano & Owen 2010) photons, respectively. Although the data do not yet distinguish between these models, the blue-shifted line represents a clear detection of a slow, ionized wind, and provides unambiguous evidence of on-going photoevaporation of the TW Hya disc.

resulting mass-loss profile has a peak at 5–10AU, and also rises towards the disc outer edge (≳100AU); in some cases the mass-loss at large radii dominates. The integrated wind rate depends mainly on the FUV luminosity incident on the disc, and for fiducial parameters \( \dot{M}_w \sim 10^{-8} M_\odot \text{yr}^{-1} \) (Gorti & Hollenbach 2009).

**X-rays:** Young solar-mass stars have long been known to be bright X-ray sources, with typical luminosities \( L_X \sim 10^{30} \text{ergs}^{-1} \) (e.g., Feigelson et al. 2007). X-rays are primarily absorbed by K-shell ionization of heavy elements, and the resulting \( \sim \text{keV} \) photoelectrons collisionally ionize and/or heat the disc gas. As in the FUV case, the disc atmosphere is not well characterised by a single value of \( T_{\text{hot}} \), but hydrodynamic models find that most of the resulting wind originates in the atomic layer, at temperatures of 3000–5000K (Owen et al. 2010), and consequently the mass-loss profile peaks at around 3AU. The mass-loss rate scales close to linearly with \( L_X \), and for typical parameters we again find \( \dot{M}_w \sim 10^{-8} M_\odot \text{yr}^{-1} \) (Owen et al. 2011, 2012).

These different “flavours” of photoevaporation models all predict mass-loss at rates that are large enough to play a significant role in disc evolution, but it is not clear which heating mechanism dominates in real discs. Some theoretical issues remain, but the dominant uncertainty in these models is the irradiating flux which reaches the disc surface (which is an input parameter in the models). Recent work has attempted to resolve this question observationally, by looking for emission line diagnostics which trace the wind structure. Photoevaporative winds are characterised by relatively low-density gas which is at least partially ionized, and are therefore expected to produce copious forbidden line emission (Hollenbach & Gorti 2009). Hydrodynamic models show that the characteristic signature
Figure 2. Surface density evolution of a characteristic viscous/photoevaporation disc model, colour-coded to denote the three stages of the disc’s evolution. For most of the disc’s lifetime (few Myr, blue curves) photoevaporation is negligible and the disc follows the behaviour of a classical viscous accretion disc. However, once the disc accretion rate drops sufficiently photoevaporation is able to open a gap in the disc at \( \approx R_c \), after which the inner disc rapidly accretes on to the star (\(~10^5\) yr, green curves). Once the inner disc becomes optically thin, photoevaporation clears the disc from inside out (few \( \times 10^5 \) yr, red curves). [Figure adapted from Alexander (2008a), using the median disc model of Alexander & Armitage (2009).]

of such a wind is a small blue-shift in the emission from face-on discs (Font et al. 2004), and the \([\text{Ne} \text{II}]\) 12.81\,\mu m line is arguably the most promising such diagnostic (Alexander 2008b). In this case the line originates close to the base of the wind, and models predict small blue-shifts (\( \Delta v \sim 5-10\,\text{km}\,\text{s}^{-1} \)) for both EUV & X-ray heated winds (Alexander 2008b; Ercolano & Owen 2010). This blue-shift was first detected in the TW Hya disc (Pascucci & Sterzik 2009, see Fig. 1), and has subsequently been seen in around a dozen other sources (Pascucci et al. 2011; Sacco et al. 2012; Baldovin-Saavedra et al. 2012). The observed line profiles do not yet distinguish between the different photoevaporation models (see Fig. 1), but these observations represent an unambiguous detection of a slow wind which is at least partially ionized. When combined with line profiles from other species (such as \([\text{O} \text{I}]\), e.g., Rigliaco et al. 2013), or other wind diagnostics such as free-free emission (Pascucci et al. 2012), these observations offer the prospect of measuring photoevaporation rates empirically.

2.2. Secular evolution

These photoevaporative mass-loss rates, inferred from both theory and observations, are comparable to the observed accretion rates for many T Tauri stars, and are clearly large enough to play a role in the later stages of disc evolution. However, in order to understand how protoplanetary discs evolve we must consider accretion and photoevaporation simultaneously. This was first done by Clarke et al. (2001), who coupled EUV photoevaporation to a 1-D viscous accretion model. Models of this type have subsequently been
refined and improved by a number of different groups (e.g., Alexander et al. 2006b; Gorti et al. 2009; Owen et al. 2010; Bae et al. 2013), and although the details depend on the choice of model parameters and input physics, all photoevaporation/viscous evolution models display the same characteristic behaviour. Initially the accretion rate is much higher than the photoevaporation rate and the disc evolves viscously, accreting mass inwards and spreading at large radii. However, as the disc mass is finite the disc accretion rate declines with time, and after a few Myr the accretion rate becomes comparable to the rate of photoevaporative mass-loss. At this point the wind opens a gap in the disc at \( \simeq R_c \), and once photoevaporation overcomes accretion the inner disc (inside \( R_c \)) is cut-off from re-supply and drains on to the star on its (short) viscous time-scale (\( \sim 10^5 \)yr). Once the inner disc has been drained photoevaporation rapidly clears the outer disc, and the entire disc is dispersed on a time-scale of a few \( \times 10^5 \)yr. This three-stage evolution scenario (illustrated in Fig. 2) is consistent with disc observations across a wide range in wavelength (e.g., Alexander et al. 2006b), and successfully reproduces the two-time-scale behaviour demanded by observations. We note also that this type of inside-out clearing inevitably gives rise to a short “inner hole” phase, and represents a possible mechanism for producing some of the observed transitional discs (e.g., Cieza et al. 2008).

Although different photoevaporation models result in qualitatively similar behaviour, the different mass-loss rates and profiles do give rise to significant quantitative differences in how discs evolve. In particular, the large wind rates predicted for X-ray- and FUV-dominated photoevaporation imply that mass-loss is significant even at early times, and
that the total mass “lost” to photoevaporation over the disc’s lifetime may be larger than that accreted on to the star. These models also result in distinctly different properties during the clearing phase (most notably higher disc masses and accretion rates than in the EUV case), and therefore make very different predictions for the properties of transition discs. However, the dominant uncertainties in this picture of disc evolution remain our lack of understanding of angular momentum transport, and also of protoplanetary disc initial conditions (in particular the initial angular momentum distribution).

Despite these uncertainties, this evolutionary scenario is now supported by mature theoretical models and a wealth of observational evidence. For much of the protoplanetary disc lifetime the evolution of the gas disc is essentially governed by the competition between accretion and photoevaporation. During this time solid material in the disc is agglomerating into progressively larger bodies, and planets are forming and migrating through the disc; this scenario is illustrated schematically in Fig. 3. The processes driving disc evolution and dispersal essentially compete with those of planet formation, depleting the disc and ultimately terminating the epoch of (giant) planet formation.

3. Implications for planetary systems

Having reviewed the physics of protoplanetary disc evolution, I now consider the implications of these results for the formation and evolution of planetary systems. Disc evolution and dispersal influences planet formation in a number of different ways, by altering the disc’s chemistry, changing the dust-to-gas ratio, and ultimately by starving forming planets of the gas reservoir from which they accrete (e.g., Shu et al. 1993; Throop & Bally 2005; Guillot & Hueso 2006). Disc evolution also influences the architectures of young planetary systems: the migration of low-mass planets is very sensitive to the local disc structure, and disc clearing halts gas-driven planet migration. These issues are discussed in detail in Alexander et al. (2014); for reasons of length, I focus here on the effects of disc clearing on giant planet migration.

3.1. Halting planet migration

It was recognised more than 30 years ago that planets in discs migrate (Goldreich & Tremaine 1979, 1980), and when large numbers of exoplanets were discovered in relatively short-period orbits \((P \ll 1\,\text{yr})\), migration offered a natural explanation. However, the predicted migration time-scales are invariably much shorter than observed disc lifetimes (see, e.g., Kley & Nelson 2012), so planet migration must be slowed or stopped in order to produce planets at their observed locations. Type I migration, which applies to low-mass planets \((\lesssim 0.1M_{\text{Jup}})\), can be halted or even reversed by changes in the local disc structure (Paardekooper & Papaloizou 2009), but Type II migration, which applies to giant planets \((\gtrsim 0.5M_{\text{Jup}})\) is driven by the viscous accretion flow and continues as long as the disc accretes. Disc dispersal offers an obvious means of stopping Type II migration, but the manner and time-scale of disc clearing can have a strong influence on the dynamics of migrating planets.

Armitage et al. (2002) showed that the distribution of then-known exoplanets was consistent with a simple model, where giant planets form outside the snow-line (at 5–10AU) and undergo Type II migration before being stranded when the disc is dispersed. Subsequent population synthesis calculations have considered the processes of planet formation in much more detail (e.g., Ida & Lin 2004; Mordasini et al. 2009), but have generally adopted highly simplified models of disc dispersal (such as exponential depletion at all radii). Alexander & Armitage (2009) considered the migration of giant planets in a viscous/photoevaporation disc model, and ran large numbers of models to gener-
Figure 4. Exoplanet semi-major axes: the shaded histograms show current exoplanet data (taken from exoplanets.org; Wright et al. 2011); the dashed histograms show model predictions from Alexander & Pascucci (2012). The “pile-up” of ∼Jupiter-mass planets is plausibly explained by the interaction between migrating planets and gap-opening due to photoevaporation. [Figure adapted from Alexander & Pascucci (2012).]

ate statistical samples. They tested their model against observations of both discs and exoplanets, and showed that this simple scenario successfully reproduces the observed properties of protoplanetary discs and the distribution of giant planets at ∼AU radii. However, as disc dispersal by photoevaporation is not scale-free, it has the potential to leave a signature on the resulting planet distribution. Alexander & Pascucci (2012) revisited this issue, and found that photoevaporative gap-opening alters planet migration in a manner that is rather sensitive to both the planet mass and the efficiency of accretion across the planet’s orbit. The migration rate of planets close to $R_c$ changes when the gap opens, and this results in mass-dependent “deserts” and “pile-ups” in the distribution of planet semi-major axes (see Fig. 4). Alexander & Pascucci (2012) tentatively associated the observed pile-up of ∼Jupiter-mass planets at ∼1AU with this effect, and suggested that we may be able to link features in the observed exoplanet distribution to the physics of protoplanetary disc evolution. Other mechanisms, such as planet traps (e.g., Hasegawa & Pudritz 2012), can also lead to similar planet pile-ups, and it remains to be seen whether such features in the initial planet distribution are preserved over Gyr time-scales. However, these results suggest that it may be possible to use the observed properties of exoplanets to inform our understanding of both disc evolution and planetary accretion.

4. Summary

In summary, we now understand that planets form in evolving protoplanetary discs, and that disc evolution plays a major role in both the formation of planets and the subsequent dynamical evolution of planetary systems. The main processes driving disc evolution on
Myr time-scales are accretion, driven by MHD turbulence, and photoevaporation by high-energy radiation from the central star (though magnetically-driven winds may also play a role). Both accretion and photoevaporation are now observed directly, and there is good agreement between these observations and theoretical predictions. However, while our picture of protoplanetary disc evolution and dispersal is now relatively mature, its implications for planet formation have yet to be explored in detail. Disc evolution alters the conditions under which planets form, while disc clearing halts planet migration and may leave footprints in the observed distribution of exoplanets. Future observations and models of both discs and planets will explore these issues further, and allow us to build up a comprehensive picture of the early evolution of planetary systems.

Acknowledgements I am grateful to many colleagues and collaborators, especially Ilaria Pascucci, Sean Andrews, Phil Armitage & Lucas Cieza, for a number of useful and interesting discussions. I also thank Ilaria Pascucci and James Owen for providing some of the data used in Fig. 1, and Phil Armitage for the turbulent disc image used in Fig. 3. My research is supported by the Science & Technology Facilities Council (STFC), through an Advanced Fellowship (ST/G00711X/1) and a Consolidated Grant (ST/K001000/1).

References
Alexander, R. 2008a, New Astron. Rev., 52, 60
Armitage, P. J. 2011, ARAA, 49, 195
Feigelson, E., Townsley, L., Güdel, M., & Stassun, K. 2007, Protostars and Planets V, 313

Fromang, S., Latter, H., Lesur, G., & Ogilvie, G. I. 2013, \textit{A\&A}, 552, A71


Li, K. 2003, \textit{PASA}, 20, 337

Lodato, G. 2007, Nuovo Cimento Rivista Serie, 30, 293


Wright, J. T., Fakhouri, O., Marcy, G. W., et al. 2011, \textit{PASP}, 123, 412