An overview of our current understanding of the formation and evolution of star clusters is given, with the main emphasis on high-mass clusters. Clusters form deeply embedded within dense clouds of molecular gas. Left-over gas is cleared within a few million years and, depending on the efficiency of star formation, the clusters may disperse almost immediately or remain gravitationally bound. Current evidence suggests that a small percentage of star formation occurs in clusters that remain bound, although it is not yet clear whether this fraction is truly universal. Internal two-body relaxation and external shocks will lead to further, gradual dissolution on time scales of up to a few hundred million years for low-mass open clusters in the Milky Way, while the most massive clusters (> $10^5 M_\odot$) have lifetimes comparable to or exceeding the age of the Universe. The low-mass end of the initial cluster mass function is well approximated by a power-law distribution, \(dN/dM \propto M^{-2}\), but there is mounting evidence that quiescent spiral discs form relatively few clusters with masses \(M > 2 \times 10^5 M_\odot\). In starburst galaxies and old globular cluster systems, this limit appears to be higher, at least several \(\times 10^6 M_\odot\). The difference is likely related to the higher gas densities and pressures in starburst galaxies, which allow denser, more massive giant molecular clouds to form. Low-mass clusters may thus trace star formation quite universally, while the more long-lived, massive clusters appear to form preferentially in the context of violent star formation.

Keywords: galaxies: star clusters; globular clusters: general; open clusters and associations

1. Introduction

The appeal of star clusters as tools to study (extra)galactic star-formation histories is at least twofold: the most massive clusters tend to be long-lived, and therefore potentially carry information about the entire star-formation histories of their host galaxies. Furthermore, they are bright and can be observed at much greater distances than individual stars. Young, massive, compact clusters have now been observed in many external galaxies, and it is commonly assumed that at least some of these are young counterparts of the ancient globular clusters (GCs) which are ubiquitous in all major galaxies. Thus, the view that GC formation...
required unique physical conditions in the early Universe (e.g. Peebles & Dicke 1968) has largely been abandoned. However, it remains much less clear how direct the link is between star formation and the formation of (massive) clusters.

In general terms, we might ask the question, What are the conditions for the formation and survival of massive clusters? Clearly, finding an answer is essential if we wish to use such clusters effectively as probes of galaxy formation and evolution. This contribution begins with a broad (and, by necessity, highly incomplete) overview of what is known about the overall properties of star cluster systems. It should then become clear that it is difficult to answer the question as stated above, as there is no clear-cut physical criterion that allows us to decide when a cluster should be classified as ‘massive’. It is therefore useful to rephrase the problem in a way that makes it more tractable. The remainder of the current review will thus focus on three main themes: (i) the general problem of cluster formation, (ii) dynamical evolution, and (iii) the shape of the initial cluster mass function (ICMF).

2. Basic observational results

(a) Open and globular clusters in the Milky Way

In the context of this volume, it is appropriate to recall that the first comprehensive discussion of the properties of star clusters was given by Sir William Herschel in a series of papers published in the Philosophical Transactions of the Royal Society of London. Herschel noted significant differences in the visual appearances of clusters and used the term GCs to describe the richest and most concentrated of them (Herschel 1814). The term open cluster emerged during the early twentieth century (Shapley 1916) as a common label for all non-GCs. Originally, this classification was purely morphological, based simply on the visual appearance of a cluster through a telescope or on a photograph. Differences in spatial distribution, with the open clusters concentrated near the Galactic plane and the GCs tending to avoid it, were recognized early on (Shapley 1916 and references therein). The open clusters are, in general, metal-rich with metallicities similar to or even exceeding the solar value (Friel et al. 2002), while the Milky Way GC metallicity distribution is bimodal, with both peaks at subsolar values (logarithmic iron abundance, relative to solar, of \[\text{[Fe/H]} \approx -1.5\] and \[-0.5\] dex; Zinn 1985). In modern terms, these differences reflect the association of the open clusters with the disc of our Galaxy and the GCs with the spheroid (bulge/halo).

While the GCs are all ancient, with ages of the order of \[10^{10}\] years and a spread of perhaps a few \[\times 10^9\] years (Marín-Franch et al. 2009), the open clusters are mostly younger than a few \[\times 10^8\] years (Wielen 1971), although some older open clusters are also known (Friel 1995). The lack of young GCs in the halo and bulge can be attributed to a cessation of star formation in these components long ago, but the field stars in the Galactic disc have a continuous range of ages, and open clusters are likely to have formed there also in the distant past. The relative deficit of old open clusters, therefore, illustrates that cluster dissolution is important.

The mass functions (MFs) of open and globular clusters are strikingly different. The MF of young open clusters can be fitted by a power law, \[dN/dM \propto M^{-2}\], down to a few hundred M\(_\odot\) (Elmegreen & Efremov 1997; Piskunov et al. 2008). In contrast, the GC MF is approximately flat at low masses, \[dN/dM\] approximately
constant for $M < 10^5 \, M_\odot$ (McLaughlin & Pudritz 1996), while the high-mass end can be fitted by a power law with slope approximately $-2$, as for the open clusters. Unless the GCs were born with a different MF from young clusters today, this flattening is another hint at the importance of dynamical evolution.

Even the most massive open clusters identified in the Galactic disc have masses below approximately $10^5 \, M_\odot$ (Davies et al. 2007; Brandner et al. 2008; Froebrich et al. 2009), an order of magnitude lower than the most massive old GCs (Meylan & Mayor 1986). Clearly, this difference cannot easily be attributed to dynamical evolution, and seems all the more puzzling given that the mass of the Milky Way disc exceeds that of the spheroid by about an order of magnitude (Dehnen & Binney 1998). We return to this issue below, but note here that GCs and open clusters are subject to very different selection biases.

Extinction of optical light by interstellar dust in the Galactic plane, combined with the high stellar density (‘crowding’) along a given line of sight, strongly limits our ability to detect distant open clusters. In fact, the discrepancy between ‘diameter distances’ (unaffected by extinction) and ‘photometric’ distances of open clusters led to one of the first quantitative estimates of the amount of dust extinction in the Galactic plane (Trumpler 1930). Current catalogues of open clusters can only be considered reasonably complete within approximately 1 kpc of the sun (Lamers et al. 2005; Piskunov et al. 2008). It is reasonable to assume that the most massive and most luminous objects can be detected out to greater distances, but the actual completeness of current surveys remains poorly quantified. GCs are, instead, easier to find. Some may still remain hidden in the plane, but most are found at higher Galactic latitudes, where extinction and crowding are less severe, and the spatial distribution of known GCs shows no tendency to concentrate near the sun (e.g. Frenk & White 1982).

(b) The Local Group

The Magellanic Clouds are our nearest large extragalactic neighbours. It has long been known that both Clouds, and the Large Magellanic Cloud (LMC) in particular, are home to large numbers of star clusters. Although the LMC is about a factor of 10 less luminous than the Milky Way (van den Bergh 1999), we can view it in its entirety. The angular extent is about $5^\circ \times 5^\circ$, corresponding to about $5 \times 5 \, \text{kpc}^2$ (adopting a distance of about 50 kpc), a significantly larger area than covered by open-cluster surveys in the Milky Way.

The nomenclature used for LMC clusters has been a source of some confusion. Many of the richest clusters were listed as globular when discovered by John Herschel, while Shapley (1930) identified eight ‘true’ LMC GCs. However, all but one of these (NGC 1835) are now known to be much younger than the Milky Way GCs. Many modern studies tend to use old age as a defining criterion when identifying the GCs in the Clouds (and elsewhere). By this metric, the LMC has about 13 GCs (e.g. Schommer et al. 1992) while the Small Magellanic Cloud (SMC) has one. This criterion works well because of a large gap between ages of approximately $3 \times 10^9$ and $10^{10}$ years that naturally separates the old LMC GCs from younger clusters (Rich et al. 2001).

The luminosity function (LF) of the old LMC clusters is similar to that of Galactic GCs (Harris 1991), suggesting that the MFs are also similar. The most massive young LMC clusters have masses of approximately $10^5 \, M_\odot$.
(Fischer et al. 1992), somewhat exceeding the most massive young clusters currently known in the Milky Way, but this might simply be a size-of-sample effect because of the larger area surveyed and the higher surface density of clusters in the LMC (Larsen 2002). There is no evidence of significant differences in the actual MFs and LFs of the young LMC and Milky Way clusters, with both following similar power-law distributions (Elson & Fall 1985; Hunter et al. 2003; de Grijs & Anders 2006).

An interesting contrast to the rich cluster systems of the Magellanic Clouds is provided by the dwarf irregular galaxy IC 1613. In spite of some ongoing star formation, this galaxy contains very few star clusters in comparison with the Clouds, even when accounting for its somewhat lower luminosity (van den Bergh 1979).

The two other Local Group spirals, M31 and M33, host populations of young and old star clusters that appear to be roughly equivalent to the Milky Way open and globular clusters (Galleti et al. 2006; Sarajedini & Mancone 2007). The catalogues remain highly incomplete, however, and the nature of many candidates remains to be verified. To date, about 350 GCs have been confirmed in M31, while about 30 old, GC-like objects have been identified in M33 (Schommer et al. 1991; Chandar et al. 2001; Huxor et al. 2009). Identification of clusters superimposed on the discs is challenging because of crowding and confusion issues (Cohen et al. 2005), so that information about the global properties of young cluster populations in the discs of M31 and M33 is still relatively scarce. However, both spirals host rich, young star cluster populations similar to those observed in the LMC, with estimated masses of up to $10^4$–$10^5 \, M_\odot$. The LFs and MFs are not well constrained, but appear consistent with those in the Milky Way and the LMC (Chandar et al. 2001; Sarajedini & Mancone 2007; Caldwell et al. 2009).

(c) Beyond the Local Group

Observations of extragalactic young clusters have been reviewed on many previous occasions (e.g. Whitmore 2003; Larsen 2006). Identification of star clusters in star-forming galaxies beyond the Local Group is challenging on the basis of ground-based observations. In their study of young clusters in external galaxies, Kennicutt & Chu (1988) listed data for 14 galaxies, of which half are members of the Local Group. The launch of the Hubble Space Telescope (HST) led to a revolution in the field, starting with the discovery of a large number of bright, blue compact star clusters in the galaxy NGC 1275 (Holtzman et al. 1992). With careful modelling of the HST point-spread function, a typical cluster with a half-light radius of approximately 3 pc remains recognizable as an extended object out to distances of at least 40 Mpc (Harris 2009). This leads to a formidable increase in the number of galaxies accessible to detailed study of their cluster populations: Larsen (2006) listed 92 young systems for which data were available as of 2004.

Many of the extragalactic systems that have been studied in detail are starburst and merging galaxies, of which the best-studied case is arguably the ‘Antennae’ system, NGC 4038/4039. Like other ongoing major gas-rich mergers, this pair is experiencing vigorous star formation, including in a large number of luminous, compact star clusters (Whitmore et al. 1999). Another well-studied system is the nearby starburst M82, which also hosts many luminous young clusters, although detailed analysis of their properties is hampered by heavy extinction as the system
is viewed nearly edge-on (O’Connell et al. 1995; de Grijs et al. 2005a; Smith et al. 2007). These cases are fairly typical of the many systems that have been studied with the HST. Where cluster masses have been derived, they are often in the range $10^4$–$10^6 \, M_\odot$ or higher, comparable to the most massive old GCs (Zhang & Fall 1999; McCrady & Graham 2007), with the lower end of the range usually being set by detection limits.

An increasing amount of data for normal spiral galaxies have also become available. Young clusters in the mass range $10^5$–$10^6 \, M_\odot$ have been found in some spirals (Larsen & Richtler 2000, 2004), showing that such objects are not unique to starbursts and interacting systems, although they may be more common there. It is worth emphasizing that most spirals in which young cluster populations have been studied are of Hubble-type Sb or later, while little is known about young clusters in Sa-type spirals.

The above discussion has concentrated on age and mass as the main parameters characterizing star cluster properties. Another useful parameter is the half-light radius, $R_h$, which is expected to remain approximately constant over the lifetime of a cluster (Spitzer 1987). This is typically a few pc for both open and globular clusters, independent of mass, although for GCs $R_h$ tends to correlate with galactocentric distance and GCs as large as $R_h > 20$–$30$ pc exist in the outer haloes of the Milky Way, M31 and M33 (van den Bergh et al. 1991; Huxor et al. 2008, 2009). For very massive clusters, there appears to be a more significant positive correlation between cluster mass and size (Harris 2010). Unusually extended ($R_h > 7$ pc) old clusters have also been found in several S0-type galaxies (Larsen & Brodie 2000; Hwang & Lee 2006; Peng et al. 2006). These ‘faint fuzzy’ clusters (FFs) are distinctly different from the outer-halo GCs in the Local Group spirals, as they are clearly associated (kinematically and spatially) with the discs of their parent galaxies in at least a few cases (Burkert et al. 2005). Their LFs also differ from those of normal old GCs by showing no ‘turnover’ near absolute visual magnitude $M_V \sim -7.5$ (current data do not constrain the LFs of FFs fainter than $M_V \sim -6$ mag). This may provide an important clue to the importance of different disruption mechanisms in the discs of S0 galaxies, as such extended clusters would be less affected by internal dynamical evolution owing to two-body relaxation, but more sensitive to external shocks (Vesperini 2010). It remains largely unknown whether FFs form through a special channel that operates predominantly in the discs of S0 galaxies, or whether this environment is particularly favourable for their survival. An interesting possibility is that FFs may have formed by the merger of smaller subunits (Burkert et al. 2005; Fellhauer & Kroupa 2005), perhaps similar to the star-cluster complexes observed in some galaxies with ongoing cluster formation (Bastian et al. 2005a).

3. Formation of clusters

The problem of cluster formation is intimately linked with that of star formation. There are many excellent reviews on this topic (e.g. Lada & Lada 2003; Mac Low & Klessen 2004; McKee & Ostriker 2007), and the discussion in this section will concentrate on a few issues of relevance to the global properties of cluster populations. More detailed discussion can be found elsewhere in this volume (Clarke 2010; Lada 2010).

*Phil. Trans. R. Soc. A* (2010)
(a) From giant molecular clouds to (embedded) star clusters

Star formation is closely associated with dense molecular gas, and clusters are observed to form deeply embedded within giant molecular clouds (GMCs). The structure of these GMCs is self-similar on a wide range of scales down to individual protostellar cores, which can be grouped into cluster-forming clumps (Williams et al. 2000; McKee & Ostriker 2007). The GMCs are themselves part of a larger hierarchy of structure in the interstellar medium (Elmegreen & Falgarone 1996; Elmegreen 2007), and tend to be organized into giant molecular complexes (Wilson et al. 2003), which are located along the spiral arms in spiral galaxies (Vogel et al. 1988). Averaged over an entire GMC, the density of molecular gas is low \( n_H \sim 10^2 \text{–} 10^3 \text{ cm}^{-3} \) and stars form only in the densest regions \( n_H > 10^5 \text{ cm}^{-3} \). Globally, star formation is therefore an inefficient process, and only a few per cent of the mass of a given GMC is converted into stars before the cloud is dispersed (Williams & McKee 1997). The GMC-wide star-formation efficiency, \( \epsilon_{\text{GMC}} \), should not be confused with the local star-formation efficiency, \( \epsilon_{\text{cl}} \), within the cluster-forming clumps, which must be at least 20–30% to produce a bound cluster (§4).

An interesting question is how the MF of GMCs is related to that of the clusters forming within them. The GMC MF in the Milky Way has a characteristic upper mass of approximately \( 6 \times 10^6 \text{ M}_\odot \). Below this mass, it can be approximated by a power law, \( dN/dM \propto M^{-1.7} \), but it declines steeply at higher masses (Williams & McKee 1997). For \( \epsilon_{\text{GMC}} \) approximately 5 per cent, the upper GMC mass would correspond to a cluster mass of approximately \( 3 \times 10^5 \text{ M}_\odot \), although the ICMF is unlikely to be a simple scaled-down version of the GMC MF, as a single GMC may form more than one cluster (e.g. Kumar et al. 2004). The mass spectrum of clumps within GMCs may be more relevant. This appears to follow a similar power law to that of the GMCs, \( dN/dM \propto M^\alpha \), with \( \alpha \approx -2 \) but with a large uncertainty (Mac Low & Klessen 2004). Nevertheless, an upper limit of a few \( \times 10^5 \text{ M}_\odot \) is consistent with other constraints on the ICMF in spiral galaxies (§5; Larsen 2009).

It is difficult to see how the most massive clusters observed in some external galaxies, with masses of \( 10^6 \text{ M}_\odot \) or higher, could form from Milky Way-like GMCs. For a \( 10^6 \text{ M}_\odot \) cluster with a half-mass radius of 3 pc, the current mean density within the half-mass radius corresponds to \( \langle n_H \rangle = 2 \times 10^5 \text{ cm}^{-3} \). This is a strict lower limit to the density of the gas from which the cluster must have formed, as \( \epsilon_{\text{cl}} < 1 \) and clusters expand following gas expulsion (Goodwin & Bastian 2006; Scheepmaker et al. 2007; Bastian et al. 2008; Pfalzner 2009). The formation of the most massive clusters requires collecting the amount of gas typical of a massive Galactic GMC within a volume only a few pc across, essentially turning such a cloud into one big clump. A likely key element to understanding how such dense, massive clumps can exist is the high gas densities in starburst galaxies. This may allow denser and more massive GMCs to condense (Escala & Larson 2008), perhaps further aided by shock compression in mergers (Jog & Solomon 1992; Ashman & Zepf 2001). There is observational evidence that GMCs in M82 are indeed compressed to higher densities than their Milky Way counterparts by the ambient pressure (Keto et al. 2005), so that in these clouds massive clusters may form at high \( \epsilon_{\text{GMC}} \). In even more extreme environments, such as in ultraluminous infrared galaxies, GMCs may be
both denser and more massive (Murray et al. 2009), consistent with the presence of clusters with $M > 10^7 M_\odot$ in some merger remnants (Maraston et al. 2004; Bastian et al. 2006).

(b) The embedded phase

Observations of the embedded phase are challenging (see Lada 2010) because of its short duration, combined with the high gas column densities in GMCs ($N_H \sim 1.5 \times 10^{22} \text{ cm}^{-2}$; McKee & Ostriker 2007) and corresponding large amounts of dust extinction ($A_V \sim 8 \text{ mag}$). The duration of this phase is uncertain, as it is difficult to determine the age of (unresolved) embedded clusters and the assumption of a single age may be questionable in the first place. Upper limits may be set by age dating clusters that have already become optically visible. The R136 cluster in the LMC contains some stars as young as 1–2 Myr which have mean extinctions of only $A_V \sim 1.2 \text{ mag}$ (Massey & Hunter 1998). Whitmore & Zhang (2002) found optically visible counterparts for about three-quarters of the brightest radio-continuum sources in the Antennae. Among these, clusters older than approximately 2.5 Myr all have low extinctions of $0.5 < A_V < 2.5 \text{ mag}$. Similarly, Reines et al. (2008) found that radio-detected clusters in NGC 4449 with ages in the range of 3–5 Myr already have low extinctions, $A_V = 0.5–1.5 \text{ mag}$. The bright cluster NGC 1569–A, with an age of approximately 5 Myr (Origlia et al. 2001; Maoz et al. 2001), is essentially free of dust extinction. From these examples, it is clear that the embedded phase lasts at most a few $\times 10^6$ years.

What physical mechanism is responsible for expelling the gas? The ionizing radiation from massive stars will produce an HII region, but the thermal pressure in the ionized gas may be insufficient to overcome self-gravity for clusters more massive than approximately $10^5 M_\odot$ (Kroupa & Boily 2002). Various alternatives are discussed by Krumholz & Matzner (2009). Supernovae could easily provide enough energy but appear after several $\times 10^6$ years, probably too late to explain the observed short duration of the embedded phase. Another candidate is winds from massive stars, which have velocities exceeding 1000 km s$^{-1}$ and also provide more than sufficient energy to unbind even massive ($M > 10^6 M_\odot$) clusters. However, the efficiency of such winds may be low. Krumholz & Matzner (2009) finally concluded that radiation pressure from massive stars is most likely the dominant gas-evacuation mechanism in massive clusters.

4. Dynamical evolution

(a) Early dynamical evolution: ‘infant mortality’

The formation of an embedded cluster does not guarantee that it will remain bound after gas expulsion (see also the discussion in Lada (2010)). If the stars and gas are initially in virial equilibrium, the velocity dispersion of the stars will be too high to match the shallower potential once the gas is expelled. If gas expulsion happens instantaneously and $\epsilon_{cl} < 50\%$, the cluster will dissolve completely, independent of the initial mass (Hills 1980). In practice, cluster expansion does not occur instantly so the stars have some time to adjust to the new potential, while gas expulsion is probably not instantaneous either. Simulations suggest that at least some fraction of the stars may remain bound for

Phil. Trans. R. Soc. A (2010)
lower star-formation efficiencies, perhaps as low as $\epsilon_{\text{cl}} \approx 20\text{–}30\%$ (Goodwin 1997; Boily & Kroupa 2003; Baumgardt & Kroupa 2007). The time scale for the cluster to settle into a new equilibrium may be as long as several $\times 10^7$ years (Goodwin & Bastian 2006).

This picture is supported by the observation that about 95 per cent of clusters formed in Milky Way GMCs dissolve in less than $10^8$ years (Lada & Lada 2003). However, the universality of this ‘infant mortality’ (IM) fraction is poorly quantified. The term ‘mass-independent disruption’ (MID) is also sometimes used to distinguish this process from the secular, mass-dependent dissolution that occurs on longer time scales and which will be discussed below. The age distribution of mass-limited cluster samples in the Antennae galaxies is approximately $\frac{dN}{d\tau} \approx \tau^{-1}$ for ages ($\tau$) of up to $10^8\text{–}10^9$ years (Fall et al. 2005), suggesting that approximately 80–90% of the clusters disappear per decade in age, independent of mass (Whitmore et al. 2007). However, over such a large age range, it is unlikely that disruption can still be attributed to gas expulsion. One difficulty with estimating the disruption parameters in an interacting system like the Antennae is that the star- and cluster-formation rates may not have been constant in the past. Bastian et al. (2009) found that the age distribution of clusters in the Antennae can also be fitted by a model in which the cluster-formation rate has increased over the past few $\times 10^8$ years (as suggested by simulations of the ongoing interaction), with MID required only over a period of approximately $10^7$ years. Evidence of a large IM fraction (approx. 70%) has also been claimed in M51 (Bastian et al. 2005$b$), but this may be partly due to age-dating artefacts around $10^7$ years (Gieles 2009). In the SMC, Chandar et al. (2006) derive an age distribution of $\frac{dN}{d\tau} \sim \tau^{-0.87}$ for $\tau < 10^9$ years, similar to that seen in the Antennae, but Gieles et al. (2007) argue that this may be caused by fading below the detection limit at old ages and instead conclude that no significant MID is needed to explain the age distribution of SMC clusters (see also de Grijs & Goodwin (2008) for evidence supporting the latter scenario).

\[(b)\text{ The cluster-formation efficiency}\]

In starburst and merger systems, young clusters often account for a large fraction (10–20%) of the total blue/ultraviolet flux (Meurer et al. 1995; Zepf et al. 1999; Tremonti et al. 2001; Fall et al. 2005), consistent with essentially all stars forming in clusters (see also Johnson et al. 2009). In the Milky Way, the majority of stars form in embedded clusters (Lada & Lada 2003). However, most stars eventually end up belonging to the field. Observationally, it is difficult to tell whether \textit{all} stars were born in clusters, with a large fraction dispersing almost immediately, or whether some stars were born in genuinely dispersed mode.

For practical purposes, one may still define a cluster-formation ‘efficiency’ as the ratio of the number of stars that end up in clusters relative to the field. As this ratio will depend on age, ideally some age range should be specified. For optically visible clusters, Bastian (2008) found a constant efficiency of $\Gamma \sim 8\%$ in galaxies spanning six orders of magnitude in star-formation rate (SFR). On the other hand, the fraction of the total $U$-band light in galaxies originating from clusters correlates with the area-normalized SFR of the parent galaxy, ranging from well below 1 per cent in quiescent systems to the high numbers found in starbursts (Larsen & Richtler 2000). The near absence of clusters in
IC 1613 and the non-universal GC-specific frequency (number of GCs per unit host-galaxy luminosity; Harris 1991, 2010) are other hints that variations in $\Gamma$ may exist.

Even if a single indicator of star formation (e.g. the far-infrared or ultraviolet luminosity) is used, different systematic errors will affect galaxies differing in dust content, metallicity, ratio of current to past SFR and other parameters (Kennicutt 1998). Such systematic errors on the parent-galaxy SFRs are probably different from those associated with the cluster-formation rates, which are typically inferred from direct observations of cluster populations but still subject to uncertainties owing to disruption, potential confusion with other objects and completeness effects. Consequently, $\Gamma$ remains difficult to constrain.

(c) Secular evolution

Clusters that survive IM will continue to evolve dynamically on longer time scales as a result of internal two-body relaxation, external shocks and mass loss owing to stellar evolution (Vesperini 2010). This ‘secular evolution’ will lead to the gradual evaporation of any star cluster and, eventually, its total dissolution. Two-body relaxation causes the velocities of the stars in the cluster to approach a Maxwellian distribution, and stars with velocities above the escape velocity will gradually evaporate from the cluster. The two-body relaxation time scales as $t_{\text{rel}} \propto \sqrt{M R_{\text{h}}^{3/2}}$, but the actual evaporation time, $t_{\text{ev}}$, may scale nonlinearly with $t_{\text{rel}}$ (Baumgardt & Makino 2003). In addition, external shocks will lead to dissolution on a time scale $t_{\text{sh}} \propto M R_{\text{h}}^{-3}$ (Spitzer 1987). Shocks may be due to encounters with spiral arms, GMCs or, for GCs on eccentric orbits, passages through the Galactic disc or near the bulge. Finally, stellar evolution causes mass loss as stars are turned into much less massive remnants. Over a 10 Gyr time span, about one-third of the initial cluster mass is lost this way (e.g. Bruzual & Charlot 2003).

For many practical purposes, secular dissolution may be conveniently characterized by a single dissolution time scale $t_{\text{dis}}$, such that mass is lost at a rate $(dM/dt)_{\text{dis}} = -M/t_{\text{dis}}$. The dissolution time scale may be parameterized as $t_{\text{dis}} = t_4 (M/10^4 M_\odot)^{\gamma}$, where $t_4 \sim 10^9$ years and $\gamma \sim 0.65$ for clusters in the solar neighbourhood (Lamers et al. 2005). The relatively larger number of old clusters in the Magellanic Clouds suggests that disruption is less efficient there (Elson & Fall 1985; Hodge 1987; Girardi et al. 1995; de Grijs & Anders 2006), while the disruption time scale in the central regions of M51 appears much shorter than in the solar neighbourhood (Boutloukos & Lamers 2003). This may be caused by different densities of GMCs in these environments (Wielen 1985; Terlevich 1987; Gieles et al. 2006b).

As low-mass clusters disrupt faster, the MF will flatten over time and, given sufficient time, $\tau \gg t_{\text{dis}}$, the MF will tend towards $dN/dM \propto M^{\gamma-1}$. Hence, observations of the MF in cluster systems of different ages can potentially be used to constrain the disruption law. Good fits to the GC LF in the Milky Way and other galaxies can indeed be obtained if cluster MFs similar to those observed in young cluster systems in merging galaxies are evolved with $t_4 \sim 10^8$–$10^9$ years and $\gamma = 0.7 – 1$ (Fall & Zhang 2001; Jordán et al. 2007; McLaughlin & Fall 2008; Kruijssen & Portegies Zwart 2009). However, radial variations in the GC LF are expected, as both shocks and tidal fields will be weaker at large galactocentric radii. The fact that such variations are not observed in old GC
systems is a potential difficulty (Vesperini et al. 2003). Interestingly, observations of the intermediate-age (approx. $3 \times 10^9$ years) merger remnant NGC 1316 do show a radial variation in the MF with a higher turnover mass near the centre (Goudfrooij et al. 2004).

5. The ICMF

The basic stellar dynamical mechanisms responsible for the evolution of the MF are the same for all clusters, even though the relative importance of different mechanisms (e.g. two-body relaxation versus shocks) may differ. However, the ICMF on which these mechanisms operate might still vary with environment. Once observational selection effects are accounted for, it is relatively straightforward to derive the LF of a cluster sample, assuming the distance is known. However, the interpretation of LFs in terms of the physically more fundamental MF is complicated by the fact that not all clusters have the same mass-to-light ratio, $\Upsilon$. A direct conversion from LF to MF is therefore not usually possible. Dissolution makes reconstruction of the ICMF even more challenging.

For samples that are large enough to derive statistically meaningful MFs, the only practical approach is to derive the ages of individual clusters and then make use of ‘simple stellar population’ (SSP) models that tabulate $\Upsilon$ versus age to convert the luminosities to masses (e.g. Bruzual 2010). Ages are typically derived from integrated broadband colours (e.g. $UBVRI$). However, these are sensitive to both age and other parameters such as extinction and metallicity. In principle, the use of multiple filters allows us to solve for all of these parameters, although using different SSP models and fitting methods can still lead to large systematic differences in the derived ages (at least a factor of 2; de Grijs et al. 2005a,b; Scheepmaker et al. 2009). In addition, stochastic effects owing to the finite number of stars in a cluster can lead to large departures from the predicted colours, especially for low-mass, young clusters (Girardi et al. 1995; Bruzual & Charlot 2003; Cerviño & Luridiana 2006; Maíz Apellániz 2009). The requirement for multiple filters is also costly in terms of observing time, especially at blue and ultraviolet wavelengths where detectors tend to be less sensitive.

Determinations of the MF are available for only a few young cluster systems. Generally, they are well represented by power laws, $dN/dM \propto M^\alpha$, with slopes of $\alpha \sim -2.0$, but the mass ranges over which these slopes are derived vary considerably (see Larsen (2009) for references). Figure 1 shows a comparison of the MFs for young clusters in spiral galaxies (taken from Larsen (2009)) and the Antennae system (Zhang & Fall 1999). Ground-based data for 17 spirals have been combined to improve statistics, although 75 per cent of the clusters belong to the two most cluster-rich galaxies, NGC 5236 and NGC 6946. The combination of data for many spirals is justified as the MFs in different subsamples are statistically indistinguishable (Larsen 2009). The Antennae clusters span the range $25 \times 10^6 < \tau < 160 \times 10^6$ years, while clusters younger than $2 \times 10^8$ years are included for the spirals. Completeness limits restrict the useful mass range to $10^5 < M/M_\odot$, but the figure already hints at differences between the MFs, with relatively more high-mass clusters in the Antennae. This is confirmed by a Kolmogorov–Smirnov test, which yields a very small probability ($P = 0.00032$) that the two samples are drawn from the same parent distribution.

Phil. Trans. R. Soc. A (2010)
Figure 1. Mass functions for young (< $2 \times 10^8$ years) clusters in spiral galaxies (Larsen 2009) and the Antennae system (Whitmore et al. 1999). Solid line, spirals; dashed line, Antennae.

The high-mass end of the MF in old GC systems is well approximated by a Schechter (1976) function

$$\frac{dN}{dM} \propto \left( \frac{M}{M_c} \right)^{\alpha} \exp \left( -\frac{M}{M_c} \right),$$

with cut-off mass $M_c >$ several $\times 10^6 M_\odot$ (McLaughlin & Pudritz 1996; Burkert & Smith 2000; Jordán et al. 2007). This should not be confused with the turnover at approximately $10^5 M_\odot$, which is most likely a result of dynamical evolution. A Schechter-function fit to Antennae data in figure 1 yields $M_c = (1.7 \pm 0.7) \times 10^6 M_\odot$ for fixed $\alpha = -2$ (see also Jordán et al. 2007), although a uniform power law with no truncation is also consistent with the data (Whitmore et al. 2007). A fit to the spiral data instead gives $M_c = (2.1 \pm 0.4) \times 10^5 M_\odot$, and a uniform $\alpha = -2$ power law is ruled out at high confidence level (Larsen 2009). This again indicates a dearth of high-mass clusters in spirals, compared with the Antennae.

Although the MF and LF are not the same, they are of course related. If $\psi_i(M_i)$ is the ICMF, normalized to unit mass over some range of initial cluster mass $M_{\text{low}} < M_i < M_{\text{up}}$, the LF is (Larsen 2009)

$$\frac{dN}{dL} = \int_{\tau_{\text{min}}}^{\tau_{\text{max}}} \psi_i[M_i(L, \tau)] \times \frac{dM_i}{dM} \times \Upsilon \times \Gamma \times \text{SFR} \times f_{\text{surv}}(\tau) \, d\tau. \quad (5.2)$$

The rate of star formation in clusters is expressed as $\Gamma \times \text{SFR}$. IM is formally included as a mass-independent survival fraction, $f_{\text{surv}} = (\tau/\tau_0)^{\log(1-\text{IMR})}$, where $\tau_0$ marks the onset of IM and IMR is the fraction of clusters lost per decade in age. For $\tau < \tau_0$, $f_{\text{surv}} = 1$ and IM may be switched off at some time $\tau_{\text{IM,max}}$, after which $f_{\text{surv}}$ is constant. The initial cluster mass $M_i$ is related to the current mass $M = \Upsilon L$ through the assumed disruption law. If there is no disruption and the ICMF is a uniform power law, $\psi(M) \propto M^\alpha$, then it follows from equation (5.2) that the LF is a power law with the same slope. In general, however, the shape of the LF will differ from that of the underlying MF because of disruption and the age-dependent mass-to-light ratio.

*Phil. Trans. R. Soc. A* (2010)
In figure 2, the LFs of clusters in the spiral galaxy M51 (Haas et al. 2008) and the Antennae system (Whitmore et al. 1999) are compared with model LFs for Schechter ICMFs with $M_c = 2 \times 10^5$ (solid line) and $2 \times 10^6 \, M_\odot$ (dashed line). In both cases, an IMR of 80 per cent for ages $(5-30) \times 10^6$ years is assumed, and secular dissolution is modelled using $t_4 = 5 \times 10^8$ years and $\gamma = 0.65$. SFRs are assumed constant, but the model LFs have been shifted vertically to match the data. The relatively small number of clusters in the Antennae compared with M51 originates from using the Whitmore et al. (1999) ‘PC: cluster-rich region’ sample, which covers only a small fraction of the full galaxy pair (Whitmore et al. 1999, fig. 11d). It is clear that the $M_c = 2 \times 10^5 \, M_\odot$ model LF matches the M51 data quite well, while the $M_c = 2 \times 10^6 \, M_\odot$ LF is too shallow at the bright end. For the Antennae, however, the $M_c = 2 \times 10^6 \, M_\odot$ LF provides a better fit. The Antennae LF flattens further below $M_V \sim -8$ mag, but Whitmore et al. (1999) indicate that the selection of cluster candidates becomes less reliable below this limit. M51 is perhaps not the most typical spiral, as it is mildly interacting with a nearby companion. However, the LF comparison suggests that the MF is more similar to that in non-interacting spirals than in the merging Antennae galaxies.

Figure 2 shows that the LF is expected (and observed) to steepen towards the bright end, so that a single power law generally provides a poor fit over an extended magnitude range. This steepening has been noted in other datasets, including the merger NGC 3256 (Zepf et al. 1999) and various spiral galaxies (Dolphin & Kennicutt 2002; Larsen 2002). Typical power-law slopes are between $-2.0$ and $-2.5$. From similar modelling of the LF, Gieles et al. (2006a) inferred a truncation of the MF around $M_{\text{up}} \sim 10^5 \, M_\odot$ in M51 and another spiral, NGC 6946, and it had already been suggested by Zhang & Fall (1999) that the steepening of the Antennae LF brighter than $M_V \sim -10$ mag might indicate truncation of the MF around $10^6 \, M_\odot$. Clearly, the LF has some diagnostic power, although not without relying on model assumptions.
Figure 3. (a) $M_V$ of the brightest clusters in galaxies versus the galaxy-wide SFR. Circles, spiral galaxies; triangles, mergers/interacting galaxies; asterisks, others. Random sampling relations for $M_c = 2 \times 10^5$ (solid line) and $2 \times 10^6 M_\odot$ (dashed line) are also shown (see text for details). The dotted line is the best fit from Weidner et al. (2004). (b) Distribution of differences, $\Delta M_V$, between the observed brightest magnitude $M_V$ (brightest) and the Weidner et al. (2004) fit. The solid curve is the $\Psi$ distribution based on equation (5.3).

(a) A few notes on size-of-sample effects

It is well known that a tight correlation exists between the luminosity of the brightest cluster, $L_{\text{max}}$, in a galaxy and the total number of clusters or the overall SFR (Billett et al. 2002; Larsen 2002; Whitmore 2003). An updated version of the $L_{\text{max}}$–SFR relation is shown in figure 3a, with spiral galaxies shown as filled circles and interacting systems as triangles (Larsen 2002; Bastian 2008). A few dwarf galaxies are indicated by asterisks. The latter are well-known outliers in this context that have been discussed extensively in the literature (Billett et al. 2002; Larsen 2002; Whitmore et al. 2007; Bastian 2008).

This relation is just the type of effect that is expected if cluster luminosities are drawn at random from an LF which decreases towards the bright end. It does not lead trivially to the conclusion that there is a physical correlation of $M_c$ versus SFR. Statistically, the luminosity of the brightest cluster may be estimated by solving $\ln 2 \sim 0.7 = \int_{L_{\text{max,med}}}^{\infty} (dN/dL) dL$, where the constant on the left-hand side is chosen such that $L_{\text{max,med}}$ is the median luminosity of the brightest cluster. For a power-law LF, this leads to the relation $L_{\text{max,med}} \propto N^{-1/(\alpha+1)}$, where $N$ is the number of clusters brighter than some minimum luminosity, $L_{\text{min}}$. The dotted line in figure 3 is the fit $L_{\text{max}} \propto \text{SFR}^{0.748}$ (Weidner et al. 2004), which implies $\alpha \sim -2.3$ if $N \propto \text{SFR}$. This is again consistent with the bright-end slope of the LF being steeper than the low-mass end of the ICMF (see also Whitmore et al. 2007).

If $L_{\text{max}}$ is determined by random sampling, it is not a single number for a given $N$ but a random variable. For a power-law LF, the mean number of clusters brighter than some luminosity $L_{\text{max}}$ will be $\mu_b = N(L_{\text{max}}/L_{\text{min}})^{(1+\alpha)}$. 

*Phil. Trans. R. Soc. A* (2010)
The probability that there are no clusters brighter than a particular $L_{\text{max}}$ is then $P_b = e^{-\mu_b}$. It follows that the distribution of brightest-cluster luminosities, $\Psi(L_{\text{max}}) = dP_b/dL_{\text{max}}$, is

$$
\Psi(L_{\text{max}}) = -\frac{(1 + \alpha)}{L_{\text{max}}} (L_{\text{max}}/L_{\text{ref}})^{1+\alpha} \exp\left(-[L_{\text{max}}/L_{\text{ref}}]^{1+\alpha}\right)
$$

(5.3)

for $L_{\text{ref}} = L_{\text{min}}N^{-1/(\alpha+1)}$ (for a slightly different approach, see Maschberger & Clarke (2008)). The tail of this distribution approaches a power law with the same slope $\alpha$ as the LF itself at high luminosities. It can be shown that $L_{\text{ref}}$ is the mode of the log $L_{\text{max}}$ distribution and is related to the median of $\Psi(L_{\text{max}})$ as $L_{\text{max,med}}/L_{\text{ref}} = (\ln 2)^{1/(1+\alpha)}$. Figure 3b shows the distribution of residuals from the Weidner et al. (2004) fit compared with equation (5.3). As found in previous studies, the scatter around the fit is largely consistent with random sampling (Larsen 2002; Whitmore et al. 2007), although it may be significant that the largest residuals are mostly due to clusters in dwarf galaxies. In particular, further steepening of the LF towards the bright end would make such outliers less likely.

For Schechter-like ICMFs, the LF is not expected to be a single power law. Is this still consistent with the observed $L_{\text{max}}$–SFR relation? The solid and dashed lines in figure 3a show the $L_{\text{max,med}}$ relations for the model LFs in figure 2. We have further assumed $\Gamma = 0.50$, so that about 14 per cent of all stars formed are still in clusters after $3 \times 10^7$ years, at the end of the IM phase. Again, a single $M_c = 2 \times 10^5 M_\odot$ Schechter ICMF fits all spiral data quite well, but fails to reproduce the interacting systems. These clearly require a higher upper mass limit, and are instead well fitted by the $M_c = 2 \times 10^6 M_\odot$ model. The detailed model parameters are poorly constrained and other combinations of $\Gamma$ and IM provide equally good fits. However, the observed $L_{\text{max}}$–SFR relation is consistent with other constraints on the ICMF discussed in previous sections. It neither requires $M_c$ to scale in a simple way with the SFR, nor that the LF and ICMF are completely untruncated, uniform power laws. The fact that the $L_{\text{max}}$–SFR relation holds over such a relatively large dynamic range, even for Schechter-like ICMFs, is in part due to the fact that the age of the brightest cluster (and hence $\Upsilon$) will be a decreasing function of $N$ (or SFR) for fixed $M_c$, so that $L_{\text{max}}$ is sensitive to size-of-sample effects even if the MF is sampled up to near $M_c$ (Larsen 2009).

6. Concluding remarks

The use of star clusters as tracers of extragalactic stellar populations relies on a close link between star formation in general and cluster formation. While several recent studies have converged on a fraction of approximately 10 per cent of stars forming in clusters that remain bound for at least a few $\times 10^7$ years, the ratio of clusters to field stars varies enormously in old GC systems, both from galaxy to galaxy and within galaxies (Harris 1991, 2010). It is hard to rule out that cluster dissolution is partly responsible for these differences, particularly if it is mass independent (Whitmore et al. 2007), but so far there is no robust way to predict what fraction of GCs may have been lost over a Hubble time in this way. While
the timing of, for example, a major burst of star formation may be inferred from a peak in the cluster age distribution, the strength of such a burst remains much more poorly constrained.

Apart from differences in the formation efficiency of clusters, the mass spectrum may also vary with environment. There are hints that the ICMF in quiescent discs may be less top heavy than in violent starbursts, so that massive clusters predominantly trace the latter. Although GC formation is no longer viewed as ‘special’, this suggests that GCs formed under conditions that are more similar to those in present-day starbursts than in discs. Low-mass clusters, formed under quiescent conditions long ago, may have dissolved by now.

In the coming years, observations of molecular gas in external galaxies with the Atacama Large Millimeter/submillimeter Array are likely to provide a tremendous boost in our understanding of cluster formation under conditions that differ from those in local star-forming regions. The newly refurbished HST is more capable than ever, and will allow more detailed constraints on the age and mass distributions of clusters in different galaxies, so that the role of environment in determining the ICMF, disruption and the cluster-formation efficiency can be better quantified. On the theoretical front, cosmological simulations will provide a more detailed picture of galaxy formation and evolution, down to scales where individual clusters can be followed (e.g. Bournaud et al. 2008; Prieto & Gnedin 2008).

References


Phil. Trans. R. Soc. A (2010)
Review. Young massive star clusters


Review. Young massive star clusters


