The dispersal of planet-forming discs: theory confronts observations

Barbara Ercolano\textsuperscript{1,2} and Ilaria Pascucci\textsuperscript{3,4}

\textsuperscript{1}Universitäts-Sternwarte München, Scheinerstraße 1, 81679 München, Germany
\textsuperscript{2}Excellence Cluster Origin and Structure of the Universe, Boltzmannstrasse 2, 85748 Garching bei München, Germany
\textsuperscript{3}Lunar and Planetary Laboratory, The University of Arizona, Tucson, AZ 85721, USA
\textsuperscript{4}Earths in Other Solar Systems Team, NASA Nexus for Exoplanet System Science

Discs of gas and dust around million-year-old stars are a by-product of the star formation process and provide the raw material to form planets. Hence, their evolution and dispersal directly impact what type of planets can form and affect the final architecture of planetary systems. Here, we review empirical constraints on disc evolution and dispersal with special emphasis on transition discs, a subset of discs that appear to be caught in the act of clearing out planet-forming material. Along with observations, we summarize theoretical models that build our physical understanding of how discs evolve and disperse and discuss their significance in the context of the formation and evolution of planetary systems. By confronting theoretical predictions with observations, we also identify the most promising areas for future progress.

1. Introduction

Planets form from the dust and gas contained in discs around young stars. These discs are a natural consequence of the star-formation process, implying that all young stars have the potential to host a planetary system. Circumstellar discs are observed to evolve and finally disperse over a timescale of a few million years (§2), which is comparable with the timescales for planet formation by the core accretion process [1] and to migration timescales for giant planets [2]. This implies that the processes driving the evolution and dispersal of discs play a crucial role in shaping new planetary systems and likely contribute to the observed exo-planet diversity [3–5].

In the standard picture, the evolution of the surface density of young circumstellar discs is controlled by viscous accretion...
while the late evolution and final dispersal by star-driven photoevaporation (see §3 where we also discuss possible recent revisions). A class of objects that are thought to be in an intermediate state between possessing an optically thick (dust) disc and being disc-less, known as ‘transition discs’, may provide us with important insights on the disc dispersal mechanisms, and has been the subject of recent observational [6] and theoretical [7] reviews. It is, however, becoming clear that transition discs are in reality a diverse class of objects. Some of them may not actually be short-lived objects caught in the act of dispersing their discs, but rather produced by a different rarer and longer lived phenomenon [7].

This paper aims at making a stronger connection between theory and observations, by exploring the predicting power and limitations of current disc evolution and dispersal models and by discussing how observations inform us on how discs disperse. In §2, we summarize empirical constraints on the evolution and dispersal of circumstellar gas and dust, while §3 is an overview of theories of disc dispersal. Section 4 confronts theoretical predictions with observations, while §5 explores the implications of disc dispersal processes for planet formation and planetary architectures. We conclude with §6 by summarizing current theoretical successes and shortcomings and by discussing which investigations are most promising to clarify how circumstellar discs disperse.

2. Observations of dispersing discs

Young stars are historically classified either from their broad-band continuum emission at infrared wavelengths (SED), probing circumstellar dust, or from optical spectroscopy, probing surrounding gas. The first classification groups young stars into four classes that trace from the formation of the star and a massive disc (Class 0 and I) through the so-called protoplanetary disc phase (Class II) into the dispersal of the disc (Class III) [8]. Recent observations hint at the presence of giant planets in Class II discs [9] and structures in Class I sources that indicate substantial disc evolution linked to planet formation [10]. As such, we will not use the historical term protoplanetary discs for Class II sources but rather call them planet-forming discs. The second classification groups young stars into classical (C) or weak-line (W) T Tauri stars (TTs) mostly based on the strength of the Hα emission line [11]. CTTs are actively accreting disc gas. It was noted early on that most WTTs are associated with Class III spectral energy distributions [12]. Thus, understanding how circumstellar gas and dust disperse requires observations that can trace the evolution from Class II to Class III and from CTTs to WTTs. As X-ray emission is elevated for all TTs, X-ray surveys of star-forming regions are also an important complement to optical and infrared observations to recover the true pre-main sequence stellar population [13].

2.1. Disc lifetime versus dispersal time

As circumstellar discs form, gas should account for almost all of the disc mass while dust only for about 1%. However, most of our understanding of how discs evolve and disperse is based on observations tracing the dust component. In addition, planet-forming discs are located at hundreds of parsecs and are just a few hundred AU in size, meaning that most observations do not spatially resolve them. One of the goals of this section is to summarize what we know about disc lifetimes, i.e. how long it takes to completely disperse a Class II disc, and about dispersal time, i.e. how long it takes for a disc to transition from Class II to Class III. We should note upfront that these times are linked to the ages of young star-forming regions and associations, which are not known in absolute sense approximately below 20 Myr [14].

2.1.1. Dust diagnostics

The advent of infrared observatories such as the Spitzer Space Telescope has enabled sensitive surveys of nearby star-forming regions and associations that revealed the population of young stars with dust discs, i.e. those displaying excess emission above the stellar photosphere at infrared wavelengths. The most recent and comprehensive analysis of these infrared data has been carried out in a recent review [15] and we refer the reader to their Section 6 for a comparison to previous work. In short, the fraction of Class II discs shows a clear decay with time and reaches zero approximately at 10–20 Myr. When assuming that the dust disc fraction decreases exponentially with time, the e-folding time approximately is 2–3 Myr at 3–12 μm and 4–6 Myr at 22–24 μm (see also figure 1a). The e-folding time of approximately 2–3 Myr is often reported as the disc lifetime. While 3–12 μm wavelengths mostly trace disc material at approximately 1 AU around sun-like stars, 22–24 μm trace dust further away at tens of AU. The longer e-folding time at longer wavelengths could result either from inside-out clearing of primordial dust or
from second-generation dust produced by collisions of kilometre-size planetesimals that formed in the first few million years of disc evolution [17]. It is worth noting that the first scenario implies a long disc clearing time of a few million years. The dust disc lifetime as a function of stellar mass is less well characterized but appears to be shorter for higher- ($\geq 2M_\odot$) than lower-mass stars [18,19]. Expanding upon these findings, it was shown that there is an overabundance of discs around lower-mass stars ($\leq 0.6M_\odot$) in the 5–12 Myr old Collinder 69 cluster [20], although no difference is found in the spatial distribution and disc fraction in younger clusters [21,22].

Disc fraction estimates are not extended to far-infrared and longer wavelengths as stellar photospheres become undetectable at typical distances of hundreds of parsecs. Nevertheless, even early millimetre surveys could show that the distribution of dust disc masses (hereafter, $M_{\text{dust}}$) around solar-mass stars in the young (approx. 2 Myr) Taurus star-forming region is significantly different from that around 10–30 Myr-old stars [23]. The sensitivity of ALMA has recently enabled a survey of all Class II discs in the nearby Lupus [24], Chamaeleon I [25] and Upper Sco [26] regions, thus complementing the previous SMA Taurus millimetre survey [27]. When adopting the same assumptions to analyse these datasets, it is found that the mean $M_{\text{dust}}$ in the 1–3 Myr-old regions of Taurus, Lupus and Chameleon I is about three times higher than in the 5–10 Myr-old Upper Sco association and that the three young regions share the same $M_{\text{dust}}$–$M_*$ scaling relation\(^1\) while Upper Sco has a steeper relation [25]. This latter finding may result from efficient inward drift of millimetre grains around the lower-mass stars and does not contradict the longer disc lifetimes inferred from infrared observations. Interestingly, the dust mass of Class III objects in Upper Sco, most likely debris or evolved transition discs [26], is at least 40 times lower than the average disc mass of 1–3 Myr-old Class II objects [25]. Similarly, discs around higher mass Herbig Ae/Be stars with ages $\geq 10$ Myr, most likely debris discs, are at most $0.5M_\odot$ while younger stars can have discs more than an order of magnitude more massive [28]. Overall, observations demonstrate that dust discs evolve significantly approximately from 1 to 10 Myr. The fraction of Class II discs drops approximately from 70% to 10% and the dust mass of the surviving 10 Myr-old Class II discs is reduced by a factor of approximately 3 on average. These observables are consistent with the fast evolution and dispersal of dust in the protosolar nebula as inferred from studies of meteorites and from constraints on the formation time of asteroids, planets and moons in the Solar System (e.g. fig. 9.5 in [16]).

\(^1\)\(M_{\text{dust}} \propto (M_*)^{1.3-1.9}\) for Chamaeleon I. The range in the power law index reflects two extremes of the possible relation between the average disc temperature and stellar luminosity.
The dust clearing time, the time for a source to transition from Class II to III, is measured via multi-wavelength infrared observations. The classical definition of transition discs goes back to early work [29] which included in this group young stars with significantly reduced or no near-infrared excess but large mid- to far-infrared excess emission (figure 1b for comparison with SEDs). This SED type hints to the development of an inner dust cavity that may represent the initial step of an inside-out disc clearing process. While this approach has clear limitations (e.g. in the hole size that can be detected and its sensitivity to small dust grains [6,30]), SEDs remain the most used tool to identify discs in different evolutionary stages. Based on SEDs, it was realized early on that the fraction of these transition discs is small, of order 10% of the Class II discs [31] implying a clearing time of a few $10^5$ years, much shorter than the disc lifetime. It has recently been reported that the fraction of transition discs increases from 8% at $\lesssim 3$ Myr to about 45% at approximately 10 Myr [32]. As the transition disc fraction is defined as the number of transition discs over the total number of discs and the overall disc fraction decreases approximately by a factor of 5 in regions older than approximately 10 Myr, the above results are expected for a time-independent dust clearing time. The existence of another type of transition discs, so-called homologically depleted discs has been proposed [33]. This definition has been applied to discs that have overall weaker infrared excess emission than Taurus discs but stronger emission than debris discs (see also [34]). This group would include discs whose surface density decreases homogeneously at all radii with time thus representing a different path of clearing. The disc clearing time, when combining transition and homologically depleted discs, is approximately 1 Myr [35], a significant fraction of the disc lifetime. However, it has been successively shown by means of detailed radiative transfer modelling that most of the homologically depleted discs are consistent with optically thick discs where dust has settled to the disc midplane rather than discs transitioning from Class II to III through a radial homogeneous draining [36]. A re-evaluation of the SEDs of over 1500 sources in 15 nearby star-forming regions, now considering settled dust discs, finds that less than 2% of the discs lie in the homogeneous draining regime [22]. This result, combined with the aforementioned observations, suggests that most discs disperse via inside-out clearing and the disc clearing time is indeed short, a few $10^5$ years.

Recently, a comprehensive review of the observational properties of transition discs has been presented [6]. What transpires is that the SED-classified transition discs are a heterogeneous group of objects. While several of the SED-inferred dust cavities have been now confirmed via millimetre imagery [37–39], some cavities are also seen in polarized near-infrared light (e.g. RX J1604-2130 [40]) while others are not (e.g. SR 21 [41]). About half of the transition discs have high millimetre fluxes indicative of large disc masses ($M_{\text{disc}} > 5 M_\odot$) while the other half are weak millimetre sources [42]. Their mass accretion rates (discussed in the next section) can be as high as that of full discs [43] or much lower [44]. This diversity points to different physical mechanisms responsible for a transition-like SED. Since some of these mechanisms may not lead to disc dispersal, the disc clearing time of a few $10^5$ years is an upper limit and we will discuss in §4 ways to identify those discs that are truly in transition. We caution that the large number of dust-depleted cavities at millimetre wavelengths (1 in 3 for the mm-bright discs [38]) should not be used to estimate the disc clearing time since such cavities can be induced by gas pressure bumps at snowlines or by forming planets [45], physical mechanisms that are not truly dispersing disc material (see §3).

2.1.2. Gas diagnostics

Similar to the infrared surveys, the gas lifetime can be obtained by measuring the fraction of stars accreting disc gas as a function of cluster age. Several accretion indicators have been discussed in the literature with the optical/UV excess emission from shocked gas at the stellar surface being the most reliable one but also the most challenging to observe (see [46] for a comparison of accretion indicators). By contrast, the H$\alpha$ line at 656.28 nm is bright, easy to observe with current facilities, and its luminosity correlates well with the optical/UV excess emission [47], hence its equivalent width (EW) and/or width are often used to identify the population of accreting stars. This diagnostic has been used to show that the fraction of accreting stars declines with cluster age approaching zero at approximately 10 Myr and, assuming an exponential decay, the e-folding time is 2.3 Myr, the same as the dust e-folding time measured at 3–12 $\mu$m [48] (see previous subsection and figure 1a). The rate at which disc gas accretes onto the star (hereafter, $M_{\text{acc}}$) is also found to decline with time [16,49], as expected by viscous evolution, but the spread of $M_{\text{acc}}$ at any age is large and the ages of young stars are poorly known to critically test theoretical predictions. Overcoming the age issue, a correlation between $M_{\text{acc}}$ and $M_{\text{dust}}$ has been reported in Lupus [50], with a ratio that is roughly consistent with the expected viscous timescale when assuming an interstellar medium gas-to-dust ratio. However, the observed correlation appears
to be mainly driven by the underlaying correlation with $M_*$ [51]. Indeed, $\dot{M}_{\text{acc}}$ is found to scale rather steeply with stellar mass with a relation close to a power law of two in several star-forming regions: $\dot{M}_{\text{acc}} \propto (M_*)^2$ [46,47,52,53]. The steepness of the relation has challenged viscous disc models from early on [54] and has prompted a number of theoretical attempts to explain it. These include Bondi–Hoyle accretion [55] (see also [56]) and dependance on the initial conditions of the parent cloud from which the protoplanetary disc formed [57,58]. Recent work has shown that the relation may also be a consequence of viscous photoevaporative disc clearing driven by X-ray radiation from the central star [59]. In this case, it was shown that the relation between accretion and stellar mass simply reflects the relation between the X-ray luminosity and the stellar mass, as measured by several surveys including the Chandra Orion Ultradeep Project [60].

Accretion indicators probe the reservoir of disc gas very close to the star; hence it is natural to ask whether gas at tens or hundreds of AU from the star has a similar lifetime. Unfortunately, surveys at infrared and millimetre wavelengths probing those radii are still limited in sample size, but their constraints of a gas lifetime of approximately 10–20 Myr are consistent with the value obtained from accretion indicators [61–63] (see also figure 1a). In addition, modelling of multiple gas diagnostics from a sample of 15 solar-type stars that lack signatures of accretion but possess optically thin dust emission shows that the gas surface density at 1 AU has dropped to less than 0.01% of the minimum mass solar nebula [62] in most systems, implying too little gas to even circularize the orbits of terrestrial planets.

The gas clearing time is yet unconstrained observationally as resolved images of gas cavities are just beginning to be possible. However, there may be an indication for gas depletion in spectrally resolved gas lines. After taking into account stellar mass and disc inclination effects, the SED-identified transition discs have narrower UV $H_2$ florescent emission lines than full discs [64] and a much weaker or absent broad component in the optical [OI] forbidden lines [65] and in the M-band CO rovibrational lines [66,67]. As transitional objects have a dearth of dust grains in their inner disc and might be, on average, lower accretors (e.g. [68], but see also [69] for a possible stellar-mass bias), it should be demonstrated that a reduced gas scale height due to less heating from dust grains and accretion does not result in undetectable broad, i.e. high velocity, emission. A total of six transition discs have been recently imaged at millimetre wavelengths with ALMA in CO and in the continuum at a spatial resolution sufficient to test if gas is depleted within the known dust cavities. In all cases gas cavities, i.e. where the gas surface density decreases by a factor larger than 10, are found to be smaller than the dust cavity [70]. This finding is taken as evidence for embedded planets because planets can clear their orbits in the gas and trap millimetre grains at the outer edge [71]. However, other physical processes can also cause a pressure gradient in the gas that will trap dust grains at larger radii compared with the gas cavity. It is also worth noting that the main CO isotopologue may not be a good gas density tracer in the inner disc as it readily becomes optically thick, e.g. the case of LkCa15 for which the $^{12}$CO(6-5) transition indicates gas depletion inside the dust disc, whereas the HCO$^+$ (4-3) line does not [72]. Finally, all discs imaged so far have large dust cavities (more than 25 AU and up to 140 AU) and are not representative of the entire population of transition discs [42] (see also §4). ALMA observations of the gas content of more typical transition discs would be extremely valuable.

### 2.2. Disc winds

Direct evidence of gravitationally unbound/outflowing material from the disc+star system is achieved by measuring velocity shifts in the emission lines of ionic, atomic and molecular lines.

Class I protostars are known to power parsec-long collimated jets that are bright in optical forbidden lines [73]. Similar gas tracers applied to Class II CTTs reveal a high-velocity component, typically blueshifted by 30–150 km s$^{-1}$, which forms in micro-jets, a scaled-down version of the Class I jets [74]. However, typical ratios of jet mass flux to accretion rate are low, less than or equal to 0.1 [75,76], hence jets/outflows do not seem to play a major role in dispersing Class II discs. More promising are slow (less than 30 km s$^{-1}$) disc winds, also traced by optical forbidden lines, most frequently by the [OI] line at 6300 Å [77–79] (see also figure 2a). It has been recently found that such low velocity emission is present in all TTs with dust discs, even those classified as WTTs, and can be described by a combination of a broad and a narrow component [65]. Most of the broad component emission arises approximately within 0.5 AU and shows the expected disc wind signature, i.e. larger blueshifted centroid velocities associated with narrower lines and lower disc inclinations (see fig. 27 in [65] and predicted profiles in §3).

$^2$Optical [OI] broad components have FWHM $>40$ km s$^{-1}$ while M-band CO broad components have a FWHM typically larger than approximately 50 km s$^{-1}$. 
Figure 2. Examples of disc wind diagnostics. (a) The [OI] 6300 Å profile of CW Tau is from [65]; note that the high-velocity component (HVC) is associated with fast jets. The [OI] 5577 Å transition is weaker than the 6300 Å but shows a very similar low-velocity profile [65]. (c) The CO-M band profile of AS205 is from stacked CO rovibrational lines around 4.7 μm [67]. (b) The [NeII] 12.8 μm profile is the mean TW Hya profile [80] and (d) the CO J = 2–1 profile of HD163296 is from Klaassen et al. [81]. All profiles, except that of AS205, are in the stellocentric reference frame and emission shifted in velocity with respect to the stellar velocity is indicative of unbound gas. M-band CO profiles do not have an absolute velocity calibration; indication of a slow wind in AS205 comes from spectro-astrometry [82].

Such winds must be magnetically driven given that the emitting region is well inside the gravitational radius even for gas at 10 000 K (see also §3). The narrow component is overall less common but always present in transition discs. It traces gas further away (approx. 0.5–5 AU) and may be associated with photoevaporative winds as indicated by the blueshift in half of the sample, although the expected trend between blueshifts and disc inclinations is not seen.

Stronger evidence for thermally driven winds comes from mid-infrared high-resolution spectra of [NeII] at 12.8 μm. The low-velocity emission has blueshifts between 2 and 12 km s\(^{-1}\) and FWHM approximately up to 50 km s\(^{-1}\) that are consistent with photoevaporative winds (see Sect. 3.2 in [83] and references therein). Among these objects it is worth pointing out TW Hya, a nearby low-mass star with a well-known transition disc [84]. Direct evidence of on-going photoevaporation is provided by multi-epoch blueshifted (approx. 5 km s\(^{-1}\)) and asymmetric [NeII] 12.8 μm emission, most of which arises beyond the dust cavity [80] (see also figure 2b). All other observables are consistent with the gap being opened by star-driven photoevaporation [85].

Molecular tracers of outflow activity are common in the protostellar phase. They also probe a low velocity (less than 10 km s\(^{-1}\)) component whose intensity decreases while the opening angle increases going from the Class 0 through to the Class II phase [86]. This slow wide-angle component, around a much faster jet, is naturally produced in magnetohydrodynamical (MHD) disc winds launched out to several AU [87,88]. Corroborating this picture, ALMA recently mapped a slow CO disc wind around the fast jet approximately from the 4 Myr-old Herbig Ae star HD 163296 [81] (see also figure 2d). Interestingly, the estimated mass loss rate from the molecular wind is similar to the mass accretion rate onto HD 163296 suggesting that winds contribute significantly to disc dispersal. Like optical forbidden lines, M-band CO rovibrational lines also show a broad and a narrow component [67] with FWHMs slightly larger than the [OI] 6300 Å lines suggesting that CO traces gas at smaller disc radii around TTs [65]. While some
profiles are double-peaked as expected if emission arises in Keplerian discs, others have a single peak with a broad-based emission [89] reminiscent of the optical forbidden lines. It is yet unclear if the single-peaked M-band CO profiles trace a disc wind as spectra published so far lack an absolute radial velocity calibration. The spectroastrometric signal detected for two Class II sources (RU Lup and AS 205N) is intriguing and can be explained by a wide-angle disc wind [82] (see also figure 2 for the CO M-band profile from AS205N). In the case of AS 205N, follow-up high-resolution ALMA observations also show deviations from Keplerian rotation in millimetre CO rotational lines which could be due to a low velocity disc wind or tidal stripping by its companion AS 205S [90].

3. Theories of disc dispersal

Of all mechanisms that have been brought forward to explain transition-like features in discs, only two can be considered true global disc dispersal mechanisms: accretion and disc winds. These processes together shape the evolution of the surface density of the disc and remove material from it by respectively depositing it onto the central object or launching it away from the system. There is strong observational evidence that both processes occur in discs, which has been discussed in the previous sections. In fact accretion and mass loss via a wind compete in determining the surface density of the disc at a given radius. To zeroth order a gap will be formed in a disc if the wind is faster at removing material from a given region than accretion is at pushing new material into it.

It is classically assumed that for the largest part of their lives, the evolution of the surface density of discs can be well described by simple viscous theory [92,93]. These models predict a slow, homogeneous dispersal of the disc. Observations, however, show that the dispersal is not a continuous process: after having evolved slowly for a few million years, discs regularly seem to disappear quite abruptly, implying a disc dispersal timescale about 10 times faster than the global disc lifetime (§2.1). These observations have motivated the development of theoretical models able to match this two-timescale and inside-out dispersal modus. Photoevaporation from the central star is currently accepted as an important player in the late evolution of discs and has seen several dedicated theoretical efforts [94–101]. All models of photoevaporation show that radiation from the central star heats the disc atmosphere, where a thermal wind is established. The mass loss rate of the wind must exceed the accretion rate in the disc for dispersal to set in. Young discs accrete at a vigorous rate, which naturally decreases as time goes by, until, after a few million years accretion rates fall to values smaller than the wind rates, allowing photoevaporation to take over the further evolution of the disc. Once the dispersal sets in the disc is then quickly eroded from the inside out (§3.2). This scenario is based on the assumption that the mass accretion rate is radially constant throughout.

While the community seems to agree on this broad brush picture, quantitatively speaking, the dispersal mechanism is still largely unconstrained. In fact, photoevaporation by energetic radiation from the central star is not the only way to produce a disc wind. Extended magnetically launched disc winds appear to be necessary to explain the properties of jets in young stars [102] and models have been developed since the 1980s [103]. However, only recently non-ideal MHD effects have been incorporated, albeit with computational costs that have mostly limited the calculation to small regions of the disc. Some of these recent models, based on local simulations, suggest that even a weak vertical magnetic field can launch a wind [104]. The wind can be so vigorous as to compete with photoevaporation for the dispersal of the disc and may even provide an efficient channel to remove angular momentum from the material in the disc, hence driving the accretion process. Global MHD discs simulations have until very recently only been possible in the ideal limit [105,106], or including only Ohmic diffusion [107,108]. Generally, the Ohmic term is sub-dominant to the Hall term in the inner disc (inside 10 AU), and sub-dominant to ambipolar diffusion further out.

The most recent simulations in the ideal limit [109], which improve on resolution and convergence compared to previous work, do not show winds that significantly contribute to the evolution of the surface density of the disc under the assumption made. At the time of writing only one set of global simulations of protoplanetary discs including all three non-ideal MHD effects (Ohmic and ambipolar diffusions, and the Hall drift) has been performed [110]. Discs are found to accrete (in which case a disc wind is also present) only for given configurations of the large-scale magnetic field, which thus remains an important uncertainty.

3Stellar winds appear to be less efficient than accretion and disc winds [91].
The nature of disc accretion and the driving mechanisms behind disc winds are currently a rapidly developing and very active area of research. An attempt at summarizing the state of the art at the time of writing follows.

3.1. Accretion

Accretion in discs is often phenomenologically described by the so-called $\alpha$ prescription [111], where the parameter $\alpha$ is used to describe the viscosity of the material, and it represents the ratio of stress over pressure. The source of the viscosity is, however, still a matter of debate. The standard paradigm for circumstellar disc accretion invokes the magneto-rotational instability (MRI) to drive turbulence [112]. The conditions to trigger MRI are that the gas is sufficiently ionized, that the disc is weakly, but non-negligibly magnetized and that the angular frequency decreases with radius. While the last two conditions are always satisfied in discs, there are uncertainties about the ionization structure of the gas. It is indeed often pointed out in the literature that large disc regions may be actually MRI-inactive, due to poor coupling of the gas to the magnetic fields. In general, regions lying under large column densities of material in the inner disc may be virtually neutral, the so-called ‘dead zones’ [113], where MRI turbulence is practically absent. Accretion in these cases is thought to happen in the outer layers surrounding the dead zones. These regions, however, may be dominated by non-ideal MHD effects (e.g. ambipolar diffusion, Ohmic diffusion) which may make MRI transport very inefficient [114].

Recent non-ideal MHD simulations [104,115] show that in the presence of even a weak net vertical magnetic field threading the disc, MRI is completely suppressed. As a result however a vigorous magnetocentrifugal wind is launched in these simulations. The wind in this case also removes angular momentum from the disc and can thus drive accretion. Using the results of local simulations a 1+1D global model of a disc evolving under the influence of MHD winds can be constructed [104,116]. Based on this approach, recent work has found that accretion is completely dominated by winds in most regions of the disc, with the MRI perhaps only playing a part in the outer regions, e.g. at $R \gtrsim 50$ AU for the fiducial model [116]. In this scenario, the accretion and wind mass loss processes are strongly coupled by the field strength which drives both. While the suggestion of MHD winds driving accretion in discs is certainly tantalizing, it is currently too early to assess its validity. In particular, all attempts at constructing global models for the evolution of discs rely on a number of important assumptions about how critical parameters scale with radius [116] and are sensitive to the amount of magnetic flux assumed to thread the disc.

Regardless of the detailed process(es) that provide the viscosity needed to drive accretion in discs, if the phenomenologically formulated alpha-description is roughly correct, then one expects the surface density to decrease as a power law of time and the radius to increase, as a result of viscous draining and spreading, i.e. $\Sigma_{\text{disc}}(t) \sim M_{\text{disc}}(t)^{-1.5}$ [92]. The implication is that viscously evolving discs should become progressively faint simultaneously at all wavelengths as a function of time. Owing to the temporal power law, the evolution should become progressively slower, such that discs should spend most of their lifetime in a homogeneously draining transition phase. To date there is no convincing observational evidence of any homogeneously draining disc. Current data support instead a fast dispersal phase which proceeds from the inside out (§2.1). Viscosity is thus not the end of the story, rather an additional dispersal process must take over at advanced stages of disc evolution. Mass loss via a disc wind, driven by photoevaporation by the central star, likely plays an important role, such that the evolution of the surface density of a disc switches from an accretion-dominated to a wind-dominated regime, marking the beginning of the final, rapid disc dispersal phase. However, if MHD winds dominate angular momentum transport in a significant part of the disc, the evolution of the surface density and of the mass accretion rate is unlikely to be well described by an alpha-formalism [116,117]. While current observations are reported to be broadly consistent with the alpha-disc scenario, they cannot unequivocally prove it, as discussed in more detail in §2.1.

3.2. Photoevaporative/thermal winds

Theoretical work on photoevaporation has been recently reviewed by a number of authors [83,118,119]. We will only summarize the main mechanisms here, focusing on the current open questions and uncertainties, as well as on the specific predictions and the predictive power of current models.

Radiation from the central star penetrates the disc atmosphere and deposits energy in (i.e. heats) the gas. To zeroth order a thermal wind is established if the temperature of the gas at a given location
becomes higher than the local escape temperature. Far-ultraviolet (FUV, $6 \text{ eV} \lesssim E \lesssim 13.6 \text{ eV}$), extreme-ultraviolet (EUV, $13.6 \text{ eV} \lesssim E \lesssim 100 \text{ eV}$) and soft X-rays ($0.1 \text{ keV} \lesssim E \lesssim 2 \text{ keV}$) are all able to efficiently heat the gas in the disc atmosphere. As shown in figure 3, the penetration depth of the photons depends, however, strongly on wavelength, with the EUV being absorbed out completely within a column of neutral hydrogen of approximately $10^{20} \text{ cm}^{-2}$ [121]. FUV and soft X-rays have larger penetration depths, of order $10^{21}, 10^{22} \text{ cm}^{-2}$, depending on the abundance of small dust grains and polycyclic aromatic hydrocarbons in the case of the FUV and the spectral shape in the case of X-rays [120]. High energy X-rays ($E \gtrsim 2 \text{ keV}$) can penetrate very large columns, but they do not heat the dense gas efficiently enough to drive a wind. Again to zeroth order, the surface mass loss rate at a given disc radius is given by $\rho_b \times c_s$, where $c_s$ and $\rho_b$ are the sound speed and the density of the gas at the base of the flow, respectively. Given that the temperature of the gas only enters as a square root in the calculation of the sound speed, and that $\rho$ varies steeply as a function of vertical depth $z$ in the disc, the surface mass loss at a given radius is much more sensitive to the location of the base of the flow than to the temperature of the gas [122]. The FUV and soft X-rays are thus expected to drive the strongest winds [97–101,120]. There is still a debate in the literature as to what type of radiation may be the main driver of the wind, although EUV photons alone can be excluded in a few sources [123]; see also §3.4. This is a fundamental question as the mass-loss rates implied by the different models can differ by orders of magnitude. Mass-loss rates crucially determine the timescales of disc dispersal for given initial disc conditions.

Hydrodynamical simulations show that a wind driven by EUV radiation only would result in mass-loss rates of order $10^{-10} M_\odot \text{ yr}^{-1}$ for a solar-mass star irradiated by photon luminosity of $\phi \sim 10^{41} \text{ photons s}^{-1}$ [96]. Radiation-hydrodynamic models including both X-ray and EUV radiation (XEUV) yield instead mass-loss rates of order $10^{-8} M_\odot \text{ yr}^{-1}$ for a disc around a solar-type star irradiated by a star with X-ray luminosity of $L_X \sim 10^{30} \text{ erg s}^{-1}$ [98–100]. The scaling of the mass loss rate in the two cases is also very different: $\dot{M} \propto \sqrt{\phi}$ for pure EUV and $\dot{M} \propto L_X$ for the XEUV case. Thus, the mass-loss rates predicted by the XEUV model for solar-type stars may vary strongly depending on the X-ray property of the central object. X-ray surveys of young stars show approximately 2 orders of magnitude scatter in the X-ray luminosities of young solar-mass stars. The contribution of short term variability to the observed scatter is minor [124], with intrinsic differences in stellar rotational velocities or internal structure being instead the dominant factor. The same scatter will then be reflected in the mass-loss rates, and thus in the expected lifetimes of their discs. The same is not true for an EUV-only scenario where the dependence of the mass-loss rate on the EUV luminosity is much weaker. thermochemical models including EUV, FUV and X-ray heating also find mass-loss rates of order $10^{-8} M_\odot \text{ yr}^{-1}$ for solar-type stars [101]. These models, however, do not perform a hydrodynamical calculation to obtain a solution.
for the wind. The mass loss rate from the 1+1D hydrostatic equilibrium models at a given radius is estimated using the approximation discussed above \((\rho b c_s)\), where the location of the base of the flow is chosen so that \(\rho b c_s\) is maximized. It is difficult to estimate the uncertainty introduced by the method, thus a comparison with hydrodynamical models [96,98–100] is of limited relevance.

The wind profile, which determines the region of the disc that is most affected by photoevaporation, is also very different in each scenario [83,118]. The X-ray profile is more extended than the EUV profile, which predicts mass loss only from a very narrow range of disc radii, centred at the gravitational radius. The FUV model is again very different, showing mass loss from the outer regions of the disc and predicting in some cases an outside-in mode of dispersal. The detailed profile of the photoevaporative wind has important consequences for the formation and migration of planets in the wind. As an example, it has recently been shown that changing the wind profile yields completely different distributions for the semi-major axes of giant planet in otherwise equal populations of discs losing mass globally at the same rate [5].

While the physics of photoevaporation is reasonably well understood, all current models are somewhat incomplete. The EUV model focuses on hydrodynamics and assumes isothermal gas [95,96, 125], hence being limited to the EUV-ionized layer only, which yields roughly constant temperatures around a value of 10000 K. The available FUV models focus on chemistry, but do not perform hydrodynamical calculations [101,126]. Current radiation hydrodynamic calculations of X-ray driven winds use realistic gas temperatures obtained from X-ray photoionization calculations [120], but they do not include chemistry and ignore the dust phase [98–100]. Indeed none of the existing models take into account dust evolution in the underlying disc and entrainment of grains in the wind self-consistently.

Table 1 summarizes the main ingredients included by the models, as discussed above. There are further important differences in the implementation of the various heating channels, as well as on assumptions made by the different codes. A full technical discussion of the different choices adopted and their influence of the derived mass loss rates and profiles is still at this stage, since not all ingredients are included in all models. A benchmarking exercise would be nevertheless useful at this stage in order to converge on a roadmap for future development.

### 3.3. Magnetohydrodynamical winds

Before we begin our discussion on magnetically supported/driven winds, it is worth noting that there exists a large body of literature on MHD models developed to explain outflows from young stars, particularly those of Class I. One set of these models were based on disc/magnetosphere interactions and mostly deal with the ideal MHD region of the disc [102,127]. Including a comprehensive discussion of these early works is beyond the scope of this review. Here we will focus on MHD winds only in the context of disc dispersal.

As mentioned in §3.1, MRI [112] is currently considered a likely mechanism by which angular momentum in discs is redistributed to allow accretion of matter through the disc and onto the central object. Classically, MHD simulations in a local shearing box have been used to study the properties of turbulence driven by the MRI [128]. More recently, vertically stratified local shearing box simulations have been used to investigate the possibility of MRI-driven disc winds [129,130]. Their model differed from previous work, that used initially toroidal and zero-net vertical flux magnetic fields [131], by assuming vertical magnetic fields and outgoing boundary conditions. This was motivated by the assumption that discs may be threaded by net vertical magnetic fields that are connected to their parental molecular cloud. These authors find that vigorous winds can be driven by the MRI, which may
disperse the discs with timescales as short as 4000 years at 1 AU and $6 \times 10^5$ years at 30 AU. These rather short timescales, which are clearly in contrast with disc observations, do not account for global viscous accretion, which would slow down the process. Very recent work expands on these results to present one-dimensional models of disc evolution including the effects of viscous heating, in addition to the loss of mass and angular moment by the disc wind [132]. The focus of this work is more on the early stages of evolution, when accretion heating is important and can give rise to density structures which show a positive radial slope for the surface density in the inner disc regions. While this may have important implications for planetesimal formation and migration models, its relevance to the dispersal of discs is limited. The latest global ideal MHD simulation performed, at the time of writing this review, is for thin accretion discs threaded by net vertical magnetic fields [109]. This work suggests that only very weak and episodic disc winds can be driven and these are not efficient at carrying angular momentum.

As mentioned in the previous section, however, local non-ideal MHD simulations suggest instead that in the presence of a weak net vertical magnetic field, MRI is completely suppressed, while a strong magnetocentrifugal wind is launched, which carries away disc angular momentum so efficiently to account for the measured accretion rates [104]. One-dimensional, vertically integrated, disc evolution models including angular momentum redistribution and wind angular momentum loss have been constructed using a parametrization of the wind stress parameter from the vertically stratified ambipolar disc simulations [133,134]. The simulations modelled MRI turbulence at 30 AU in an FUV-ionized disc and led to the conclusion that, depending on the value of the initial net flux, discs may in some cases undergo a two-timescale dispersal behaviour, similar to that achieved by photoevaporation [134]. For low initial magnetic field strengths discs are almost inviscid and very long lived. However, even in the cases where the initial field is sufficient to instigate rapid disc dispersal, it is not possible to predict what kind of transition disc morphology would be obtained or whether the disc would then disperse from the inside-out, as observations suggest (see §2.1). The behaviour of MHD dispersal models depends upon the radial evolution of the net field and how this couples to disc evolution, which is currently unknown [134]. A simplified (1+1D) approach, motivated by numerical simulations [104] has been used to model MHD wind driven disc evolution [115,116]. This study shows that, under the assumptions made, wind-driven accretion and mass loss, rather than MRI, dominate disc evolution. Later work also tries to incorporate some thermodynamical effects using simplified prescriptions for disc temperature and the depth of the FUV penetration [116]. In this work it is found that, while FUV penetration can have significant effects on the wind-driven accretion rates and fractional wind mass loss rates, the key parameter controlling the disc evolution timescales in their models is the amount of magnetic flux threading the disc. This depends both on the initial strength of the magnetic field and its evolution with respect to the surface density evolution of the disc. This is in agreement with previous assessment [134], and it shows that understanding the behaviour of magnetic fields in discs is key to predicting the impact of MHD winds in the evolution and dispersal of discs. While direct observations of magnetic fields in planet-forming discs remain challenging [135,136], disc winds are commonly observed (§2.2) and their properties may be used to indirectly constrain magnetic fields.

3.4. Prediction of wind structures

Gas flowing into a wind has a very clear non-Keplerian kinematical signature, which can be observed in the profile of emission lines that are produced within it.

The physics of photoevaporative winds is relatively well understood and detailed calculations of wind structures, that will improve the ability of models to confront observations, are slowly maturing. Examples of predicted emission line profiles for four different orientations of a full disc (figure 4, left panels) and one with an inner cavity of approximately 14 AU (figure 4, right panels) are shown in figure 4, which is a re-rendering of previous radiation-hydrodynamics calculations [137]. The top panels show density maps, where superimposed is the location of the 85% emission region of the [OI] 6300Å line (yellow contour) and [NeII] 12.8 µm line (purple contour), and the arrows show the direction and magnitude of the velocity vector. The lower panels show the emission line profiles for the same lines at four different disc inclinations: 0° (blue dotted line), 30° (red dashed line), 60° (black dashed-dotted line) and 90° (green solid line line). Note how the wind emission dominates for close to face-on orientations (0°), leading to a large blueshift and a small FWHM, as seen in some disc wind diagnostics (§2.2).

On the other hand, MHD wind models that try to predict the mass and angular momentum loss profiles and flow topology from the disc rely on a number of assumptions about how critical parameters scale with radius [115]. Furthermore, as discussed in the previous section, these models are extremely sensitive to details of the magnetic field strength, topology and evolution that are poorly known. As
such, a detailed comparison with candidate wind diagnostics is still lacking. However, one important difference that can already be identified using current models of photoevaporative and MHD winds is that the latter are likely to include an unbound higher velocity component launched close to the star (\( \lesssim 1-2\) AU). Photoevaporative outflows, instead, drop off rapidly in strength inside a radius of approximately 0.2\( GM_*/c_s \sim 2\) AU, where \( c_s \) is the sound speed \([134]\). Therefore, while the identification of an unbound component launched close to the star immediately points to MHD winds \([65]\) (see also §2.2), slower winds driving the mass loss in the 2–10 AU regime may be MHD or thermal in nature.

As observations of disc winds have grown in the past few years (§2.2), they also revealed that multiple diagnostics are needed to determine wind mass loss rates. For instance, predicted [N\( \text{eII} \)] 12.8 \( \mu \)m line profiles for an EUV-driven wind \([138]\) and an X-ray-driven wind \([137]\) are very similar although the mass loss rates differ by 2 orders of magnitude. This is because the location of the base of the flow is much deeper in the disc in the X-ray case (see §3b). However, combining this diagnostic with an upper limit on the EUV luminosity reaching the disc from the centimetres emission in excess of the thermal dust emission \([139,140]\) rules out EUV photoevaporation in at least three cases pointing to mass loss rates larger than \(10^{-10} M_\odot \) yr\(^{-1}\) \([141]\). Similarly, in the context of thermal wind models, the [OI] 6300 Å line cannot be used alone to infer the underlying mass loss rate \([142]\). This is because, if it is collisionally
excited, it has a strong temperature dependence imposed by the Boltzmann term in the emissivity and mostly traces the hot layer of the wind, not the bulk. Molecular lines should be more sensitive to the mass loss rate since they sample a significant area of the wind launching regions. However, the exploitation of molecular tracers is currently severely hampered by the lack of a suitable hydrodynamic wind model coupled to chemistry and to dust evolution models (which dominate the opacity in the wind) to interpret the observations. While a number of chemical models exist of the deeper, denser regions of discs, no model is currently available for the optically thinner disc winds. Current detailed chemical calculations which extend to the disc atmosphere use a hydrostatic disc model analysed in a 1+1D fashion [101]. Some studies of chemistry along MHD disc wind streamlines for Class 0–II have been presented however and show promising results [88,143]. In the future, one can thus hope to use similar techniques coupled to a more generalized approach for the calculation of disc wind to better exploit molecular observations to constrain the wind driving mechanism and directly measure important wind properties.

4. Observations versus theory: the true population of dispersing discs

Before comparing observations and theory to identify discs that are truly caught in the act of dispersing, we briefly summarize the salient results discussed in previous sections. We also schematically illustrate the three main stages of disc evolution and dispersal in figure 5. Disc dispersal processes, accretion and winds, act throughout the disc lifetime as evinced from observations of objects in different SED classes. Once accretion becomes undetectable most of the dust and gas at large disc radii are also dispersed. Disc dispersal in low-mass star-forming regions appears to occur from the inside-out as indicated by the detections of dust cavities and hints of gas cavities and gaps. There is also tentative evidence for the evolution of disc winds, with the possible disappearance of MHD winds as discs evolve (see right panels in figure 5). The transition from disc-bearing to disc-less is fast, approximately 10% of the disc lifetime.

While theoretical models combining accretion and disc winds are consistent with these basic observations and reproduce the lifetime and dispersal of discs, several authors have pointed out that the theory fails to explain a subset of the observed transition discs [7]. But are all transition discs truly caught in the act of dispersing?

We make use of the largest classification of transition disc candidates [70], which is based on Spitzer and complementary long-wavelength photometry, to identify different populations of transition discs in the mass accretion rate ($\dot{M}_{\text{acc}}$) versus hole size ($R_{\text{hole}}$) plot [144]. Stellar parameters are also provided in this work [70], including spectral type and luminosity (Table A.2 of [70]), an estimate of $R_{\text{hole}}$ from SED modelling, which is insensitive to $R_{\text{hole}} \leq 1$ AU, and total disc masses (assuming a gas-to-dust ratio of 100), also obtained from SED modelling (Table A.3 of [70]). From this sample of approximately 200 transition disc candidates, we identified a subsample of 72 objects with literature mass accretion rates including upper limits and estimated dust holes larger than 1 AU. This subsample is plotted in figure 6 and values used for the plot, including references, are provided in Table A.1 of [70]. $^4$Objects are colour coded by disc mass, spectral types G and earlier are surrounded by a light blue circle, and upper limits in $\dot{M}_{\text{acc}}$ have a down-pointing arrow. We decided to highlight early spectral type stars because they appear to be mostly high accretors surrounded by discs with large dust holes. Grey squares are snapshots of models of EUV- plus X-ray-driven photoevaporating discs [99] while grey lines are evolutionary tracks for the same photoevaporating discs with an embedded giant planet [145]. The solid and dashed grey lines correspond to two different assumptions for the dust dynamics in the inner disc at the time of gap opening. Solid lines represent models where the dust in the inner disc continues to be replenished from the outer disc after the formation of the gap. In this assumption, small dust grains (in the Epstein regime) are able to move across the gap together with gas. The vertical dashed lines correspond to models where the planet is assumed to trap all grains in the outer disc, such that as soon as a gap forms, the disc is observed to have a dust cavity of size equal to the radius at which the gap was opened by the planet. These latter models also assume that grains in the inner disc drain onto the stars on much shorter timescales than the gas in the inner disc. These models illustrate the range of possible transition discs predicted by photoevaporation. Note how embedded planets can not only open larger dust gaps but can also reduce the transition time [145].

Twenty-nine out of 72 discs (approx. 40%) fall in the area covered by grey squares and within the solid grey lines and, as such, are consistent with cavities being opened by photoevaporation. If these are the

$^4$Note that there are sources with the same $\dot{M}_{\text{acc}}$ and $R_{\text{hole}}$, only 68 individual points can be recognized in figure 6. These sources are (i) J163115.7-24340 and DoAr28, $\log(\dot{M}_{\text{acc}}) = -8.4$ and $R_{\text{hole}} = 20$ AU; (ii) J162802.6-23550, 03442156+321509 and J130455.7-7739, $\log(\dot{M}_{\text{acc}}) < -11$ and $R_{\text{hole}} = 2$ AU; (iii) ASR118 and IC348-67, $\log(\dot{M}_{\text{acc}}) = -10.2$ (non-detection for the second source) and $R_{\text{hole}} = 2$ AU.
only discs caught in the act of dispersing, the disc dispersal timescale is about 40% of that measured from counting all transition discs. However, this number should be taken as a lower limit as upper limits on mass accretion rates are more rarely reported in the literature than mass accretion rate measurements. Overall, these results confirm previous reports that the dispersal timescale is short, of order $10^5$ years. Once the inner disc has drained onto the central star, stage 3 in figure 5, theoretical models predict a roughly equal amount of time to consume the outer disc [99]. At this stage, the star is not accreting (WTT), the hole would be anywhere from a few to approximately 100 AU, and, assuming no grain growth, there should be still enough mass in small grains for the disc to be optically thick. Thirteen out of the 72 discs (approx. 20%) in figure 6 could be in this stage, a factor of 2 lower than the accreting transition discs consistent with photoevaporation. Again, this fraction should be taken as a lower limit because mass accretion rates are often not reported in the literature for non-accreting objects; see the example of the transition disc around J160421.7 with approximately 70 AU hole and non-accreting based on the $H\alpha$ EW.
but no upper limit on the mass accretion rate [70]. In addition, disc photoevaporation models do not currently include grain growth and migration, which does occur in approximately 1–10 Myr period over which discs evolve and disperse (see §2.1 for observables). This means that the fraction of predicted optically thick dust discs is in fact an upper limit, thus reducing the discrepancy between theory and observations for the stage 3 discs.

The population of discs that is most difficult to explain, not only via photoevaporation models, is that of discs that have large dust cavities (greater than 20 AU) and are accreting vigorously, at rates not too dissimilar from those of full discs (approx. $10^{-8} M_\odot$ yr$^{-1}$ [43]: figure 6). This population is dominated by millimetre-bright, i.e. massive, dust discs with many of them around early-type stars [144]. These are unlikely to be short-lived objects caught in the act of dispersing their discs [7]. Their nature is hotly debated in the literature. It has been shown that three to six giant planets are needed to open gas gaps with the observed widths (15–45 AU) and depletion (factors of 10–1000) combined with a very low, rather unusual, disc viscosity (less than 0.001) [146]. In addition, reproducing the approximately 10% transition disc fraction would require extending the incidence of giant planets within approximately 2 AU out to 3–30 AU, which seems unlikely based on current direct imaging statistics [147]. An alternative explanation requires a dead zone, combined with an MHD wind to remove the inner gas in order explain the gap sizes and depletions [148]. However, predicted mass accretion rates are too low in comparison with observed values, a shortcoming also encountered by photoevaporation models. The failure of most models in reproducing the accretion rates boils down to the fact that an $\alpha$-type accretion is assumed in the inner disc, meaning that large gas surface densities are required at the low velocities implied by typical $\alpha$ values. It has recently been proposed that gas accreting at transsonic speeds may help in solving this problem [149]. In this scenario, very low gas surface densities (e.g. they use a midplane density of $n_H = 10^8$ cm$^{-3}$) are required in the inner disc, which may be consistent with recent observational constraints [70]. The argument is based on the fact that accretion driven by MHD winds may indeed produce the required accretion speeds if the low density gas in the inner disc has a level of ionization such that it is sufficiently coupled to the magnetic field, but not so much as to drag the magnetic flux inwards. However, they also point out that this picture cannot be applied to full discs or in the outer regions of discs with cavities as the implied gas surface densities are too low in comparison with those inferred observationally via millimetre observations.

5. Consequences for planet formation

The implications and consequences of disc dispersal for the formation and evolution of planetary systems have been discussed in a recent review published in the Protostar and Planets VI (PPVI) conference proceedings [83]. In what follows we aim at providing an update, covering work that has appeared since this review. For the sake of consistency we structure the discussion according to the (i) chemical effects and impact on planet formation and (ii) dynamical effects.

5.1. Chemical effects and impact on planet formation

One aspect of the influence of photoevaporation on the formation of planetesimals which was considered in the PPVI review is the selective removal of virtually dust free material. Grains in the wind launching region will be lifted in the wind according to the force balance between the drag force [150], gravity and the centrifugal force. Small (approx. micrometre size) grains are expected to be entrained in the wind [151–153], but the majority of the solid mass will remain in the disc. This effect is further amplified by considering that due to settling [154] the wind launching region will only be populated by sub-micrometre grains which can be brought up to the base of the wind via turbulence [155]. Photoevaporation is thus expected to increase the dust to gas ratio in the disc, which is an important factor in triggering planetesimal formation via the streaming and gravitational instabilities [156]. There have been studies investigating the role played by internal EUV + external FUV photoevaporation [157] and internal EUV-only photoevaporation [158] in the formation of planetesimals. Both studies concluded that while planetesimal formation could be triggered in both cases, it would be at a later time in the disc evolution, when the gas surface density is too low to allow the formation of giant planets. While recent observations of ringed structures in young protoplanetary discs suggest that the planet formation process may have already begun, well before photoevaporation becomes dominant, this process remains relevant for the formation of terrestrial planets or debris disc [159].

Elements of a one-dimensional dust evolution model [160] have been bootstrapped onto a 1+1D chemical model to investigate the interplay between dust evolution and disc dispersal [126]. This led...
Figure 7. The colour map shows the vertically integrated dust-to-gas ratio in the disc at times just before and after the opening of the gap by X-ray photoevaporation. The overlain black contours show the location where the Stokes numbers are between values of 0.01 and 0.1.

to the conclusion that dust evolution hardly affects the disc lifetimes, while on the other hand, large reductions (approx. 2–10) in the gas/dust mass ratio are achieved due to photoevaporation, suggesting that this process may make conditions favourable for the formation of planetesimals by instabilities.

It should be noted that dust to gas ratio is not the only necessary ingredient to allow convergence of radial drift to form particle clouds leading to the streaming instability. The size of the particles (or better the Stokes number) also plays an important role. It has been shown that for metallicities (column integrated dust-to-gas ratios) approximately greater than $10^{-2}$, Stokes numbers of order $10^{-2}$ are sufficient to trigger the instability [161]. More recent work extends the condition down to Stokes numbers of order $10^{-3}$ for similar metallicities [162]. Very recent calculations of dust evolution and radial drift in a viscously evolving disc affected by X-ray photoevaporation [163] suggest that these criteria may easily be met for standard disc and dust models [98,164]. Figure 7 shows a colour map of the dust to gas ratio with overlain a black contour showing the location where the Stokes numbers are between values of 0.01 and 0.1. In this model, the low dust to gas mass ratio at 2 Myr ($\sim 10^{-4}$ instead of the canonical $\sim 10^{-2}$) is due to radial drift which has removed the large particles (which contain most of the dust mass) from the disc. The gap opens shortly after 3 Myr and creates a region of pressure maximum inside the gas-rich inner edge of the outer disc where the conditions may become favourable for triggering the streaming instability. This region moves out as the hole grows, possibly leaving behind a population of planetesimals which may provide the seeds for a terrestrial planetary system in tight configuration, reminiscent of the compact multi-transiting planet systems discovered by Kepler (but note that a similar configuration can also be obtained through different physical processes [165]). More detailed models and more thorough investigation of the parameter space are now needed before more robust conclusions can be drawn. The larger mass loss rates and the more extended photoevaporation profile, characteristic of an X-ray driven wind, solve the problem encountered by previous calculations where, without an efficient particle trapping mechanism [166], radial drift depleted the outer region of an evolving disc before photoevaporation could operate [150,167].

Further effects of internal EUV and external FUV photoevaporation on the chemical composition of the material in the disc available for planet formation have also been explored [168,169]. These works show that the preferential removal of H/He rich gas by photoevaporation could result in the gradual enrichment of refractory elements in the disc midplane. This process could explain the larger than expected Ar, Kr and Xe abundances measured in the atmosphere of Jupiter by the Galileo probe [170]. The models presented to date do not include internal X-ray and FUV photoevaporation, which, due to their larger rates, may yield significant enrichment without the need to invoke an external FUV component. In a very recent study, models have been presented for the formation of the Solar System incorporating disc evolution, pebbles and gas accretion, type I and II migration, simplified disc photoevaporation and Solar System chemical measurements [171]. The models show that photoevaporation of the protoplanetary disc was needed to explain the formation of Jupiter and Saturn with all their constraints. These findings motivate future work to incorporate a more realistic photoevaporation prescription in models for the formation of our Solar System.
5.2. Dynamical effects on the formation and evolution of planetary systems

Trapping dust particles by vortices in discs is an attractive avenue to create ‘hotspots’ for planet formation. The possibility of vortices created at the gap edges in transitional discs by planet formation or photoevaporation has recently been investigated [172,173]. In theory both scenarios involve a pile up of gas at the gap edge which could be then subject to the Rossby-wave instability, leading to the formation of a vortex.

The removal of gas from the protoplanetary disc spells the end for gas giant formation, and it provides a natural stop to migration processes involving exchange of angular momentum between the planets and the gas in the disc. The general consequences of gas removal for the formation and evolution of planetary systems have been explored in a number of seminal works reviewed during PPVI [83]. In the last few years, progress has been made on the modelling of giant planet migration within a photoevaporating disc [174], and it has been shown that the details of the photoevaporation profile, i.e. the driving radiation, as well as magnitude of the mass loss rates may leave detectable signature on the observed semi-major axis distribution of giant planets. That is to say that photoevaporation does not leave a generally predictable signature, as the effects depend on the profile and the total mass loss rates [5]. These models expand on previous work which explored the effects of EUV photoevaporation on planetary system architectures [3,175] by including disc dispersal by X-ray photoevaporation. Recent simulation work by Coleman & Nelson [176] aims at understanding the diversity of observed planetary systems, in particular of the two observed giant planet populations: hot and cold Jupiters. They conclude that a radially structured protoplanetary disc, due to the combination of magnetospheric cavities and photoevaporative winds is needed to at least qualitatively explain the observations. The predictive power of these models remains, however, limited due to uncertainties implicit in the approach, specifically on the rate of mass and angular momentum accretion across the planet’s gap [177].

Moving onto later times of dynamical evolution, recent work has shown that the disc dispersal mechanism can influence migration histories of planetary embryos, finally determining the formation of hot super-Earths versus giant planet cores [178]. This results from resonant chains being disrupted by late dynamical instabilities triggered by the dispersal of the gaseous disc. The simulations can reproduce many of the exoplanet trends, including the observed giant exoplanet–stellar metallicity correlation and the lack of such correlation for super-Earths [179–181].

More recent works also highlight the importance of disc dispersal in understanding the migration histories of planetary embryos [182,183].

6. Summary and concluding remarks

We have summarized empirical constraints on the evolution and dispersal of planet-forming discs and discussed theoretical models that offer a physical interpretation. Particular emphasis was placed on the subset of discs that may be caught in the act of dispersing, often called transition discs. As it transpires from this review, observations and theory have advanced significantly in just the past few years and are challenging classical views of disc evolution and dispersal. By confronting theory with observations we can draw the following main conclusions:

— Current data are consistent with accretion being the main disc dispersal mechanism on ~Myr timescales but cannot prove whether disc viscosity or MHD winds drive accretion.

— The relatively long disc lifetime (a few Myr) but short disc, dispersal timescale (approx. $10^5$ yr) can be explained by: (i) photoevaporation taking over long-lived viscously evolving discs, (ii) MHD winds if the inner disc can retain most of its magnetic flux while the outer disc loses it during evolution, or (iii) a combination of the two, where a thermally driven wind, threaded by magnetic lines, exerts a torque on the gas remaining in the disc and drives accretion.

— Disc winds are now directly detected toward many young stars. Winds arising inside the EUV gravitational radius for 10,000 K gas are most likely MHD driven while those tracing material further out may be MHD or photoevaporative in origin.

— The SED-identified transition discs are a heterogeneous group of systems. About 50% of them can be explained by models that include viscous evolution and photoevaporation. The other 50%, those that have large mass accretion rates (approx. $10^{-8} \, M_\odot \, yr^{-1}$) and disc masses, might not be truly dispersing discs and remain difficult to account for in the classical $\alpha$ disc paradigm.
— Gas diagnostics are starting to reveal cavities and gaps at radial distances comparable to dust cavities in transition discs providing valuable constraints on the ways discs evolve and disperse.

At this stage, clarifying the role of accretion and disc winds in the evolution and dispersal of planet-forming discs is critical to make progress in the field and connect the first stages of planet formation to mature planetary systems. Measurements of turbulence in discs will be key to assess the role of viscous accretion. Luckily, they are becoming possible thanks to the sensitivity and spatial resolution of ALMA and multiple approaches are being tested to understand uncertainties and biases [176]. Expanding the set of wind diagnostics to molecular tracers will be also invaluable to clarify the origin of disc winds beyond the EUV gravitational radius as only FUV-driven photoevaporative winds can be molecular and these are launched beyond approximately 2 AU [101]. Finally, extending the measurements of magnetic fields to disc sizes and mapping their evolution in time seem necessary to test if MHD winds can truly drive accretion and disperse discs from inside-out as observed. As such measurements remain challenging even with current facilities, in the short term empirical constraints may come from mapping the evolution of MHD and photoevaporative winds traced via gas emission lines.

From the theory side, there are also several important directions that need to be pursued. As discussed in §3.2, photoevaporative models are mature but still incomplete. A longer-term but challenging goal would be to build models that include all stellar heating sources, treat disc chemistry and dust evolution, and include hydrodynamics to properly determine wind rates and profiles as well as to predict emission line profiles. Global MHD disc simulations, with enough vertical extent, are still needed to demonstrate that MHD winds can drive accretion and disc dispersal. Predictions that could distinguish this mode of evolution from viscously evolving photoevaporating discs are also necessary to make progress in the field. The relative extent of the gas and dust and gas depletions are some of the diagnostics that are becoming accessible to observers and might be promising to establish how discs evolve and disperse.

Finally, the last few years have seen several attempts at quantitatively accounting for disc dispersal in planet formation and evolution models. What is apparent is that disc dispersal is not only a boundary condition to planet formation models, but many of the processes involved can play a powerful role in the assembly of material to make planetesimals, as well as in the evolution of young planetary systems. Thus, the era of quantitative planet formation modelling can only truly begin once a more quantitative picture of disc dispersal is established via comprehensive and observationally constrained theoretical modelling.

Data accessibility. The datasets supporting this article have been uploaded as part of the electronic supplementary material.

Authors’ contributions. B.E. contribution focused on the description of theory, whereas I.P. contribution focused on the description of the observations. All authors gave final approval for publication.

Competing interests. We declare we have no competing interests.

Funding. B.E. is funded by a professorship at the Ludwig-Maximilians University of Munich, Germany. I.P. acknowledges support from an NSF Astronomy and Astrophysics Research Grant (ID: 1312962). The review also benefitted from collaborations and/or information exchange within NASA Nexus for Exoplanet System Science (NExSs) research coordination network sponsored by NASA Science Mission Directorate.

Acknowledgements. We thank Cornelis Dullemond and a second anonymous referee for their comments which helped improve the review. We also thank J. Owen, G. Picogna and G. Rosotti for providing transition disc model tracks and synthetic line profiles and S. Edwards, U. Gorti and N. van der Marel for useful discussions.

References


